Variations in Time-Dependent Forgetting: Insights From Directed Forgetting and Retrieval Practice Effects

Inaugural-Dissertation zur Erlangung der Doktorwürde der Fakultät für

Humanwissenschaften der Universität Regensburg

vorgelegt von

Anna Theresia Nickl

aus Nürnberg

Regensburg 2023

Erstgutachter (Betreuer): Prof. Dr. Karl-Heinz T. Bäuml
Zweitgutachter: Prof. Dr. Klaus W. Lange

Acknowledgments

This thesis would not have been possible without a number of people: First and foremost, I want to thank my supervisor, Prof. Dr. Karl-Heinz Bäuml, who encouraged me to take my work into a more methodological direction and to incorporate my interest in coding. I am very grateful for all the expertise and support he offered me as well as the confidence he showed in me to overcome obstacles. I am aware that the naturalness with which I was supported at the lab cannot be taken for granted.

I would also like to thank my (current and former) colleagues at the lab, who always lent me their moral, methodological and intellectual support: Petra Witzmann, Magdalena Abel, Oliver Kliegl, Lisa Wallner, Michael Wirth, Lukas Trißl, Verena Kriechbaum, Johannes Bartl, Sarah Meixensperger and Marilena Hirsch. Special thanks are in order to Petra Witzmann, who was immensely helpful in all things administrative and who always had an open ear during the years, as well as to Lena Abel and Oli Kliegl, who supported me hands-on when I needed it the most. Without you, the last year would have been even more challenging. Big thanks need to go the the lab assistants for all their help (among them Nicole Meisner, Stefanie Agar, Lorenz Kick and Sophia Weigl) and of course to all the participants who took part in my experiments and thankfully reappeared for their second scheduled meeting.

Last but not least, I want to thank my family and friends who have had my back in all my endeavors for years now and who managed to remain interested in what I was doing throughout this project, or at the very least feigned their interest very well - and who rallied around me when life unexpectedly got quite a bit harder for me last year. Ohne euch hätte das alles nicht funktioniert. Danke.

Contents

Parts of the present thesis have been published as:

Nickl, A. T., & Bäuml, K.-H. T. (2023). To-be-forgotten information shows more relative forgetting over time than to-be-remembered information. *Psychonomic Bulletin & Review*. Advance online publication. https://doi.org/10.3758/s13423-023-02330-1

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

Abstract

Memory performance for previously studied materials typically declines as time since study progresses, a phenomenon called time-dependent forgetting. Since the beginnings of memory research, how and why we forget has been an important area of interest, but many questions remain unanswered. One of them is whether there are variations in the rate of time-dependent forgetting depending on different experimental manipulations. This was examined within two fields in this thesis: Directed forgetting, where participants are asked to forget some, but not all previously studied information, and retrieval practice effects, where previously studied information is then either retrieved or restudied before a final test. These manipulations typically result in reduced memory for to-be-forgotten materials and increased memory for retrieval practiced materials compared to their respective control conditions. Retention of studied materials was measured at several delay intervals per experiment over the span of one week (Exp. 1-4) or three days (Exp. 5), and rates of forgetting were estimated by fitting retention data with a power function of time using maximum likelihood methods. Experiments 1 and 2 showed that time-dependent forgetting is reduced for to-be-remembered vs. to-be-forgotten information in item-method directed forgetting, where individual items are cued to be remembered or forgotten, while Experiment 3 showed that both kinds of information are forgotten at the same rate in list-method directed forgetting, where whole lists of items are cued to be remembered or forgotten. Experiments 4 and 5 demonstrated that retrieval practice, with and without feedback, can reduce time-dependent forgetting compared to restudy, but that restudy does not reduce time-dependent forgetting compared to a no-practice control condition. In addition, the findings demonstrate that the relative rate of forgetting can vary between conditions for some but not all experimental manipulations that introduce differences in memory performance between conditions and that differences in forgetting can also emerge when initial memory performance is similar. Variations in the rate of forgetting could be contingent on differences in degree of learning or on the involvement of specific processes that alter the underlying structure of the memory traces. Theoretical and methodological implications of the findings are discussed.

Chapter 1

Time-dependent Forgetting: Theoretical Overview and Research Methods

Humankind has long been fascinated with remembering and forgetting but has often held the two processes in very different regard. In Greek mythology for example, both a goddess of memory, Mnemosyne, and a goddess of forgetting, Lesmosyne, or Lethe, were recognized (see e.g. Sheard & MacLeod, 2005), but the former was more revered, illustrated by Hesiod's placement of Lethe among spirits with a negative connotation such as Hardship or Starvation in his Theogony, while Mnemosyne, mother of the Muses, is listed among the first named gods (Hesiod, ca. 700 B.C.E/1914). This fascination with both processes is not surprising given that remembering is a key factor of our sense of self (Conway, 2005; Klein & Nichols, 2012): Knowing who we are, where we come from and where we want to go all depend to a large extent on memory. According to Klein and Nichols (2012) this goes beyond the content of memories but extends to knowing that those memories are indeed our own. This dimension makes degenerative diseases that impair memory such as Alzheimer's particularly daunting and distressing (Werner et al., 2021), especially for episodic memories that involve personal experiences and contain information about the time and space those experiences occurred in (Tulving, 1972, 1993).

For most people, forgetting is sometimes desirable in the case of unwanted or unpleasant memories but is mostly seen as something to be battled. Accordingly, the classification of forgetting as something purely unfortunate is widespread, but within memory research it has long been agreed that forgetting is in fact a highly adaptive process that enables remembrance of important information (e.g. R. A. Bjork, 2011; McGeoch, 1932), already noted by William James (1890/1952), who wrote "if we remembered everything, we should on most occasions be as ill off as if we remembered nothing" (p. 445). Indeed, being able to forget is not without its merits: Everyday life is becoming increasingly complex, especially considering the amount of information that is always available at our fingertips in the form of constantly updated news, social media posts and digitized research resources. Remembering every single bit of information we encounter in a given day would very soon become overwhelming and would hamper targeted retrieval of required information. A classic and impressive example is the case of the mnemonist S described in detail by Luria (1968), who seemed to be unlimited both in the "capacity ... and the durability of the traces he retained" (p. 18), but who had trouble forgetting information that was no longer relevant for him, going so far as writing down and burning obsolete content in an attempt to forget. Similarly, Parker et al. (2006) described a woman, AJ, who felt dominated by her constant, autobiographical recollections and who stated "it is non-stop, uncontrollable and totally exhausting. ... Most have called it a gift but I call it a burden" (p. 35). For most of us, however, remembering and forgetting progress very differently: We are able to filter or update our memories when information is revealed to be outdated and most of the time, do not feel burdened by a surplus of memories. Quite the opposite: Indeed, anybody who has ever tried to remember something for some crucial event in the future has probably wondered whether there are ways to prevent or at least slow down forgetting, or has noticed that some memories are retained while others seem to be forgotten more quickly.

1.1 Mechanisms of Forgetting

There have been impressive demonstrations of (mostly semantic) acquired information enduring in memory for very long retention intervals, for example Bahrick's (e.g. 1979, 1984) studies on retention of Spanish vocabulary learned at school and the geography of cities over the span of several decades, which has been taken as evidence for a permanent storage of at least some memories (but see Neisser, 1984). On the other hand, it has also been demonstrated that memories can be distorted and changed over time: In a classic study, Bartlett (1932) described, among other lines of research, repeated retrieval attempts of previously studied materials, such as folk tales or pictures. His participants overwhelmingly introduced more and more changes into their recall, mostly conforming to stereotypes and schemas until, over time, only very little of the original memory was retained. According to Bartlett, remembering should be seen as a constructive process shaped by the presence, not as a mere reproduction of past information. How and why memories change, deteriorate over time, or become less accessible has been studied extensively since the end of the 19th century. Most of the findings are at odds with popular beliefs about our memory as a sort of diary or inner camera and instead paint a picture of malleability and fallacies, starting with the finding that memory performance usually decreases as time after study progresses (time-dependent forgetting; e.g. Ebbinghaus, 1885).

A factor that has been shown to influence forgetting, or the phenomenon that previously recallable information is not recallable at the present moment (Wixted, 2005), is the match – or mismatch – in retrieval cues. Studies investigating the role of test format have demonstrated that information that cannot be retrieved in a free recall format often can still be correctly recognized from an array of lures (Carpenter & DeLosh, 2006; Hogan & Kintsch, 1971). Additionally, the availability of appropriate retrieval cues, such as category names, can increase the proportion of correctly recalled information, illustrating the importance of distinguishing between availability and accessibility of information (Tulving & Pearlstone, 1966). Retrieval can also be influenced by information that was present but not focused on during encoding of other target information, which has been referred to as context (Klein et al., 2007; Murnane et al., 1999). Such contextual cues have been found to enhance retrieval of target information when made available during test and this holds both for physically (e.g. Godden & Baddeley, 1975) and mentally reinstating the study context (Sahakyan & Kelley, 2002; Wallner & Bäuml, 2017). At the same time, a mismatch in available context cues – when certain contextual information is present during study but not at test – can impair retrieval (e.g. Godden & Baddeley, 1975). Contextual information can involve environmental information, such as odors (Isarida et al., 2014), natural physical environments (Godden & Baddeley, 1975; but see Murre, 2021), and virtual reality environments (Shin et al., 2021), but also mental or physical states, such as status of drug use (Rickles et al., 1973). As context, both internally and externally, gradually changes over time (contextual drift; Howard & Kahana, 2002), a delay between study and test will introduce a mismatch between study and test context, thereby impairing access to and retrieval of studied materials as time since study progresses. Contextual fluctuation over time is therefore seen as a major contributing factor for forgetting (Mensink &

Raaijmakers, 1988, 1989), and has been incorporated into several theoretical models of memory (Karpicke et al., 2014; Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1981).

Another factor that has been used to explain many phenomena in remembering and forgetting is interference, where memory for target information is impaired by the presence of similar, competing information. For most of the 20th century, interference theory was the dominant view of why information becomes inaccessible (M. C. Anderson, 2003; M. C. Anderson & Neely, 1996; Wixted, 2004b). From the view of interference theory, memory for target information decreases over time not because of any inherent decay processes but because with time, additional, interfering information is encoded (M. C. Anderson, 2003) that competes for retrieval, introducing cue-overload (e.g. Wixted, 2004b, 2005) and impaired retrieval. Important findings were those of retroactive and proactive interference, the observation that the encoding of competing information both before (proactive interference; Underwood, 1957) and after (retroactive interference; Müller & Pilzecker, 1900) target information is studied can impair recall of the latter information compared to when no interfering information is encoded. Another particularly influential paradigm has been that of retrieval-induced forgetting, where participants first study a set of materials, typically items paired with their semantic categories (e.g. M. C. Anderson et al., 1994). Before a final recall test for all studied materials, some items out of some categories are retrieved, establishing three subgroups of items: retrieved items, related unretrieved items, and unrelated unretrieved (control) items. While retrieved items are recalled at a higher rate on the final recall test, related unretrieved items are impaired in comparison to control items, a finding that has been attributed to interference between retrieved and unretrieved related items (M. C. Anderson, 2003). Observations such as these demonstrate that forgetting can proceed differently for studied information depending on encoding and retrieval of associated information.

Early on into memory research, and competing with a purely interference-based view of forgetting, it was proposed that memory traces do decay over time according to how much they are used (law of disuse; Thorndike, 1913). This idea was discounted by many researchers, in part based on findings by Jenkins and Dallenbach (1924), who had their participants study a list of items and then either sleep or remain awake during the retention interval before the test. Memory for studied materials was better after sleep than after remaining awake, which was taken as support for interference theory, as interference of competing memory traces should be reduced during sleep, but as at odds with a decay perspective as such a process should have progressed also during sleep (see also McGeoch, 1932). An equally early idea was that of some sort of stabilizing process that protects memories from further decay or the assumption that decay lessens with time, first proposed by Müller and Pilzecker (1900) who observed a temporal gradient of retroactive interference in their experiments: Interfering information encoded early into a retention interval was found to be more detrimental to the previously encoded target information than interfering information encoded late into the retention interval. This idea of initially fragile memory traces that have to be stabilized for long-term retention was specified within consolidation theory (e.g. Dudai, 2004; McGaugh, 2000), which is frequently disregarded within memory psychology, but is much more widely accepted in neuroscience (e.g. Dudai et al., 2015; McGaugh, 2000; see also Wixted, 2004b). Indeed, several psychological models for memory exist that aim at explaining findings usually taken as support for consolidation with a consolidation-free mechanism (e.g. Brown & Lewandowsky, 2010; Yonelinas et al., 2019). These ideas involving natural decay and consolidation assume ongoing, underlying processes that impact the memory trace itself to explain rates of forgetting, and are thereby in contrast with ideas centering context-fluctuation and interference which focus predominantly on failures at retrieval.

Consolidation is assumed to transform newly encoded, initially fragile memory traces into stable long-term memory representations (Dudai et al., 2015; Müller & Pilzecker, 1900), with many researchers distinguishing between cellular (or synaptic) and systems consolidation (Dudai, 2004; Dudai et al., 2015). Cellular consolidation refers to processes immediately and within the first few hours after encoding that make the memory trace resistant against competing information and pharmacological interference (Dudai et al., 2015). In contrast, systems consolidation is thought to take place over a much longer time span of several days or even years (Dudai, 2004) and involves the reorganization of memories, whereby the initial hippocampus-dependence lessens and distributed networks in the neocortex become more important (Squire et al., 2015). Based on beneficial effects of sleep after study on subsequent retention, it has been proposed that sleep plays an important role in memory consolidation (e.g. Stickgold, 2005), attributed to repeated reactivation of hippocampal areas that had been active during study especially during slow-wave sleep (Dudai et al., 2015; Squire et al., 2015). A large array of findings has been taken as support for consolidation, starting with the observation of a temporal gradient in retrograde amnesia (Ribot, 1881/1882), where younger memories are impaired or lost to a higher degree than older memories. Since then, lesion studies in animals have supported an early hippocampus-dependence by demonstrating temporal gradients in induced retrograde amnesia (e.g. Clark et al., 2002; Winocur et al., 2013) and temporal gradients in brain activity after encoding have been observed in human subjects at least in some studies (Haist et al., 2001; Takashima et al., 2009; see also Squire et al., 2015). Consolidation could also adaptively vary for different kinds of memories (Cowan et al., 2021; Stickgold & Walker, 2013), shaping which memories are forgotten quickly and which are retained long-term. Such a distinction could rest on expectations whether memories will be relevant in the future or not (e.g. J. R. Anderson & Schooler, 1991; R. B. Anderson et al., 1997).

To this day, no unified theory of forgetting and long-term retention exists, and instead several different factors are thought to play some kind of role in forgetting (Wixted, 2004b), e.g. pro- and retroactive interference and context effects. Findings exist that are at odds with assumptions within standard consolidation theory (Nadel & Moscovitch, 1997; Winocur et al., 2013; see also Squire et al., 2015), and the classical view of interference with its focus on cue-overload that impairs retrieval has also been shown to be inadequate to describe natural forgetting (for overviews, see Wixted, 2004b, 2005). However, interference can also be conceptualized as unspecific activity (Müller & Pilzecker, 1900; for a modern investigation of their principal findings see Dewar et al., 2007) or as memory formation that degrades not-yet consolidated memory traces (Wixted, 2004b, 2005). This can be aligned with a naturally occurring trace decay process that is not solely due to the passage of time, which had been criticized in initial propositions of decay as a potential factor for forgetting (McGeoch, 1932). Such trace decay has for example been implemented in Wickelgren's (1974) single trace theory, where memories are assumed to decay due to interference over time and trace resistance slows down this process. More recently, Davis and Zhong (2017) proposed ongoing "active forgetting“ or erasure of unused memory traces (see also Hardt et al., 2013) as a fundamental competitor of consolidation to ensure homeostasis within the brain. Possibly, memory traces decay due to unspecific interference and need to be consolidated to be retained long-term but can also be actively erased, influencing how quickly they are perceived to be forgotten. It would be an interesting question following from this which variables affect whether memory traces are erased or consolidated. Existing research points to preferred consolidation of emotional content and motivational influences (e.g., Murty et al., 2017; Sharot & Phelps, 2004; see also Cowan et al., 2021; Stickgold & Walker, 2013), indicating overall a certain goal-relevant orientation (Cowan et al., 2021). As so many fundamental questions regarding time-dependent forgetting and variations within it remain unanswered, it might be a good idea at this point to take a step back and ”to assemble the basic facts and look for a common message“ (Wixted, 2004b, p. 241). A useful starting point could be to examine more closely what exactly the time course of forgetting looks like.

1.2 Measuring Time-dependent Forgetting

The obvious method to measure time-dependent forgetting is to collect data on retention of studied materials at different delays after study and then compare performance across delays. Memory performance can be measured in different ways, for example in ”savings“ (Ebbinghaus, 1885), odds (Wixted & Ebbesen, 1997), d’ (Donkin & Nosofsky, 2012), or, more commonly, in proportion or percentage retained (Averell & Heathcote, 2011; Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991). More importantly, the exact method of comparing performance across delays and between conditions can have substantial influences on the implications drawn from such studies. Most commonly, losses in memory are captured in absolute (retention at time 1 - retention at time 2) and relative terms (retention at time 1 - retention at time 2, relative to the amount retained at time 1) and this introduces an important chasm in the interpretation of forgetting rates. This is due to the fact that equality of absolute forgetting does not equate to equality of forgetting in relative terms (see Figure 1): If retention drops from 80 % to 60 % in group A, and from 60 % to 40 % in group B, both groups show the same absolute forgetting (-20%), but very different relative forgetting, -1/4 in group A, but -1/3 in group B. While from an ”absolute“ point of view then, forgetting does not differ between groups A and B, from a ”relative“ point of view group B shows a substantially increased rate of forgetting (see also Wixted, 2022a).

How exactly the rate of forgetting should be compared between conditions has been the point of some debate (e.g. Bogartz, 1990; Loftus, 1985; Slamecka, 1985; Slamecka & McElree, 1983; Wixted, 1990), for example whether the conditions should be compared horizontally or vertically. A comparison of absolute forgetting (vertical comparison, e.g. Slamecka, 1985; see also Figure 1A) has been indirectly favored in a lot of work examining long-term effects for different paradigms in memory research, by the primary use of ANOVAs as a method of analysis (e.g Abel & Bäuml, 2019; MacLeod & Macrae, 2001; Rivera-Lares et al., 2022; Roediger & Karpicke, 2006a). Examining interactions between the factors of condition and delay can then be used to compare absolute rates of forgetting (see Bogartz, 1990). Most of the existing studies however measured retention at only two retention intervals, which can only offer an

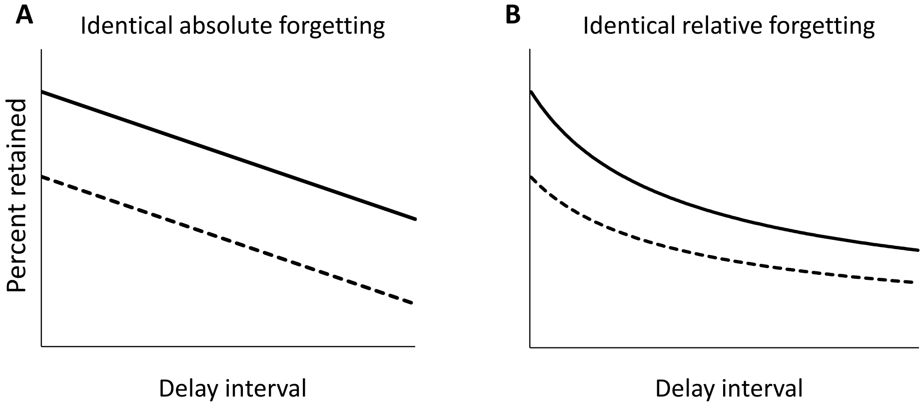

Figure 1: *Identical Absolute and Relative Forgetting.* Displayed are two conditions that vary in retrieval success at the shortest delay interval but show the same amount of absolute (A) or relative (B) forgetting . **(A)** the two conditions have the same vertical distance across the delay interval, constituting identical absolute forgetting. **(B)** the two conditions are represented by two power functions with identical proportional decay parameters and are therefore identical in relative forgetting. Over a long delay interval, the two conditions will converge.

approximation of the time course of forgetting. Wixted (2022a) recently argued that relative forgetting is of higher theoretical interest than absolute forgetting, the same way one would not use absolute losses in profit between companies of different sizes to compare how they are doing financially. He also pointed out that several existing theories on forgetting yield similar predictions for absolute forgetting but different predictions for relative forgetting, making relative forgetting more suitable to test those theories. For example, relative forgetting might be assumed to remain constant as time progresses (e.g. Mensink & Raaijmakers, 1988), speed up (Underwood & Keppel, 1963) or decrease (e.g. J. R. Anderson & Schooler, 1991; Brown et al., 2007), but in all of these cases, absolute forgetting per period of time would decrease with prolonged retention intervals.

An alternative approach to the measurement of time-dependent forgetting is to fit performance scores across delays with a mathematical function. This way, relative rates of forgetting can be compared between conditions, and the non-linear trajectory of forgetting can be described more closely. The curvilinear form of forgetting over time had first been demonstrated by Ebbinghaus (1885), who, over the course of several months (Murre & Dros, 2015), learned lists of nonsense syllables to criterion (correct recitation) and then measured how long it took him to reach this criterion again after different retention intervals spanning 30 minutes to 31 days (see Murre & Dros, 2015, for a replication). Memory performance was examined in terms of time saved on this second criterion test compared to the original criterion test (savings). Ebbinghaus remarked on the "initial speed ... [and] later slowness" (p. 103) of forgetting he measured and supposed that the effects of initial study in his case would only have disappeared completely after "infinitely long time" (p. 104).

Ebbinghaus proposed a logarithmic function to describe his data mathematically but remained noncommittal whether this function could also be used to describe forgetting in a more general way. Since then, researchers have fit a wide range of data sets, such as autobiographical memories, short-term memory in animals and both short- and long-term memory in humans (Squire, 1989; Rubin, 1982; Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991) with numerous mathematical functions, e.g. hyperbolic, linear, logarithmic, exponential and power functions, or mixtures of those (for comparisons, see e.g. Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991; see also Figure 2).

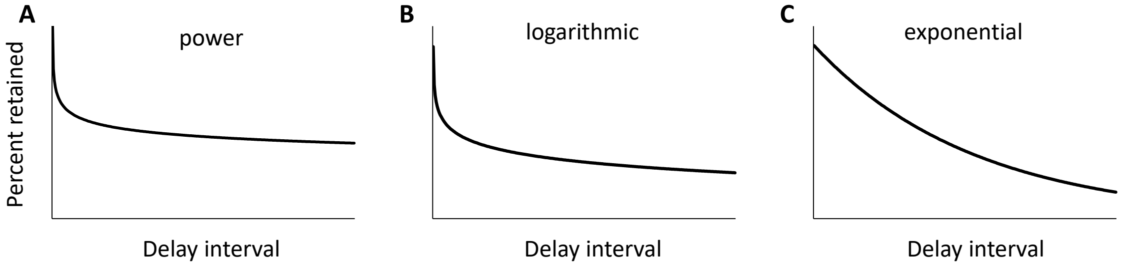

Figure 2: *Prominent Forgetting Functions.* Displayed are three commonly used functions in their two-parameter versions (see Wixted & Ebbesen, 1991), namely the power, $r(t) = at^{-b}$ (A), logarithmic, $r(t) = a - b \ln t$ (B) and exponential, $r(t) = ae^{-bt}$ (C) functions.

The choice of mathematical function has often been made by comparing goodness-of-fit criteria between functions (e.g. Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991), and can also be guided by form considerations, such as being defined at $t = 0$ or not taking on negative values (Wickens, 1998). Attempts have also been made to use theoretical constraints in this search (see also Wickens, 1998) and one of these is the assumption of a stabilizing process in memory, i.e. consolidation. As early as 1897, Jost argued in his second law that for two memories of equal strength, the older memory should show less decay than a younger memory, in line with Ribot's observation (1881/1882) of a temporal gradient in retrograde amnesia (see Wixted, 2004a, 2004b). Based on their experiments on retroactive interference, Müller and Pilzecker (1900) proposed a strengthening process that protects memories against interference. These ideas are consistent with the observation that time-dependent forgetting decelerates, with a steep initial drop in performance that then levels off until only little forgetting takes place during each time interval (e.g. Wickelgren, 1974; Wixted & Ebbesen, 1991, 1997). Another point of theoretical debate has been whether memory performance eventually degrades to 0 when measured over very long periods of time or whether a function with an asymptote greater than 0 is more appropriate (e.g. Averell & Heathcote, 2011; Rubin et al., 1999), but mostly it has been assumed that forgetting functions decay to 0 over time (e.g. J. R. Anderson & Schooler, 1991; Ebbinghaus, 1885; Wickelgren, 1972; for a theoretical discussion see Wixted, 2004a).

A promising candidate function is the power function of time which has been found to fit a wide range of data well (J. R. Anderson & Schooler, 1991; R. B. Anderson et al., 1997; Donkin & Nosofsky, 2012; Murre & Dros, 2015; Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991). In his single-trace theory, Wickelgren (1974) proposed an exponential-power function, with a power function component to capture trace fragility and long-term forgetting, as in a power function, the rate of loss decreases with time (Wixted, 2004a). Power functions can also produce data that conform to Jost's law and have accordingly been assumed to be in line with the concept of consolidation as a process that makes memory traces more resistant against forgetting over time (Wixted, 2004a, 2004b). This requirement is not met by all functions that can describe the curvilinear form of forgetting, as illustrated by the exponential function with its constant decay rate, which can be seen as incompatible with these theoretical ideas (Wixted, 2004a). The three-parametric power function (e.g. Carpenter et al., 2008; Siler & Benjamin, 2020; Wixted, 2022b) takes on the form $r(t) = a(1 + ct)^{-b}$, where a is memory recall at $t = 0$, c is a scaling unit, and b is the relative rate of forgetting. Other forms of the power function have also been used repeatedly, such as two-parametric versions of the form $r(t) = at^{-b}$, where a is memory recall at $t = 1$ (e.g. R. B. Anderson et al., 1997; Bäuml & Trißl, 2022; Wixted & Ebbesen, 1991, 1997), or $r(t) = a(1 + t)^{-b}$ (see e.g. Wixted, 2022a), where $c = 1$ and a is memory recall at $t = 0$ (for comparisons of the function versions, see Wixted, 2004a, 2022b).

1.3 Variations in Time-dependent Forgetting

While most of the research presented so far has been concerned with how relative time-dependent forgetting of a single condition can be described best, not much data exists yet on how mathematically described forgetting might vary between conditions (such as populations, materials, test formats or practice conditions). Comparisons between conditions might yield information on whether different functions are better suited to some but not other conditions, or whether the parameters within a function might differ systematically between conditions. Especially this latter area of research might in turn increase our understanding of how those conditions differ in the first place.

Clinical and developmental research has looked at differences in forgetting between groups. For example, several studies have investigated rate of forgetting for amnestic vs. healthy control subjects, reporting both normal and increased rates of forgetting for patient groups, depending on the exact kind of amnesia (e.g. Huppert & Piercy, 1979; Squire, 1981), suggesting that an impairment of forgetting rate might depend on the localization of brain damage. Similarly, forgetting of newly studied information might be increased only for patients suffering from certain types of dementia (Carlesimo et al., 1993). For the case of Alzheimer's disease specifically, the results seem to be mixed (e.g. Hart et al., 1987; Stamate et al., 2020; see also Rodini et al., 2022) and could depend on which time span is investigated. Also examined has been the role of age (e.g. Davis et al., 2003; Fjell et al., 2005; Giambra & Arenberg, 1993), with mixed results on whether rate of forgetting is increased for older individuals. Outcomes might however depend on materials, study schedules and test formats. These studies have typically used ANOVAs or numerical results to compare forgetting between groups across several delay intervals, mostly focusing on absolute forgetting. An exception is a study by Wickelgren (1975), who examined memory performance in children as well as in young and older adults in a recognition paradigm and found no differences in relative forgetting rates between age groups. In studies such as these that contrast different populations, the groups can vary on a number of dimensions, for example in how well or how quickly they can encode new materials, which can complicate the comparison of forgetting rates between groups. Overall, it is difficult to draw general conclusions about variations in time-dependent forgetting from studies like these.

Research also exists on the influence of delay on well-known memory effects, in which differences between conditions can be controlled to a higher degree. As indicated above, emotional materials seem to be forgotten at a lower rate than neutral stimuli (e.g. Sharot & Phelps, 2004), and a similar pattern can arise for materials associated with higher vs. lower reward-expectancy (Murty et al., 2017; for overviews see Cowan et al., 2021; Stickgold & Walker, 2013). Other areas of interest include for example retrieval-induced forgetting or degree of learning, but most of these studies have examined only one short and one long delay and/or focused on ANOVA and therefore absolute forgetting (e.g. Abel & Bäuml, 2019; MacLeod & Macrae, 2001; Rivera-Lares et al., 2022). This limits the informative value of these studies concerning variations in time-dependent forgetting. Less research exists that examined relative time-dependent forgetting, but the few studies so far show that variations in the rate of time-dependent forgetting might be possible under certain conditions. For example, Carpenter et al. (2008) investigated whether retrieval practice with feedback for previously studied materials can reduce forgetting compared to restudy when implemented shortly after study, but the results were mixed. Wixted (2022a) demonstrated a pattern of reduced forgetting rates for high vs. low degree of learning and Bäuml and Trißl (2022) showed that repeated selective practice of studied materials, implemented some time after study, can reduce forgetting compared to unpracticed materials. Overall, however, this issue remains under-researched, especially concerning more subtle differences between experimental conditions and not those that arise from comparing relatively extreme populations. Therefore, more research is needed on the effects of experimental manipulations on healthy, everyday forgetting.

1.4 The Current Project

Kahana and Adler (2002) argued that "the important question for theories of memory to address is how forgetting is affected by experimental manipulations and not what mathematical form the forgetting process assumes“ (p. 14) (see also Brown et al., 2007). Following this suggestion, the goal of this thesis is to further investigate whether there is evidence for different trajectories of time-dependent forgetting depending on the processes surrounding study, and under which circumstances forgetting can differ between conditions. That some information is forgotten more quickly while some other information can be retained over a long time is a very intuitive idea and would make systematic variations within forgetting rates probable. Apart from the influence of factors such as context-dependency or interference, the rate of forgetting could for example overall depend on probability of future use (J. R. Anderson & Schooler, 1991; R. B. Anderson et al., 1997; see also Cowan et al., 2021), with specific characteristics of memory traces and study conditions that can serve as proxies to gauge this probability, or the goal-relevance of memories. Finding such patterns would not only be of theoretical interest but could also be of use in the creation of lasting memories, for example in educational contexts, or in clinical applications where patient groups struggle with the retention of information.

In the five experiments presented here, participants studied a set of items and, after an experimental manipulation, were tested on their retention of the studied items after a certain delay interval. Per experiment, data from four (Experiments 1, 2, 4 and 5) or five (Experiment 3) delay intervals was used to estimate forgetting curves, by performing power function analysis. The empirical part of this thesis is split into two sections: In the first section (Chapter 2), three experiments on directed forgetting are presented, where participants are asked to remember one set of studied materials, but to forget another set. This allows for an examination of whether the explicit instruction to forget studied materials and the associated expectation that these materials will not be relevant for a future test can change the rate at which those materials are forgotten. Experiments 1 and 2 make use of the item-method of directed forgetting, where individual items are cued to be remembered or forgotten, while in Experiment 3 the list-method of directed forgetting is employed, where whole lists of items are cued to be remembered or forgotten. The typical effect of directed forgetting – better memory for information that is to be remembered – has been ascribed to different mechanisms in the two methods (e.g. B. H. Basden et al., 1993; Rummel et al., 2016): to processes acting at encoding in the item-method, and to processes acting mainly after encoding in the list-method. A comparison of both methods can thus be used to glean information about the respective relevance of both phases of memory processing for the rate of forgetting that can be observed.

The second empirical section of the thesis (Chapter 3) contains two experiments investigating the effects of retrieval practice (Experiments 4 & 5). In both experiments, studied materials were subjected to different kinds of practice after study: trying to recall the materials with or without feedback (retrieval practice), studying them again (restudy) or no practice. Previous studies have demonstrated that practicing retrieval of previously studied materials can enhance later retention compared to restudying those materials (see e.g. Roediger & Karpicke, 2006a), and this effect has been found to increase in size for longer retention intervals (e.g. Roediger & Karpicke, 2006b). In addition, several studies have shown a protective effect of retrieval practice for practiced materials, for example in the face of retroactive interference (e.g. Potts & Shanks, 2012; Halamish & Bjork, 2011). Different kinds of practice could also indicate whether materials will be relevant in the future (see J. R. Anderson & Schooler, 1991; R. B. Anderson et al., 1997; Cowan et al., 2021). The results of Experiments 4 and 5 can show whether certain practice strategies, such as retrieval practice, also induce a protective effect against forgetting and can be used to slow down the rate of time-dependent forgetting, a question that is of particular interest in educational contexts. Taken together, the results of this thesis can offer information on whether different experimental manipulations in memory research can lead to different rates of time-dependent forgetting by comparing conditions within the individual experiments, and, by taking into account the findings from all five experiments, under which circumstances forgetting rates can vary. This, in turn, can further inform our understanding of the processes governing forgetting itself. The findings of all five experiments and their implications as well as methodological issues are examined in a General Discussion (Chapter 4).

Chapter 2

Time-dependent Forgetting and Directed Forgetting

In memory research, several mechanisms have been proposed that ensure that less important or interfering information is less readily accessible than more important information at the time of its retrieval (M. C. Anderson & Hulbert, 2021), among them retrieval-induced forgetting (M. C. Anderson et al., 1994), where selective retrieval of only some studied information impairs retrieval of other studied information, and context-dependent forgetting due to a mismatch in contextual retrieval cues (Godden & Baddeley, 1975; Murnane et al., 1999). Common to these situations is that the assumption that a certain set of information will not be necessary in the foreseeable future or in the current moment is an implicit one: For example, in the case of context-dependent forgetting, there is no explicit declaration at the time of test that a set of information that was studied in a different context is unimportant or that it should be forgotten. Instead, information associated with the present test context will simply be more accessible than contextually mismatched information.

While situations such as these are quite common and a variety of them have been extensively researched (e.g. Bäuml et al., 2020; Klein et al., 2007; M. C. Anderson, 2003), there are also situations where the distinction between "important" vs. "unimportant" or "relevant" vs. "irrelevant" information is an explicit one. For example, when asked to renew a password, it is very clear that the old password is no longer relevant and can safely be forgotten - and maybe should even be forgotten to reduce future interference. Similarly, when asking for directions on the street, a passerby might start to give instructions, only to correct themselves a sentence later when they realize that they made a mistake. This second set of situations involves a more active or directed form of memory adaptation, in contrast to the more undirected factors contributing to forgetting detailed above. The idea here is that memory adaptation can be an intentional - as well as an adaptive and positive - process to reduce accessibility of interfering or redundant information.

2.1 Experimental Research Into Memory Adaptation

To assess active memory adaptation, several experimental paradigms have been used. One example is the think/no-think paradigm (M. C. Anderson & Green, 2001; M. C. Anderson et al., 2004; Hotta & Kawaguchi, 2009), in which participants first learn associations between cue and target words and are then asked for some pairs to actively suppress thinking of the associated target word when presented with the cue word (no-think pairs), while they are allowed to think of the target words for other pairs (think pairs). In a final test, recall performance for all originally studied pairs is measured, resulting in an impaired performance for no-think pairs compared to think pairs and, crucially, also compared to baseline pairs that were not presented again after initial study. This latter finding is taken as a sign of item suppression, especially as this finding has also been reported when new probes are used as cue words instead of the original cues. However, some replication attempts for these original findings have failed (e.g. Bulevich et al., 2006; Hertel & Calcaterra, 2005; Hertel & Gerstle, 2003; Meier et al., 2011), and some authors have argued that the central finding of impaired recall for no-think pairs could also (at least in part) be attributed to retroactive interference that arises when participants use the no-think trials to actively connect the cue word to a stimulus other than the associated target word (Bulevich et al., 2006; Hayne et al., 2006; Hertel & Calcaterra, 2005). The think/no-think effect therefore may not be especially robust, and, even if successfully measured, produces only relatively small performance differences (Bäuml, 2008). Additionally, suppression effects typically only emerge for a high number of no-think trials (e.g. M. C. Anderson & Green, 2001: 8-16 trials) and they seem to be rather short-lived, with the suppression effect dissipating after only a few hours in some research (e.g. Davidson et al., 2020; Fischer et al., 2011). Some participants even show rebound effects at longer delays (Noreen & MacLeod, 2014), with higher recall for no-think items than for either think or baseline items, reminiscent of findings by Wegner et al. (1987), who reported paradoxical rebound effects after participants had been asked to employ thought suppression.

2.2 Directed Forgetting: Prominent Findings and Theoretical Accounts

Another field in memory adaptation research uses the more straightforward approach of first having participants study a set of materials and then asking them to try and forget a subset of these materials. At the same time, a different subset of materials is to be remembered for an upcoming test. A final recall test for all studied materials is then used to assess memory performance for to-be-forgotten (TBF) vs. to-be-remembered materials (TBR). The area of research using this basic manipulation has sometimes been called "intentional forgetting", but most research has been done using the term of *directed forgetting* (DF). Not only are the experimental procedures in this field of research less complex than e.g. those in think/no-think research, they also typically lead to larger differences in memory performance for TBF vs. TBR materials.

Early research into directed forgetting saw a variety of experimental designs (e.g. Brown, 1954; R. A. Bjork et al., 1968; MacLeod, 1975; Muther, 1965). Since then, the research on directed forgetting has split into two main paradigms, the item-method and the list-method of directed forgetting (Figure 3; see e.g. Basden & Basden, 1996; B. H. Basden et al., 1993; MacLeod, 1998). In item-method directed forgetting (IMDF, for reviews, see M. C. Anderson & Hanslmayr, 2014; Bäuml, 2008), TBR and TBF materials are presented one at a time in randomized order, with a cue following each item denoting whether it is to be forgotten or to be remembered. The factor cue is therefore manipulated within-subject, as each participant encounters both TBR and TBF information. In contrast, list-method directed forgetting (LMDF, for reviews, see Bäuml et al., 2020; Sahakyan et al., 2013) usually consists of the presentation of two lists of items (List 1 and List 2), and a cue is given after the presentation of the first one, telling the subject whether to forget or to remember List 1. All subjects are asked to remember the second list. In this paradigm, cue is typically manipulated between-subjects, although some researchers present their participants with several pairs of lists, so that all participants encounter both a remember and a forget cue over the course of the experiment (e.g. Pastötter et al., 2012; Zellner & Bäuml, 2006). In the related paradigm of selective forgetting, only some of previously studied information is cued to be forgotten, typically resulting in reduced performance for these items only (see e.g. Delaney et al., 2009; Kliegl et al., 2018). For both IMDF and LMDF, final memory performance is generally worse for TBF information in comparison to TBR information, showing that participants can successfully forget information when cued to do so. This difference in recall between TBR and TBF information constitutes the DF effect. For LMDF, a second typical effect emerges: Recall of List 2 is often enhanced for participants cued to forget List 1 compared to participants who are cued to remember both lists (List 2 facilitation or enhancement). The contribution of demand characteristics to the impaired recall of TBF information may be assumed to be negligible for both paradigms, as monetary awards for additional recall of TBF information have been shown to have little effect (MacLeod, 1999).

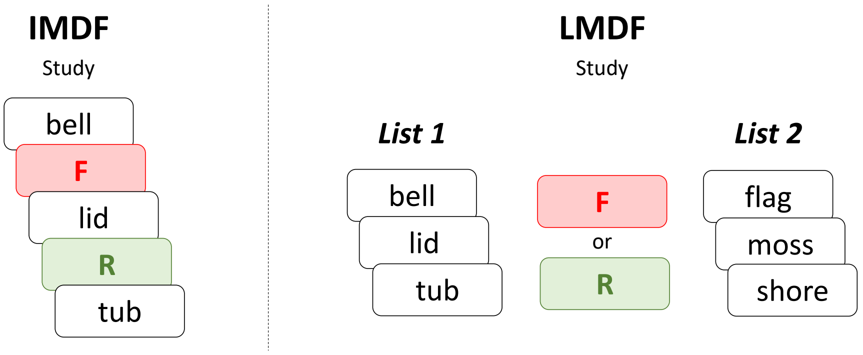

Figure 3: *Directed Forgetting Paradigms.* Displayed is the study phase of typical item-method (IMDF) and list-method directed forgetting (LMDF) experiments. Studied information is cued to be remembered (R) or to be forgotten (F). In a subsequent test phase after a retention interval of varying length, all studied items are tested, regardless of cue. Information cued to be forgotten is typically retained at a lower rate.

While early research on directed forgetting had used IMDF and LMDF interchangeably, comparative work on the two methods produced systematic differences between both methods. For example, IMDF has been observed both in free recall and in tests using item recognition (e.g. Davis & Okada, 1971; MacLeod, 1989), but LMDF is typically not present in recognition tests (e.g. Basden & Basden, 1996; B. H. Basden et al., 1993; Benjamin, 2006). Basden and Basden (1998) reported a study during which, before a delayed test, some participants were warned that TBF information would indeed be relevant for the upcoming test (contrary to the previous forget-cue). They found that this warning eliminated DF in LMDF (see also Sheard & MacLeod, 2005), but not in IMDF. Concerning theoretical ideas to explain these effects, the general consensus today is accordingly that both methods do in fact rest on very different mechanisms: In LMDF, as encoding of List 1 is completed by the time the forget or remember cue is given, any processes that shape directed forgetting in this task must therefore act after encoding. On the other hand, in IMDF the cue to remember or forget each item is given immediately after its initial presentation. Subjects may therefore engage in "maintenance rehearsal" (e.g. Fawcett & Taylor, 2008; Woodward et al., 1973) of each item while waiting for the corresponding cue, only deciding then which items to commit to memory. Arguably, processes acting on the encoding stage might therefore be contributing to IMDF, but not to LMDF. It has also been claimed that only LMDF truly captures intentional forgetting and that the forget-cue in IMDF instead functions as a do-not-learn-cue (Johnson, 1994; Sheard & MacLeod, 2005).

2.2.1 Item-method Directed Forgetting

The two most prominent theoretical accounts for IMDF accordingly both stress differences in encoding of TBR vs. TBF information: The *selective rehearsal account* assumes that items are maintained in memory until the presentation of a forget or remember cue, upon which TBF items are dropped from further rehearsal, while rehearsal for TBR items continues in anticipation of the recall test (B. H. Basden et al., 1993; R. A. Bjork, 1970). IMDF effects have been reported across various test formats, including item recognition and implicit memory tests (Basden & Basden, 1996; B. H. Basden et al., 1993), which supports the idea that IMDF is primarily a result of differential encoding.

The *inhibition account*, in contrast, assumes that TBF information is subjected to attentional inhibition (Zacks et al., 1996) or that encoding of TBF items is impaired by way of memory control processes that downregulate these items' memory representations (Fawcett & Taylor, 2008; Fellner et al., 2020; Rizio & Dennis, 2013). This line of work sees forgetting as an active process, contrary to the assumption of a merely passive dropping of TBF items from further rehearsal as it is assumed by the selective rehearsal account. Support for this more active view comes e.g. from findings that successful forgetting of TBF information requires effort (Fawcett et al., 2016; Fawcett & Taylor, 2008), as evidenced by longer reaction times to visual probes after TBF than TBR items. Additionally, brain regions that are activated for intentional forgetting have repeatedly been shown to differ from those involved in unintentional forgetting (Nowicka et al. 2011; Rizio & Dennis, 2013; Wylie et al., 2008). It has also been argued that both mechanisms, inhibition and selective rehearsal, play a role in IMDF, with the former acting on TBF items and the latter influencing TBR items (M. C. Anderson

& Hanslmayr, 2014; Fellner et al., 2020; Rizio & Dennis, 2013).

The last few years have also produced results challenging these prominent views on IMDF: Zwissler et al. (2015) found evidence in a series of experiments that TBF items were consistently better recognized than uncued items and also elicited lower false alarm rates than uncued items, in contrast with the finding of reduced memory performance for no-think items compared to baseline items in think/no-think research (e.g. M. C. Anderson & Green, 2001). Zwissler et al. (2015) did however find that recognition performance was better for TBR items compared to both other item types, demonstrating the basic IMDF effect. As the authors point out, this observation is at odds with both selective rehearsal and inhibition accounts. Schindler and Kissler (2018) were able to replicate this finding and additionally examined event-related potentials separately for the three types of information. The results indicated distinct and active early processing of both TBR and TBF information, which adds to the findings incompatible with a purely passive dropping of TBF information from rehearsal. More recently, IMDF has also been attributed to contextual processing that differs between TBR and TBF information (Chiu et al., 2021), assuming targeted unbinding of TBF items from their study context.

Other authors posit that the assumption of differential encoding for TBR and TBF information is not sufficient to explain IMDF. For example, Ullsperger et al. (2000) found quantitative differences in event-related potentials for deeply vs. shallowly encoded information, but reported qualitative differences in activation for TBR vs. TBF information, suggesting that the differences between TBR and TBF information go beyond the amount of encoding they receive. More recently, Rummel et al. (2016) used a multinomial modeling approach to dissociate the influence of storage and retrieval processes and found both to contribute to IMDF effects. Following up on this finding, Marevic and Rummel (2020) reported enhanced retrieval for semantically related TBF items when the test included a reinstatement of their category label, demonstrating that the provision of adequate retrieval cues can facilitate retrieval of both TBR and TBF information. While IMDF might therefore be due not only to processes acting at encoding, the general consensus is nevertheless that differences between TBR and TBF information in IMDF, but not in LMDF, take form already during the initial encoding phase.

2.2.2 List-method Directed Forgetting

Contrasting with the prominent theoretical accounts of IMDF, in LMDF, differences in memory performance between TBR and TBF information (List 1) are attributed to differences in retrieval processes, as for both conditions, encoding of List 1 is completed by the time the cue to remember or forget is given. This focus on retrieval processes is supported by such findings as the general absence of List 1 forgetting in item recognition testing (Benjamin, 2006; Abel et al., 2021). Several existing accounts attempt to explain both List 1 forgetting and List 2 enhancement with a single mechanism: For instance, the *selective rehearsal account* (R. A. Bjork, 1970; Sheard & MacLeod, 2005) assumes that after a forget cue only List 2 items are rehearsed while List 1 items receive no further rehearsal, leading to a decrease in memory performance for List 1 and an increase in performance for List 2 items.

However, Geiselman et al. (1983; see also Abel & Bäuml, 2019) presented their participants with lists that consisted of both intentionally and incidentally encoded items. They observed decreased retention for List 1 items following a forget cue, both for intentionally and incidentally encoded items, which is inconsistent with a pure selective rehearsal account. According to the *inhibition account* therefore, the instruction to forget leads to inhibition of the first-list study episode, thereby reducing the items' accessibility at test (R. A. Bjork, 1989; Geiselman et al., 1983). For forget-cued participants, this induces decreased recall of List 1 items, and, because of reduced proactive interference, increased recall of List 2 items. The inhibition account can explain various findings, such as the absence of List 1 forgetting under conditions of divided attention during study of List 2 (Conway et al., 2000), as inhibition is assumed to be an effortful process, or the crucial role of post-cue encoding: Several studies show that the emergence of List 1 forgetting hinges on the presentation of List 2 (e.g. Abel et al., 2021; Pastötter & Bäuml, 2007, 2010), and it has been argued that inhibitory processes should only arise in the presence of competing material. The ability of the inhibition account to explain various findings in the LMDF literature does, however, depend largely on the exact conceptualization of the inhibitory processes at play (Abel et al., 2021; Bäuml et al., 2020).

Finally, the *context-change account* assumes that the forget cue induces a mental context change that impedes recall of List 1 items due to the induced mismatch between study and test contexts (Mulji & Bodner, 2010; Sahakyan & Kelley, 2002) and at the same time enhances List 2 recall, again due to reduced proactive interference. While

LMDF and context-dependent forgetting behave similarly at short retention intervals, both paradigms differ in their results at longer intervals (Abel & Bäuml, 2017, 2019): While the effect of the forget-cue on List 1 recall could still be observed after a longer delay, the detrimental effect of a context change between study of List 1 and study of List 2 on recall of List 1 was only present after a short retention interval. This divergence in long-term behavior is at odds with a purely context-based explanation for LMDF.

Aside from the single-mechanism accounts described above, some researchers suggest two-factor accounts to explain List 1 forgetting and List 2 enhancement separately, as several studies have shown that the two effects are dissociable: For example, Benjamin (2006) used item recognition as test format and found no List 1 forgetting, but intact List 2 enhancement. Conversely, Aslan et al. (2010) examined several age groups and, for first graders, reported intact List 1 forgetting, but no List 2 enhancement. Bäuml et al. (2008) found different oscillatory EEG correlates for List 1 forgetting vs. List 2 enhancement, adding to the idea that the two effects are mediated by different mechanisms (see also Sahakyan & Delaney, 2003, 2005; Zellner & Bäuml, 2006). Sahakyan and Delaney (2003, 2005) follow the context-change account for List 1 forgetting (Sahakyan & Kelley, 2002) and, initially based on post-experimental questionnaires, additionally assume a change in encoding strategy that participants adopt after the presentation of a forget cue and that leads them to encode List 2 more efficiently. Support for this idea comes e.g. from the finding that no List 2 enhancement arose when both forget- and remember-cued participants were instructed to use a deep encoding strategy for List 2 (Sahakyan & Delaney, 2003), and from the absence of List 2 enhancement for incidentally encoded items in a between-subjects design (Sahakyan & Delaney, 2005).

It has also been argued that the forget cue results in a "reset of encoding" (Pastötter & Bäuml, 2010; for a review see Pastötter et al., 2017), whereby memory load and inattention, built up during study of List 1 (Sederberg et al., 2006; Pastötter et al., 2008), are canceled or reduced in response to the forget cue. This is assumed to improve encoding of early List 2 items, in line with findings that List 2 enhancement is typically restricted to the first few items of List 2, resulting in an enhanced primacy effect for forget-cued participants (Pastötter & Bäuml, 2010; Pastötter et al., 2012). This finding is also at odds with the strategy based account of Sahakyan and Delaney (2003, 2005), as a change in adopted study strategy should improve encoding of all List 2 items (Pastötter & Bäuml, 2010). Support for the reset of encoding hypothesis also comes from the examination of Alpha amplitude in EEG, as low Alpha amplitude had previously been linked to high encoding efficiency (Sederberg et al., 2006): Hanslmayr et al. (2012) found an increase in Alpha amplitude from List 1 to List 2 for remember-cued, but not for forget-cued participants, indicating a selective change in encoding efficiency for List 2 following a forget-cue.

2.3 The Role of Delay for Directed Forgetting

Taken together, the existing literature points to different mechanisms responsible for DF – the difference between TBR and TBF materials in IMDF, and the amount of List 1 forgetting in LMDF – under the two methods: processes acting at encoding in IMDF, and processes acting at retrieval in LMDF. Even taking into account the newer work on IMDF that argues for a retrieval component to this effect (Rummel et al., 2016; Marevic & Rummel, 2020), differences between TBR and TBF information can arise already at encoding in IMDF and for List 2 enhancement in LMDF, but only after encoding for List 1 forgetting in LMDF. The two methods therefore offer the opportunity to investigate how relative and absolute time-dependent forgetting behaves for TBR vs. TBF materials that differ (IMDF) or do not differ (LMDF) in encoding, which could improve our understanding of whether time-dependent forgetting varies systematically between conditions.

2.3.1 Previous Findings

Most studies on IMDF and LMDF so far examined directed forgetting after a short retention interval, leaving it unclear whether TBR and TBF information in the two paradigms is forgotten at different rates. However, a few studies also employed longer delay intervals between study and test, allowing for a preliminary estimation of whether both effects can be detected at longer delays and therefore whether the memory adaptation introduced in the two methods is long-lasting or not. Due to the use of ANOVA-based analyses, these studies provide an examination of absolute time-dependent forgetting. For IMDF, the effect has been reported to persist for delays of 90 min (Scullin et al., 2017) and 6 hrs (Saletin et al., 2011), using free recall as test format. A comparison of the IMDF effect across two different delay intervals has been conducted using item recognition at test, immediately after study and one or two weeks after study (MacLeod, 1975, 1989). A significant effect arose in both delay conditions, with a similar difference in memory performance between the TBR and TBF items.

Regarding LMDF, a few studies examined directed forgetting effects for a long-delay interval between 20 min and 24 hrs and compared the effect with a short-delay condition of a few minutes (Abel & Bäuml, 2017, 2019; Hupbach, 2018; Scully & Hupbach, 2020). These studies used either free or cued recall as test format. Results indicated intact List 1 forgetting also after the longer delays, with a similar size of the effect as after the shorter delay condition. Joslyn and Oakes (2005) conducted a diary study for autobiographical memories across two weeks, with some participants receiving a forget-cue for the events of the first week (List 1). The 1-week delay therefore consisted of the recording of week-2 (List 2) items. The authors reported intact LMDF for first week items, but included no short delay interval. In some studies (Hupbach, 2018; Scully & Hupbach, 2020), results also generalized from List 1 forgetting to List 2 enhancement. Two other studies reported a reduction or elimination of LMDF after a longer delay: Liu (2001) found LMDF to be numerically reduced after 22 min in comparison to a 3 min baseline condition, but reported a non-significant interaction between the factors cue and delay, indicating no influence of the length of delay on the size of the DF effect. Shapiro et al. (2006) investigated directed forgetting of product attributes and found intact LMDF after 3 min, but no significant DF effect after 18 min. At the same time, List 2 enhancement was present at both delay intervals. However, the authors may have induced a mental context change in addition to the forget-instruction, possibly skewing the results (for a discussion of delay effects in LMDF, see Abel & Bäuml, 2017).

These findings on IMDF and LMDF suggest that directed forgetting effects may be permanent in nature, and, for most of the studies, remain roughly of the same absolute size for short and long retention intervals. However, research on the issue is scarce, and, at least in LMDF, has mostly included retention intervals of not more than a single day. An exception is a study reported in a book chapter by Basden and Basden (1998): In this experiment, both IMDF and LMDF were examined at a short delay and a long delay interval of one week between study and test, with free recall as test format. In both tasks, the results showed a numerical directed forgetting effect that was reduced in comparison to immediate recall. However, in both cases, no statistical analysis was reported, leaving it unclear whether the numerical differences were significant. Taken together, these studies allow only tentative assumptions on how the size of DF in both methods changes over time, and how absolute time-dependent forgetting varies for TBR vs. TBF information. As so far, IMDF and LMDF have only been studied at one or two retention intervals within one study, and the analyses have been restricted to the use of ANOVAs, these studies allow no elaborate conclusions on whether the relative time-dependent forgetting of TBF information differs from time-dependent forgetting of TBR information.

2.3.2 Theoretical Expectations

Additionally, none of the existing theoretical accounts make any clear-cut predictions regarding possible differences in (absolute or relative) time-dependent forgetting for TBR and TBF information. For LMDF, it has been pointed out that neither the selective rehearsal account nor the inhibition account in their original form are aimed at long-term DF (Abel et al., 2021), making additional assumptions necessary for long-term effects. To illustrate, using a selective rehearsal explanation for both IMDF and LMDF, persisting DF effects seem plausible, at least when additional assumptions are made for LMDF (Abel et al., 2021): The better rehearsed TBR materials should continue to be better rehearsed at longer delays, as long as TBF information is not selectively rehearsed during the delay. A variant of the selective rehearsal account of LMDF instead assumes that additional rehearsal during the delay should continue to be selective for TBR materials (Abel et al., 2021; MacLeod et al., 2003), and such a line of argument could also be applied to IMDF. Predictions regarding relative forgetting can potentially be derived from recent work by Wixted (2022a), who re-analyzed studies that manipulated degree of learning – implemented as different numbers of study trials – and that collected data across several retention intervals. Wixted observed that relative forgetting, captured by a power function of time, was consistently smaller for stronger compared to weaker information. If this finding can be applied to DF, relative forgetting should be reduced for stronger TBR information in comparison to weaker TBF information for both methods of DF. However, as IMDF and LMDF differ in the assumed contribution of processes working at encoding, differences between both methods may nonetheless emerge.

For the inhibition accounts of IMDF and LMDF, long-term predictions depend on the exact conceptualization of inhibitory processes. It has been observed that some inhibitory effects, for example as they are assumed in retrieval-induced forgetting (e.g., M. C. Anderson et al., 1994), seem to dissipate with longer retention intervals (Abel & Bäuml, 2014; MacLeod & Macrae, 2001; see also Bäuml & Kliegl, 2017). It has also been assumed more generally that inhibition should only temporarily reduce the accessibility of affected items (R. A. Bjork, 1989; Storm et al., 2012). At the same time, inhibition in LMDF could also be assumed to be long-lasting (Abel & Bäuml, 2019; Abel et al., 2021) to explain the observations of persisting DF effects, and, if inhibition in IMDF critically alters encoding of TBF information, these effects could also be assumed to persist across longer delays. In the case of inhibition, predictions regarding relative forgetting would be speculative, and could differ between both methods of DF, as the assumed inhibitory processes likely differ markedly for IMDF and LMDF.

Clear predictions regarding long-term behavior can however be made based on the context change account of LMDF, according to which the DF effect should dissipate across longer delays: The mental context change between List 1 and List 2 introduced by the forget-cue should become negligible compared to a remember-cue for longer delays between the study phase and the test phase (e.g. Abel & Bäuml, 2017, 2019; Sahakyan & Kelley, 2002; see also Divis & Benjamin, 2014). Such a dynamic would necessitate more rapid forgetting of TBR information than of TBF information, and is at odds with accounts of persisting LMDF at longer delays (Abel & Bäuml, 2017, 2019). For the accounts that assume several contributing mechanisms (such as a contribution of both selective rehearsal and inhibition to IMDF; e.g. Fellner et al., 2020) predictions for forgetting rates depend not only on the exact conceptualization of these mechanisms but also on their balance of contribution. Accordingly, the present set of experiments is not designed to test predictions of individual accounts. Instead, the aim is to investigate what kinds of long-term behavior existing accounts need to be able to explain.

2.4 Goals of Experiments 1-3

Experiments 1-3 of this thesis were designed to investigate the role of delay for both IMDF and LMDF. Aside from purely examining whether both DF effects can be observed at longer delays, it was also of interest whether the absolute sizes of the effects change across longer delays, and, using power functions of time, whether TBR and TBF information differ in their amount of relative time-dependent forgetting between methods. This was necessitated by a scarcity on existing studies manipulating the length of the retention interval between study and test in both methods, and the absence of studies that do so for more than two delay intervals. The results of these experiments will provide an addition to the theoretical arguments pertaining to IMDF and LMDF, as any successful theoretical approach needs to be able to explain the long-term behavior of both effects. Additionally, a comparison of IMDF and LMDF in regard to rates of time-dependent forgetting can also shed light on fundamental decay processes in human episodic memory, as so far, only a few studies have contrasted relative time-dependent forgetting of different conditions (e.g. Carpenter et al., 2008; Siler & Benjamin, 2020; Wixted, 2022a). If the rate of forgetting also depends on the probability of future use (J. R. Anderson & Schooler, 1991; R. B. Anderson et al., 1997), or how goal-relevant information is (Cowan et al., 2021), TBF information should be generally forgotten at a higher rate than TBR information that is cued to be relevant for the final test.

The current experiments therefore include four (Experiments 1 & 2) or five (Experiment 3) delay intervals, spanning from 3 min to 7 days. This allows for an extension of previous results on long-term DF effects. Aside from a traditional ANOVA-based analysis, in line with most of the existing literature, a power function analysis was conducted. This second analysis aims at a comparison of forgetting rates for TBR and TBF information, to examine whether the two differ within each method. By adopting the same analysis for both IMDF and LMDF, a qualitative comparison of the two methods is also possible.

2.5 Experiment 1: Item-method Directed Forgetting (Free Recall)

2.5.1 Method

Ethical Considerations

All reported studies were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. In all five experiments reported in this thesis, all subjects gave their spoken informed consent and took part in the experiment in return for either course credit or a compensatory amount of money.

Participants

120 participants took part in the experiment ($M = 24.12$ years, range = 18-30 years, 86 female). They were recruited mainly from Regensburg University, as well as by placing online advertisements in students' groups in Germany. 80 % of the participants were currently enrolled at university, while the remaining subjects reported to be employed. In all five experiments, all subjects spoke German as their native language and reported

no neurological or psychiatric disease. Participants were distributed equally across the four between-subjects conditions, yielding $n = 30$ participants in each condition. Previous work on IMDF (using both free recall and item recognition testing) and on time-dependent forgetting mostly reported large effects of IMDF and time-dependent forgetting ($d > 0.80$; e.g., Kliegl et al., 2019; MacLeod, 1975; Scullin et al., 2017). Sample size in the present experiments was therefore determined on the basis of a power analysis with G*Power (version 3.1.9.7, Faul et al., 2009) using $\alpha = 0.05$ and $\beta = 0.20$ as well as effect sizes of $d = 0.80$ for expected time-dependent forgetting and expected IMDF.

Materials

20 concrete, unrelated nouns (4-6 characters) were drawn from the CELEX database, using Wordgen v1.0 Software Toolbox (Duyck et al., 2004) and divided into two sets of ten words each. The items of each set were all paired with either an instruction to forget or an instruction to remember them for an upcoming test.

Design

The experiment had a 2 (CUE: forget vs. remember) × 4 (DELAY: 3 min vs. 1 day vs. 3 days vs. 7 days) mixed factorial design. CUE was manipulated as a within-subject factor, whereas DELAY was varied between subjects. In each delay condition, the assignment of the item sets to either the forget or the remember instruction was counterbalanced across participants.

Procedure

Data collection took place during the Covid-19 pandemic, and, due to safety reasons, was therefore conducted via zoom meetings (Zoom Video Communications), in which subjects and experimenters were connected by live web-cam and microphone feeds. No voice or video recordings of participants were made. For all participants, the experiment began with a collection of their demographic information and an explanation of their rights.

For participants in the 3 min delay condition, the experiment took place during a single session, while for all other participants two zoom meetings were held with one, three, or seven days between them. The second meeting was always scheduled for the same time of day as the first one (± 2 hrs). During each session, the experimenter shared their screen for stimulus presentation and instructed subjects orally. All instructions were also visible on screen. The software OpenSesame (version 3.3, Mathôt et al., 2012) was used for stimulus presentation and balancing.

Regardless of condition, the experiment started for all participants with the study phase, during which all 20 items were presented individually in the middle of the screen (4 s). Each item was followed either by the instruction "FORGET" in red font or "REMEMBER" in green font (1 s). Presentation order was pseudo-randomized with each cue type presented no more than three times in succession. Subjects were informed beforehand that only items followed by "REMEMBER" would be relevant for an upcoming test at the end of the experiment and that items followed by "FORGET" could be discarded from memory. Following the study phase, all participants, except for those in the 3 min delay condition, were asked to count backwards in steps of seven from a three-digit number for 2 min. This served as a recency control and was intended to hamper active rehearsal of studied materials during the delay. Subjects were then dismissed and asked to return to their second scheduled meeting one, three, or seven days later (± 2 hrs). The second meeting began with a 3 min distractor task of solving Raven's Standard Progressive Matrices (Raven et al., 2000). The task was self-paced and subjects gave their answers orally. The participants in the 3 min delay condition immediately proceeded to this task after the study phase. Finally, all participants were given a 4 min free recall test for all items that had been presented during study, regardless of cue. Accordingly, they were informed that they had been misinformed during study and that in fact all studied items, regardless of the instruction that had followed them, were relevant for this test. Subjects typed their answers into the zoom chat, one word at a time and in any order they chose. Afterwards, subjects were debriefed, thanked, and compensated for their participation.

Fitting Procedures

Maximum likelihood methods were used to fit the three-parametric power function of time, $r(t) = a(1+ct)^{-b}$, to the recall rates of the two cue conditions using group average data (see Appendix A). Time was measured in days since the end of the practice phase. For both cue conditions, a separate power function model was generated and compared to a statistical baseline model that described the recall rates of the four delay conditions as the product of four independent binomial distributions. Parameters were estimated by maximizing the likelihood of the power function model. This likelihood was then compared to the likelihood of the baseline model using the likelihood ratio, which led to an approximate χ^2-test with two degrees of freedom to examine whether the power function described the recall rates well (Bäuml & Trißl, 2022; Trißl & Bäuml, 2022; see also Riefer & Batchelder, 1988; Myung, 2003; Wickens, 1982). Degrees of freedom for this test depend on the difference in number of parameters between the more general and the restricted model: In this case, the (more general) baseline model contains four parameters (four recall rates), while the restricted model contains only two (parameters a and b).

Before running these central analyses, however, a common scaling parameter c for the two cue conditions was estimated, following previous studies that had used power function analysis to compare experimental conditions (see Carpenter et al., 2008; Siler & Benjamin, 2020). For this, a restricted power function model, in which parameters a and b were allowed to vary freely between conditions but parameter c was restricted to be the same for both conditions, was compared to a more general model, in which all three parameters in both conditions were allowed to vary freely. The fit of this restricted model was tested using maximum likelihood methods and a χ^2-test with one degree of freedom. This restricted model described the data equally as well as the more general model, $\chi^2(1) = 1.21$, with $c = 4.90$ as the best fitting scaling parameter. This parameter estimate was then used for all further analyses. As a result, a χ^2-test with two degrees of freedom was used to evaluate the fit of the power function (with its remaining two free parameters a and b) to the (four) recall rates of each cue condition.

In the next step, it was examined whether parameters a and b of the power function varied significantly between cue conditions. For this, the data sets of both conditions were combined and the fit of a more general model that allowed parameters a and b to vary freely between conditions was compared to that of a more restricted model in which either parameter a or parameter b was restricted to be the same for the two conditions. Again, the comparison was based on maximum likelihood methods, resulting in a χ^2-test with one degree of freedom (see also Bäuml & Trißl, 2022; Trißl & Bäuml, 2022). All fitting procedures were written in R (R Core Team, 2021) and implemented in R Studio (RStudio Team, 2020), using optim() from the R package *stats* (version 4.1.1) with a Nelder-Mead method for maximization. All other analyses were carried out in IBM SPSS Statistics for Windows, Version 28.0 (IBM Corp., Armonk, NY).

2.5.2 Results

Analysis of Variance

Figure 4 shows mean recall rates for both TBR and TBF items across all four delay intervals. A 2 × 4 mixed-factors ANOVA with the within-subject factor of CUE (forget, remember) and the between-subjects factor of DELAY (3 min, 1 day, 3 days, 7 days) produced a main effect of CUE, $F(1, 116) = 375.29$, $MSE = .02$, $p < .001$, $\eta_p^2 = 0.76$, with higher recall for TBR than for TBF items, and a main effect of DELAY, $F(3, 116) = 31.72$, $MSE = .03$, $p < .001$, $\eta_p^2 = 0.45$, reflecting time-dependent forgetting. Additionally, the ANOVA showed a significant interaction between the two factors, $F(3, 116) = 10.61$, $MSE = .02$, $p < .001$, $\eta_p^2 = 0.22$, suggesting a decrease in the size of the IMDF effect with delay. In line with this finding, follow-up paired t-tests between TBR and TBF items demonstrated significant IMDF at all four delay intervals, all $ts(29) > 7.59$, $ps < .001$, $ds \geq 1.39$, with effect size d decreasing with increasing delay (from $d = 2.39$ after 3 min to $d = 1.39$ after one week).

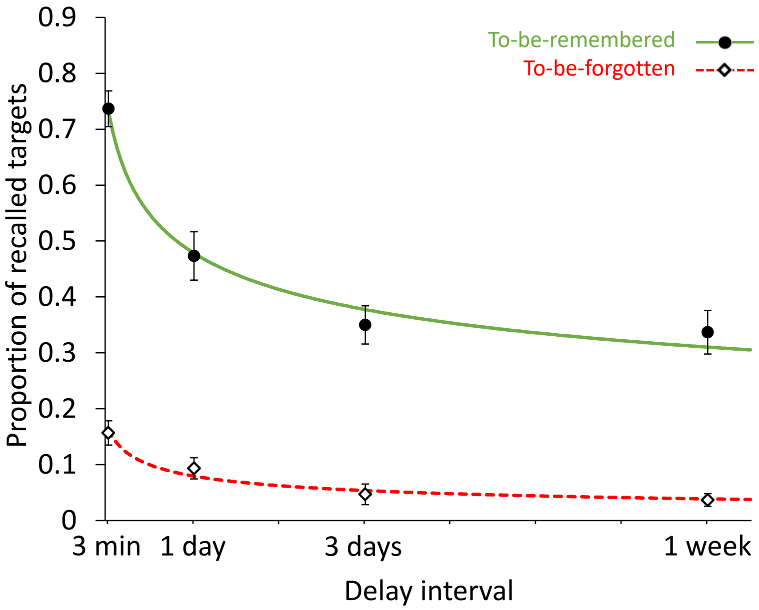

Figure 4: *Results of Experiment 1.* Mean recall rates for both cue conditions are displayed along with the best-fitting power functions. Recall of both the to-be-remembered items (TBR: filled circles, solid green) and the to-be-forgotten items (TBF: open diamonds, dashed red) showed time-dependent forgetting, described by a power function of time. The TBF items showed a higher relative rate of forgetting, reflected in a larger forgetting rate parameter, than the TBR items (see Table 1). Error bars represent ± 1 standard error.

Power Function Analysis

When fitting the power function to the recall rates of both cue conditions, results showed that the function described the recall rates of both conditions well (Figure 4), as is reflected in $\chi^2(2)$ values of ≤ 1.99 for the single cue conditions (Table 1). Indeed, the power function explained most of the variance in the data, as is represented in r^2 values of $\geq .971$ in both conditions. Estimates for parameter a were found to differ significantly between conditions, $\chi^2(1) = 222.71$, showing an effect of directed forgetting. At the same time, the cue conditions differed also in parameter b, $\chi^2(1) = 4.88$, reflecting a higher forgetting rate for TBF compared to TBR information.

Table 1

Best Power Function Fits and Explained Variance for Experiment 1

Condition	**a**	**b**	**$\chi^2(2)$**	**r^2**
To-be-Remembered	0.735	0.242	1.99	.986
To-be-Forgotten	0.162	0.403	1.13	.971

2.5.3 Discussion

Using analysis of variance and follow-up t-tests, the results of Experiment 1 show IMDF to persist for all three longer delay intervals (1 day, 3 days, 7 days), which is consistent with the few prior observations of a lasting IMDF effect with free recall testing (e.g., Basden & Basden, 1998; Scullin et al., 2017). At the same time, the present experiment goes beyond these previous findings by measuring IMDF for more than two delay intervals for the first time. When applying maximum likelihood methods, recall rates of the TBR and TBF items followed typical time-dependent forgetting and mean recall rates of both item types were well explained by the power function of time. Crucially, TBR and TBF items differed significantly in forgetting rates: Forgetting rate parameter b was larger for the TBF than for the TBR items, suggesting increased time-dependent forgetting for the TBF items.

Following up on Experiment 1, Experiment 2 was aimed to examine whether the present findings generalize from free recall to item recognition testing. It has been observed that participants tend to recall items of higher memory strength before items of lower strength in free recall (Wixted et al., 1997). This dynamic could potentially induce output interference (see e.g. Bäuml, 2008; Roediger, 1974; A. D. Smith, 1971), where recall decreases for later recalled information, thereby systematically impairing recall of the weaker items. If this pattern were to generalize to (stronger) TBR and (weaker) TBF items and participants selectively rehearsed the TBR information also during the delays (MacLeod et al., 2003), then the posited prioritization of the TBR items at test might even increase with the length of the delay interval. This might lead to an enhanced forgetting rate for the TBF relative to the TBR items. If recall dynamics like these contributed to or created the difference in forgetting rates observed in Experiment 1, then this difference should disappear with item recognition as a final test format. In item recognition, testing order is experimenter-guided and random, circumventing possible output order effects induced by strength. As item recognition tests typically yield a higher memory performance, such a test format could also reduce possible floor effects, which might have been present in Experiment 1's recall of TBF items. It should be noted, however, that the presence of a floor effect in the mean recall rates of the TBF items would imply that the actual relative forgetting rate of these items would have been even higher than the estimated parameter b suggests, as the floor effect would have created an upper bound for the size of parameter b. The presence of a floor effect thus would not affect the main conclusion from Experiment 1, i.e., that TBF items show faster forgetting than TBR items.

2.6 Experiment 2: Item-method Directed Forgetting (Item Recognition)

2.6.1 Method

Participants

120 new participants were recruited ($M = 22.26$ years, range = 18-30 years, 90 female), again mainly from Regensburg University but also by online advertisements. 92.5 % of the sample were currently enrolled at university, all other participants were either employed or doing vocational training. Participants were again distributed equally across the four between-subjects delay conditions, yielding $n = 30$ participants in each condition. Sample size followed Experiment 1.

Materials

In addition to the 20 nouns used in Experiment 1, another 60 concrete, unrelated nouns (4-7 characters) were drawn from the CELEX database, again using Wordgen v1.0 Software Toolbox (Duyck et al., 2004). 40 of these words were used during study, while the remaining 40 were used as lures during the final test. The 40 study items (20 words from Exp. 1 and 20 new words) were split into two sets of 20 words each, which were paired with either an instruction to forget or an instruction to remember them for an upcoming test.

Design

The experiment had the same 2 (CUE: forget vs. remember) × 4 (DELAY: 3 min vs. 1 day vs. 3 days vs. 7 days) mixed-factorial design as Experiment 1. CUE again served as a within-subject factor, whereas DELAY was varied between subjects. In each delay condition, the assignment of item sets to instructions was counterbalanced across participants.

Procedure

Again, data collection took place via zoom meetings. The experimental procedure was identical to Experiment 1 except for the following changes: (a) During study, items were presented for 1.5 s each to avoid ceiling effects, while cues were shown for 1 s as in Experiment 1. Presentation order was again pseudo-randomized with each cue type presented no more than three times in succession. Before the final test, participants were again informed that they had been misinformed earlier and that all items shown during study were to be considered relevant for the test. (b) At test, study and lure items were shown individually for 5 s. Subjects were asked to respond orally with “old” for words they thought they had seen during study, and with “new” for words they thought were new, irrespective of the instruction the word had been followed by during study. The experimenter recorded all responses by pressing corresponding keys on their keyboard. At test, old and new words were intermixed pseudo-randomly, with a maximum of three old or three new words presented in succession.

Fitting Procedures

The fitting procedure was identical to the one employed in Experiment 1. Again, before running more detailed analyses, a common scaling parameter c was estimated for both cue conditions, using the same procedure as in Experiment 1. Like in Experiment 1, the restricted power function model with a common scaling parameter c for both cue conditions described the hit rates equally as well as the more general power function model in which the parameter varied freely across conditions, $\chi^2(1) = 0.57$, with $c = 12.08$ as the best fitting parameter. Like in Experiment 1, this parameter estimate was used for all further analyses.

2.6.2 Results

Analysis of Variance

A numerical inspection of the data showed that mean false alarm rates ("old" responses to new items) differed between delay conditions, increasing from $M = .16$ ($SD = .11$) in the 3 min condition to $M = .26$ ($SD = .13$) after 1 day, $M = .28$ ($SD = .11$) after 3 days, and $M = .33$ ($SD = .11$) after 7 days. According to a univariate ANOVA, this increase was significant, $F(3, 116) = 11.84$, $MSE = .01$, $p < .001$, $\eta_p^2 = 0.23$, indicating that the response criterion changed across delay conditions. Following Wixted and Ebbesen (1991), the raw hit rates were therefore corrected for each cue condition by dividing them by the sum of hit and false alarm rates. Figure 5 displays the corrected hits. A 2 × 4 mixed-factors ANOVA for these corrected hits with the within-subject factor of CUE (forget, remember) and the between-subjects factor of DELAY (3 min, 1 day, 3 days, 7 days) showed a main effect of CUE, $F(1, 116) = 211.86$, $MSE < .01$, $p < .001$, $\eta_p^2 = 0.65$, reflecting typical IMDF, and a main effect of DELAY, $F(3, 116) = 21.84$, $MSE = .03$, $p < .001$, $\eta_p^2 = 0.36$, indicating time-dependent forgetting. There was no significant interaction between the two factors, $F(3, 116) = 1.99$, $MSE < .01$, $p = .119$, $\eta_p^2 = 0.05$, indicating no change in the size of the DF-effect at longer delays. Following the analysis for Experiment 1, follow-up paired t-tests were conducted for all four delay intervals to investigate the difference between TBR and TBF items. The DF-effect was significant at all delay intervals, all $ts(29) > 7.14$, $ps < .001$, $ds \geq 1.30$, with only slight variations in effect size d.

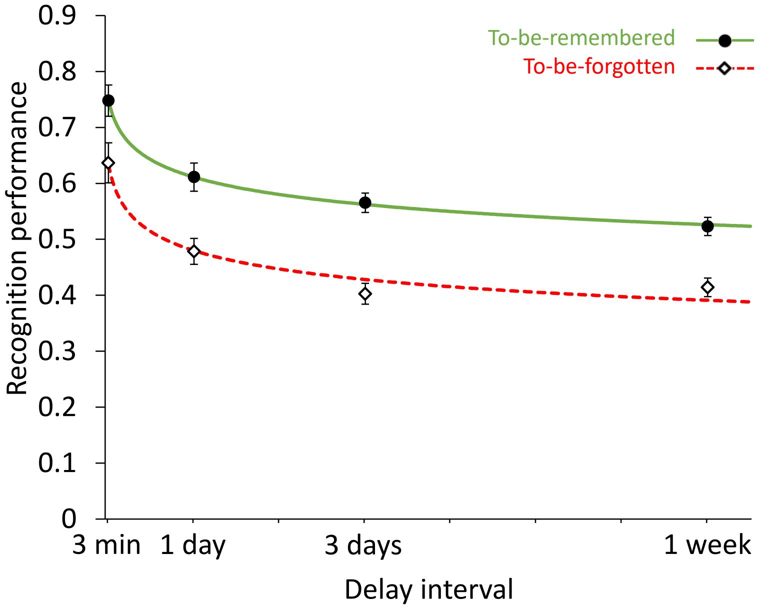

Figure 5: *Results of Experiment 2.* Mean corrected hits for both cue conditions are displayed along with the best-fitting power functions. Recognition of both the to-be-remembered items (TBR: filled circles, solid green) and the to-be-forgotten items (TBF: open diamonds, dashed red) showed time-dependent forgetting, described by a power function of time. The TBF items showed a higher relative rate of forgetting, reflected in a larger forgetting rate parameter, than the TBR items (see Table 2). Error bars represent ± 1 standard error.

Power Function Analysis

Again, the power function described the corrected hit rates of both conditions well (Figure 5), as is reflected in $\chi^2(2)$ values of ≤ 2.97 for the single cue conditions (Table 2). The power function again explained most of the variance in the data, as is represented in r^2 values of $\geq .974$ in both conditions. Estimates for parameter a were again found to differ significantly between conditions, $\chi^2(1) = 19.41$, showing an effect of directed forgetting. At the same time, the cue conditions also differed in parameter b, $\chi^2(1) = 4.13$, reflecting a higher forgetting rate for TBF compared to TBR information.

Reanalysis for d’

Another measure to examine recognition data is the sensitivity index d’ (see e.g. Snodgrass & Corwin, 1988), which was calculated as an alternative to the corrected hits reported above. For the calculation of d’, the correction suggested by Macmillan and Creelman (2005) was used for hit rates of 1 and false alarm rates of 0. A 2 × 4 mixed-factors ANOVA for d’ with the within-subject factor of CUE (forget, remember) and the

Table 2

Best Power Function Fits and Explained Variance for Experiment 2

Condition	**a**	**b**	**$\chi^2(2)$**	**r^2**
To-be-Remembered	0.750	0.080	0.05	.974
To-be-Forgotten	0.635	0.109	2.97	.975

between-subjects factor of DELAY (3 min, 1 day, 3 days, 7 days) showed a main effect of CUE, $F(1, 116) = 332.21$, $MSE = .15$, $p < .001$, $\eta_p^2 = 0.74$, reflecting typical IMDF, and a main effect of DELAY, $F(3, 116) = 29.86$, $MSE = .39$, $p < .001$, $\eta_p^2 = 0.44$, indicating time-dependent forgetting. This time, there was also a significant interaction between the two factors, $F(3, 116) = 5.24$, $MSE = .15$, $p = .002$, $\eta_p^2 = 0.12$, indicating a change in the size of the DF-effect at longer delays (see Figure 6). In line with this finding, follow-up paired t-tests between TBR and TBF items demonstrated significant IMDF at all four delay intervals, all $ts(29) > 7.73$, $ps < .001$, $ds \geq 1.41$, with effect size d decreasing with increasing delay (from $d = 1.98$ after 3 min to $d = 1.41$ after one week).

As d’ does not follow a binomial distribution, it was not possible to perform the same maximum likelihood methods as described above for the corrected hits. Instead, and following previous work on time-dependent forgetting (e.g. R. B. Anderson & Tweney, 1997; Wixted & Ebbesen, 1991), non-linear least squares were used to estimate power function parameters, using explained variance as a descriptive measure of goodness-of-fit. To obtain a common parameter c for both cue conditions, the three-parametric version of the power function was fit separately to both conditions in a first step. In a next step, the analysis was repeated with the mean of both separate c estimations as a common parameter c. The numerical pattern of results of this analysis was very similar to that found with maximum likelihood methods, even though parameter estimates for d’ naturally differed numerically from those for corrected hits (see Table 3). Again, the rate of time-dependent forgetting was higher for TBF information, indicating that the pattern of findings reported above does not depend on exactly which method is used to correct hit rates in item recognition.

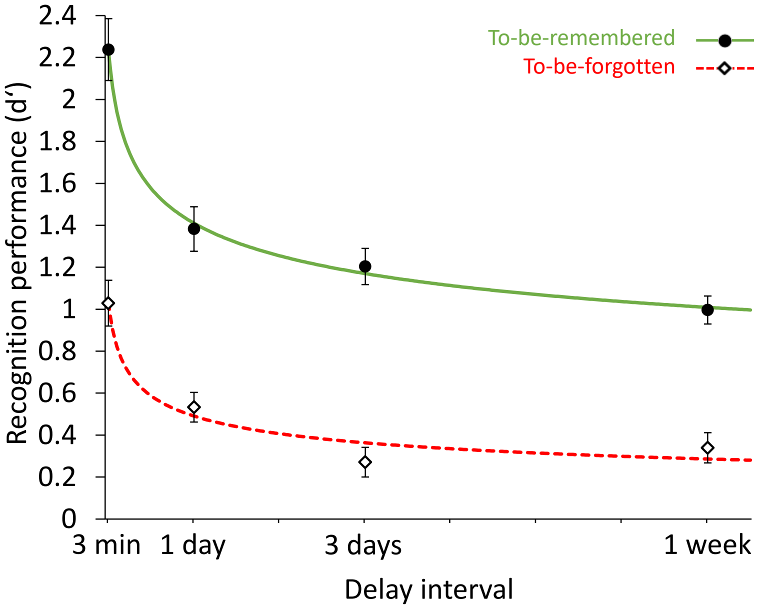

Figure 6: *Results of Experiment 2 using d'.* For both cue conditions, d' is displayed along with the best-fitting power functions derived from non-linear least squares. Recognition of both the to-be-remembered items (TBR: filled circles, solid green) and the to-be-forgotten items (TBF: open diamonds, dashed red) showed time-dependent forgetting, described by a power function of time. The TBF items showed a higher relative rate of forgetting, reflected in a larger forgetting rate parameter, than the TBR items (see Table 3). Error bars represent ± 1 standard error.

Table 3

Non-Linear Least Squares Parameter Estimates and Explained Variance for Experiment 2 (d')

Condition	**a**	**b**	**c**	**r^2**
To-be-Remembered (d')	2.245	0.178	12.642	.998
To-be-Forgotten (d')	1.039	0.287		.963

2.6.3 Discussion

Using item recognition rather than free recall testing, the results of Experiment 2 replicate those of Experiment 1. Experiment 2 again produced an IMDF effect for all four delay intervals using ANOVA, which is consistent with prior work using item recognition by MacLeod (1975, 1989), who also found IMDF to be lasting. In a power function analysis analogous to that for Experiment 1, the corrected hit rates again showed typical time-dependent forgetting for both the TBR and the TBF items, with the memory rates of both item types well described by a power function of time. Critically, and like in Experiment 1, the TBR and TBF items differed in time-dependent forgetting, with the TBF items showing a larger relative rate of forgetting (parameter b) than the TBR items.

In one regard the results from Experiments 1 and 2 differed: In Experiment 1, a significant interaction emerged between the factors cue and delay, but the corresponding interaction was not significant in Experiment 2 for the corrected hits. The ANOVA analysis accordingly implied a change in the size of the DF-effect only in Experiment 1, while for both experiments, the power function analysis indicated differences in the rate of forgetting between TBR and TBF information. As ANOVAs and power function analysis focus on different measurements of forgetting (absolute vs. relative forgetting, respectively), the two methods of analysis do not always coincide in their implication regarding differences in forgetting (e.g. Carpenter et al., 2008). Even if they do agree superficially – as was the case for Experiment 1, where both lines of analysis implied differences in time-dependent forgetting for TBR vs. TBF items – they do not necessarily match in the details: According to the ANOVA results for this experiment, TBF items show a slower decline across the retention intervals than the TBR items, as the latter show a larger absolute drop in memory performance, compared to the lower relative rate of forgetting for TBR information derived from power function analysis.

As outlined above, the results in Experiment 1 might have been influenced by a floor effect masking the actual drop in performance (both absolute and relative) for TBF information. The absence of a significant interaction in Experiment 2 might therefore be a reflection of the absence of a floor effect, but it is also possible that the change in recall format (from free recall to item recognition) influenced this difference in results. However, a re-analysis of the results for Experiment 2 with d' instead of the corrected hits yielded the same pattern of ANOVA results as Experiment 1, with a significant interaction between the factors cue and delay, and a decrease in effect size d for longer retention intervals in follow-up paired t-tests. It is possible therefore, that the absence of a significant interaction for the corrected hits in Experiment 2 was not due to the use of item recognition, but that instead the corrected hits did not adequately capture the actual time course of time-dependent forgetting in Experiment 2. This raises the question which measure of item recognition should be used for power function analysis in the case of changing response criteria across delay intervals (e.g. the current

Experiment 2). While for the present data d' might have captured the time-dependent forgetting more accurately for TBR and TBF information, d' does not allow for an analysis using maximum likelihood methods because the latter requires a binomially distributed measure of performance. Even though the same numerical pattern as for the corrected hits (and for the recall rates of Experiment 1) arose with a non-linear least squares analysis for d', in that forgetting rate parameter b was smaller for TBR compared to TBF items, it was not possible to statistically test this difference in forgetting rate.

Taken together, the results for both Experiments 1 and 2 agree on a lower relative rate of forgetting for TBR compared to TBF information in IMDF, and this finding does not seem to be dependent on the choice of recall format, even though it is unclear what the preferable approach for the analysis of recognition data might be. In a next step, it was to be investigated in Experiment 3 whether the central finding of increased forgetting for TBF information generalizes from IMDF to LMDF, as in both methods participants are informed that TBF information will be irrelevant for the final test (see e.g. Cowan et al., 2021). However, as mechanisms that explain item-method vs. list-method directed forgetting are assumed to differ, especially in their involvement of processes acting at encoding (e.g. Rummel et al., 2016), also a different pattern of results might emerge for LMDF.

2.7 Experiment 3: List-method Directed Forgetting (Free Recall)

2.7.1 Method

Participants

300 new participants ($M = 22.22$ years, range = 18-30 years, 185 female) were recruited from the same subject pool as in Experiments 1 and 2. 87.3 % of participants were currently enrolled at university, all other participants reported to be employed or doing vocational training. Participants were distributed equally across conditions, resulting in $n = 30$ for each condition. Sample size followed Experiments 1 and 2.

Materials

32 new concrete, unrelated nouns (4-6 characters) were drawn from the CELEX database, using Wordgen v1.0 Software Toolbox (Duyck et al., 2004). These words were then split into two lists of 16 words each that served as List 1 and List 2.

Design

The experiment had a 2 (CUE: forget vs. remember) × 5 (DELAY: 3 min vs. 1 day vs. 2 days vs. 3 days vs. 7 days) between-subjects design. Within each delay condition, each list served equally often as List 1 and List 2 and was assigned equally often to the forget and remember conditions.

Procedure

Again, data collection took place via zoom meetings. Before the beginning of the study phase, subjects were informed that they would be presented with two lists of words for study. It was added that they would only be told after the presentation of each list whether that list would be relevant for an upcoming memory test at the end of the experiment and that forgetting one list might help them remember other materials more efficiently. It was stressed that it would be best to memorize each list well, as those further instructions would only be given after each list. During study, words were displayed in random order for 4 s each. After study of the first list, participants in the remember condition were asked to remember the list for the upcoming test. Participants in the forget condition were told that this list could be discarded from memory as it would not be tested later. This instruction was followed by the presentation of List 2. All subjects were asked to remember this second list for the upcoming test.

Between study and test, exactly the same procedures were employed as in Experiments 1 and 2: Again, all subjects in the long delay conditions counted backwards for 2 min following the study phase and returned 1, 2, 3 or 7 days later for their second session. Immediately before the test, all participants (including those in the 3 min condition) solved Raven's Matrices (Raven et al., 2000) for 3 min. At test, subjects were asked to remember all words from the study phase, regardless of list membership and cue condition, following previous work on LMDF (e.g. Golding & Gottlob, 2005; Wessel & Merckelbach, 2006; Zellner & Bäuml, 2006). Subjects were then requested to type every item they could remember from the two lists into the zoom chat, in any order they liked. They were given five minutes to complete the task. This procedure was chosen because the standard procedure of asking subjects to first recall words only from List 1 (e.g. Abel & Bäuml, 2017; Hupbach, 2018; Sahakyan & Kelley, 2002) could have reduced or even eliminated possible List 2 benefits (see Pastötter et al., 2012; but see Zellner & Bäuml, 2006), whereas the present test procedure has been shown to produce significant List 2 benefits (e.g. Zellner & Bäuml, 2006).

Fitting Procedures

The fitting procedure was similar to the ones employed in Experiments 1 and 2. Again, before running more detailed analyses, a common scaling parameter c was estimated, but this time for both cue conditions from both lists at the same time (four conditions total), using the same basic procedure as in Experiment 1. Like in Experiments 1 and 2, the restricted power function model with a common scaling parameter c for all four conditions described the recall rates equally as well as the more general power function model in which the parameter varied freely across all four conditions, $\chi^2(3) = 1.97$, with $c = 11.06$ as the best fitting parameter. Like in Experiments 1 and 2, this parameter estimate was used for all further analyses.

2.7.2 Results

Analysis of Variance

Figure 7 shows mean recall rates at all five delay intervals for List 1 (A) and List 2 (B). A univariate 2 × 5 ANOVA for final recall of List 1 with the between-subjects factors of CUE (forget, remember) and DELAY (3 min, 1 day, 2 days, 3 days, 7 days) found significant main effects of CUE, $F(1, 290) = 53.23$, $MSE = .03$, $p < .001$, $\eta_p^2 = 0.16$, indicating directed forgetting, as well as of DELAY, $F(4, 290) = 9.70$, $MSE = .03$, $p < .001$, $\eta_p^2 = 0.12$, showing time-dependent forgetting. There was no significant interaction between the two factors, $F(4, 290) = 0.57$, $MSE = .03$, $p = .684$, $\eta_p^2 < 0.01$. Another univariate 2 × 5 ANOVA for final recall of List 2 with the between-subjects factors of CUE and DELAY also found significant main effects of CUE, $F(1, 290) = 7.95$, $MSE = .04$, $p = .005$, $\eta_p^2 = 0.03$, suggesting List 2 enhancement for forget-cued participants, and DELAY, $F(4, 290) = 7.99$, $MSE = .04$, $p < .001$, $\eta_p^2 = 0.10$, but no significant interaction between the two factors, $F(4, 290) = 0.19$, $MSE = .04$, $p = .941$, $\eta_p^2 < 0.01$.

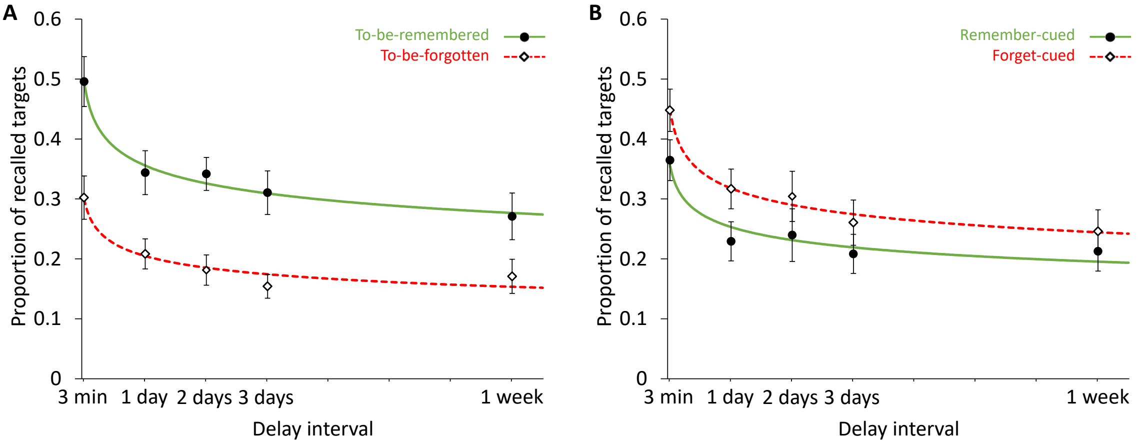

Figure 7: *Results of Experiment 3.* Mean results for first-list recall (A) and second-list recall (B), along with the best-fitting power functions are displayed. **(A)** performance for to-be-remembered items (filled circles, solid green) was higher than for to-be-forgotten items (open diamonds, dashed red). **(B)** performance for forget-cued participants (open diamonds, dashed red) condition was higher than for remember-cued participants (filled circles, solid green). In all four cue × list conditions, time-dependent forgetting emerged, described by a power function of time, with similar rates of time-dependent forgetting in all conditions (see Table 4). Error bars represent ± 1 standard error.

Power Function Analysis

As for Experiments 1 and 2, and for both List 1 and List 2, the power function described the recall rates of both cue conditions well (Figure 7), as is reflected in $\chi^2(3)$ values of ≤ 3.03 for the all four conditions (Table 4). The power function again explained most of the variance in the data, as is represented in r^2 values of $\geq .933$ for all conditions. For List 1, estimates for parameter a were again found to differ significantly between both cue conditions, $\chi^2(1) = 41.17$, showing an effect of directed forgetting. Contrasting with Experiments 1 and 2, the cue conditions did not differ in parameter b, $\chi^2(1) = 0.44$, reflecting similar forgetting rates for both TBR and TBF information. For List 2, estimates for parameter a also differed significantly between the cue conditions, $\chi^2(1) = 9.10$, reflecting List 2 enhancement for forget-cued participants. As for List 1, the cue conditions did not differ in parameter b, $\chi^2(1) < 0.01$, reflecting similar forgetting rates after both cue instructions for List 2 information. Another model that allowed parameter a to vary freely across all four Cue × List conditions while parameter b was fixed for all conditions, was compared to a model in which both parameters were allowed to vary freely. According to this comparison, parameter b did not differ significantly across the four conditions, $\chi^2(3) = 0.45$.

Table 4

Best Power Function Fits and Explained Variance for Experiment 3

Condition	a	b	$\chi^2(3)$	r^2
List 1 recall				
To-be-Remembered	0.497	0.134	0.92	.985
To-be-Forgotten	0.300	0.154	2.59	.946
List 2 recall				
Remember-Cued	0.357	0.138	3.03	.933
Forget-Cued	0.450	0.140	0.96	.984

Additional Analyses

Following previous work (e.g. Pastötter & Bäuml, 2010; Pastötter et al., 2012; Sheard & MacLeod, 2005), the recall performance of the individual List 1 and List 2 items depending on their input order during initial study was examined for all subjects. Due to experimenter error, the input order of items during study was missing for four participants (one person in the Remember 3 days condition, and one person each in the Forget 3 min, 1 day, and 7 days conditions). This serial position analysis was therefore conducted for $n = 296$ subjects. Pooled across all five delay intervals, for List 1, recall at all list positions was numerically higher for TBR information compared to TBF information, but the size of this benefit was particularly pronounced for the first half of List 1 (Figure 8A). For List 2, a numerical recall benefit emerged for forget-cued participants that was restricted to the first four list positions (Figure 8B).

To investigate these effects, a 2 (CUE: remember vs. forget) × 5 (DELAY: 3 min vs. 1 day vs. 2 days vs. 3 days vs. 7 days) × 16 (SERIAL POSITION: 1-16) mixed-factors ANOVA was conducted for each list. CUE and DELAY were manipulated between-subjects, while SERIAL POSITION varied within subjects. For List 1, the ANOVA produced significant main effects for CUE, $F(1, 286) = 7.25$, $MSE = .63$, $p = .008$, $\eta_p^2 = 0.03$, reflecting a recall benefit for remember-cued participants, DELAY, $F(1, 286) = 8.01$, $MSE = .63$, $p < .001$, $\eta_p^2 = 0.10$, indicating time-dependent forgetting, as well as for SERIAL POSITION, $F(13.05, 3732.81) = 30.53$, $MSE = .16$, $p < .001$, $\eta_p^2 = 0.10$,

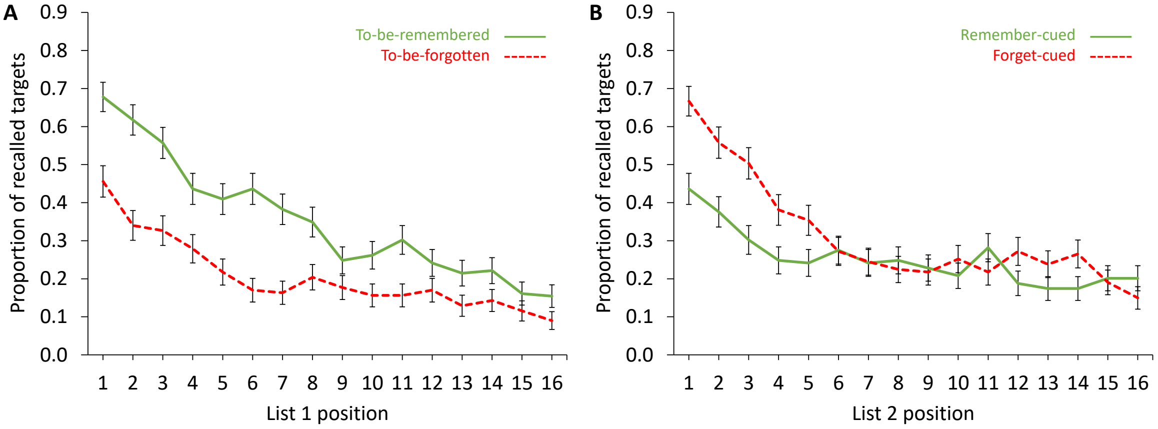

Figure 8: *Serial Position Analysis for Experiment 3.* Results for first-list recall (A) and second-list recall (B). Displayed is the mean proportion of recalled targets depending on their serial input order during study (list position, 1-16) for both remember-cued (solid green) and forget-cued (dashed red) conditions, pooled across all five delay intervals. Error bars represent ± 1 standard error.

reflecting a primacy effect for early items. Additionally, the ANOVA showed significant interactions of CUE and SERIAL POSITION, $F(13.05, 3832.81) = 2.84$, $MSE = .16$, $p < .001$, $\eta_p^2 = 0.01$, reflecting a bigger primacy effect for remember-cued participants, as well as of DELAY and SERIAL POSITION, $F(52.21, 3732.81) = 1.67$, $MSE = .16$, $p = .001$, $\eta_p^2 = 0.02$, indicating that the size of the primacy effect changed across delay conditions.[1] The three-way interaction did not reach significance, $p = .653$.

For List 2, main effects emerged for CUE, $F(1, 286) = 7.25$, $MSE = .63$, $p = .008$, $\eta_p^2 = 0.03$, reflecting a recall benefit for forget-cued participants, DELAY, $F(1, 286) = 8.01$, $MSE = .63$, $p < .001$, $\eta_p^2 = 0.10$, indicating time-dependent forgetting, as well as for SERIAL POSITION, $F(13.36, 3821.09) = 20.49$, $MSE = .16$, $p < .001$, $\eta_p^2 = 0.07$, reflecting a primacy effect for early items. Additionally, the ANOVA showed a significant interaction of CUE and SERIAL POSITION, $F(13.36, 3821.09) = 3.93$, $MSE = .16$, $p < .001$, $\eta_p^2 = 0.01$, reflecting a bigger primacy effect for forget-cued participants. No other interaction reached significance, all $p \geq .625$.

To further examine these primacy effects, paired t-tests were conducted for both cue conditions and for both lists, contrasting mean recall of the first four list items (Bin 1) and of the last four list items (Bin 4). Significant primacy effects emerged for all comparisons, all $t \geq 5.41$, all $p < .001$. An examination of effect size d showed that for List 1, remember-cued participants demonstrated a bigger primacy effect than

---

[1] Appendix B shows the descriptive results of the serial position analysis of both lists for both cue conditions individually for all five delays.

forget-cued participants (TBR: 1.11; vs. TBF: 0.64), while for List 2, the opposite pattern emerged (TBR: 0.44; vs. TBF: 1.00).

As subjects were instructed to recall items from both lists at once, self-chosen output order during the test was also examined. For each subject, list membership of the first recalled word was recorded. Overall, most subjects started recall with List 1, though fewer subjects in the forget-cued conditions did so than in the remember-cued conditions (see Table 5). Those subjects who recalled items from both lists ($n = 251$) recalled on average a proportion of .69 ($SD = .28$) of the list they started recall with before they first switched to the other list during their recall. This proportion did not vary systematically across delay and cue conditions, nor depending on which list subjects started recall with. A table depicting these results can be found in Appendix B (Table B1).

Table 5

Proportion of Sample Who Started Recall With List 1 for Experiment 3, Depending on Cue- and Delay-Condition

Cue-Condition	3 min	1 day	2 days	3 days	7 days
Remember-Cued	83.3	80.0	90.0	89.7	69.0
Forget-Cued	72.4	55.2	33.3	50.0	50.0

To examine the effects of self-chosen output order during recall, mean recall rates were calculated separately for subjects who started recall with List 1 vs. List 2 . As most subjects in remember-cued conditions started recall with List 1 but only half the sample in forget-cued conditions did so, this separation was restricted to forget-cued participants. Due to varying cell sizes across delay intervals, the data was pooled across delay intervals. Starting recall with List 2 resulted overall in lower recall rates for List 1 and higher recall rates for List 2, compared to starting recall with List 1 (see Figure 9). The influence of the forget-cue was therefore overall greater for those subjects who started recall with List 2. Nonetheless, both groups of forget-cued participants differed numerically from remember-cued participants.

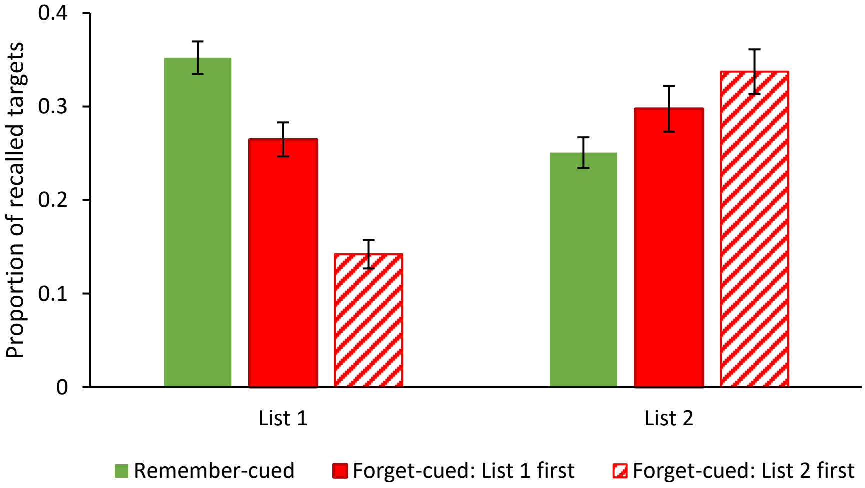

Figure 9: *Influence of Self-chosen Output Order During Recall on List 1 and List 2 Effects for Experiment 3.* Displayed is the mean recall of List 1 and List 2 items, pooled across all five delay intervals. Data for remember-cued participants is displayed in green, data for forget-cued participants is split into those who started recall with List 1 items ($n = 76$, filled red bars) and those who started recall with List 2 items ($n = 73$, striped red bars). Error bars represent ± 1 standard error.

2.7.3 Discussion

Regarding List 1, the results show persistent LMDF across all four long delay intervals (1 day, 2, 3, 7 days) with the size of the LMDF effect being largely unaffected by delay interval. The finding of a persistent LMDF effect with prolonged delay is in line with the few studies that examined the issue previously (Abel & Bäuml, 2017, 2019; Hupbach, 2018; Scully & Hupbach, 2020) and extends their findings to even longer delays and a more fine-graded analysis of the role of delay interval. The findings are however at odds with the previous observations of reduced or eliminated DF effects at longer delays (Liu, 2001; Shapiro et al., 2006). Regarding List 2, the results were analogous to those for List 1, again with a persistent effect of the forget cue on recall performance, in that the recall enhancement in response to the forget cue was comparable in size across delay intervals. However, this difference in recall between forget and remember cued participants was smaller for List 2 recall, which was also visible in a smaller effect size of the factor cue for List 2, $\eta_p^2 = 0.03$, than for List 1, $\eta_p^2 = 0.16$.

As for Experiments 1 and 2, power function analysis of the recall rates of (List 1) TBR and TBF items revealed typical time-dependent forgetting for both item types, as both were well described by the power function of time. Importantly and in contrast with the results for IMDF, forgetting rates did not differ significantly between the two types of items, suggesting similar time-dependent forgetting for TBR and TBF information. Recall rates for List 2 items also followed the power function of time and forgetting rates were also similar between the forget and remember conditions. Moreover, forgetting rates were not only similar between cue condition within lists but also across lists.

Following previous LMDF work (Pastötter & Bäuml, 2010; Pastötter et al., 2012; Sheard & MacLeod, 2005), an additional serial position analysis was conducted. The results demonstrated significant primacy effects for both lists that varied in size for the two cue conditions. For List 1, remember-cued participants showed higher numerical recall for all list positions, with a bigger primacy effect in comparison to forget-cued participants. For List 2, conversely, forget-cued participants showed a bigger primacy effect. In this case, this facilitating effect was numerically restricted to the first few list positions. For both lists, significant interactions of list position and cue emerged, in line with the results of Experiment 1 of Pastötter and Bäuml (2010). At the same time, several other studies found such an interaction only for List 2, but not for List 1 (e.g. Pastötter & Bäuml, 2010, Exp. 2 & 3; Pastötter et al., 2012; see also Lehman & Malmberg, 2009). In these studies, remember-cued participants demonstrated an even recall benefit for all list positions.

A closer look at the present results showed that solely for List 1, a significant interaction of delay and list position emerged, indicating that the size of the primacy effect changed across delay intervals. As can be seen from Figures B1 and B2 (see Appendix B), the overall pattern of serial position effects did not change much across the five delay conditions for List 2, but did so for List 1: Here, a primacy benefit of remember- over forget-cued participants emerged for some of the longer delay conditions. In the shortest delay condition (3 min) in contrast, the numerical benefit of remember-cued participants seemed to be evenly distributed across list positions. As previous studies that used serial position analyses typically used only one short delay condition (e.g. Lehman & Malmberg, 2009; Pastötter & Bäuml, 2010; Pastötter et al., 2012), this result should be viewed as preliminary. It could however indicate different processing and resulting differences in forgetting of early vs. late List 1 items across long delays in remember-cued participants. Tentative support for these findings comes from some animal studies (Bolhuis & van Kampen, 1988; Reed, 1996) that reported an influence of the length of the retention interval on the size of observable primacy effects (see Wright, 1994, for a comparison of primacy effects in animal vs. human memory). Wright et al. (1985) demonstrated a similar influence of retention interval on primacy effects both in humans and animals. The retention intervals used in these studies, however, are mostly much shorter than the ones used in the present experiment. Whether such retrieval dynamics change across delays should be investigated in future research.

Some evidence exists that List 1 forgetting and List 2 facilitation are differently affected by list output order during recall (Pastötter et al., 2012), while other studies show no influence of output order on DF (Zellner & Bäuml, 2006). According to a meta-analysis conducted by Pastötter et al. (2012), output order does not affect the magnitude of List 1 forgetting, but does influence the size of List 2 facilitation, with the latter being more pronounced when participants recall List 2 first. When no explicit recall order is used, as in the present experiment, especially forget-cued participants might start recall with the more recent List 2, and, due to output interference at test, might therefore show reduced recall of List 1 (Golding & Gottlob, 2005; Sahakyan et al., 2013). The influence of recency on self-chosen recall order might be expected to be smaller for longer delays, when the length of the retention interval is very similar for both List 1 and List 2. An examination of self-chosen recall order – implemented as list membership of the first recalled word – in the present experiment paints a different picture, however: At the shortest delay (3 min), a majority of subjects started recall with items from List 1 (remember-cued: 83.3 %; forget-cued: 72.4 %). At the longer delays (1 day, 2 days, 3 days, 7 days), conversely, the majority of remember-cued participants still started recall with List 1 items (69.0 % - 90.0 %), but a smaller proportion of forget-cued participants did so (33.3 % - 55.2 %). Accordingly, more forget-cued participants started recall with the to-be-remembered List 2 at longer delays in the present sample. This might be due to the relatively higher strength of List 2 in this condition compared with the to-be-forgotten List 1. Most participants switched between lists during recall, but most recalled a majority (69 %) of the list they started recall with before they first switched to the other list, indicating that list membership of the first recalled word described overall list output order well.

To examine whether self-chosen list output order influenced recall of List 1 and List 2 for forget-cued participants, mean recall rates were calculated separately for those participants who started recall with List 1 items and for those who started recall with List 2 items. Recalling List 2 first could result in output interference and thereby lower recall of List 1. Indeed, forget-cued subjects who started recall with List 1 recalled more List 1 items than did subjects who started recall with List 2. At the same time, forget-cued participants who started recall with List 1 recalled less List 2 items than those who started recall with List 2. As a majority of the sample (regardless of cue condition) started recall with List 1 at the 3 min delay, but less forget-cued than remember-cued participants did so at the longer delays, the different results for forget-cued participants depending on their recall order may have influenced the overall results at longer delays, increasing both the size of observed List 1 forgetting and of List 2 facilitation at the longer delays compared to designs where all subjects start recall with List 1 (e.g. Hupbach, 2018; Mulji & Bodner, 2010; Pastötter & Bäuml, 2007). This could in turn have increased the observed rate of relative forgetting for forget-cued participants in List 1 and decreased the observed rate of relative forgetting of forget-cued participants in List 2. Accordingly, the present results for forget-cued participants might contain an overestimation of List 1 forgetting and an underestimation of List 2 forgetting. Overall, the power function analysis found no significant differences in relative rates of forgetting for remember- and forget-cued participants, neither for List 1 nor for List 2. Crucially, it seems unlikely at this point that the key difference in results between Experiments 1 and 2 (IMDF) and the present LMDF experiment – increased forgetting for TBF information in IMDF, but no differences between forgetting of TBR and TBF information in LMDF – is solely due the choice of recall format in the LMDF experiment: An overestimation of List 1 forgetting for forget-cued participants would have made it easier to detect a difference between TBR and TBF information, but only a small numerical trend in this direction could be observed. Instead, it is even possible that forget-cued participants would show reduced relative forgetting in a design where all participants start recall with List 1. Future research should therefore examine the influence of list output order on both List 1 forgetting and List 2 facilitation and their long-term behavior more closely, especially as differences could emerge between self-chosen and experimentally controlled output order.

2.8 Additional Analyses for Experiments 1-3

All analyses for Experiments 1-3 up to this point have been carried out for memory scores arithmetically averaged across participants. However, it has been pointed out that doing so can produce averaging artifacts, for instance, a group function with mathematical properties that are not representative of the individual participant data (Estes, 1956; Sidman, 1952). As a way to evaluate whether or not averaging artifacts contribute to results, a reanalysis using geometric averaging instead has been proposed (e.g. R. B. Anderson & Tweney, 1997; Wixted & Ebbesen, 1997). Consistent with prior work (e.g., Wixted, 2022a; Wixted & Ebbesen, 1997), geometric averaging for Experiments 1-3 led to the same pattern of results as arithmetical averaging (see Table 6). The same fitting procedures as for arithmetically averaged data were used. Estimates for parameter *c* were numerically very similar to those derived for arithmetic averages for each experiment. For all conditions, the power function described the recall rates equally as well as the statistical baseline models. Importantly, estimates of forgetting rate parameter *b* were smaller for the TBR items than for the TBF items in both experiments using IMDF (Exp. 1 & 2), but were very similar for both TBR and TBF information in LMDF (Exp. 3). Additionally, parameter *b* was again very similar for all four conditions of Experiment 3. These findings corroborate the results presented above and indicate that the contribution of averaging artifacts, should they exist, is not solely responsible for the pattern of results found with arithmetically averaged data.

Table 6

Best Power Function Fits and Explained Variance for Geometrically Averaged Data of Experiments 1-3

Condition	**a**	**b**	**c**	**df**	**$\chi^2(2)$**	**r^2**
Experiment 1						
To-be-Remembered	0.728	0.228	6.86	2	1.37	.991
To-be-Forgotten	0.157	0.371			1.51	.976
Experiment 2						
To-be-Remembered	0.744	0.071	19.20	2	0.16	.995
To-be-Forgotten	0.630	0.097			2.65	.969
Experiment 3						
List 1, To-be-Remembered	0.481	0.137	11.28	3	1.54	.976
List 1, To-be-Forgotten	0.287	0.152			2.04	.954
List 2, Remember-Cued	0.345	0.147			2.71	.944
List 2, Forget-Cued	0.440	0.149			0.84	.987

2.9 Discussion of Experiments 1-3

The results of Experiments 1-3 show that, over the course of one week, both IMDF and LMDF effects are lasting. For IMDF, this holds both for free recall and item recognition, and the results of all three experiments thus replicate and extend prior work (e.g., Abel & Bäuml, 2017; Hupbach, 2018; MacLeod, 1975, 1989; Scullin et al., 2017). The power analysis produced additional findings of interest. First, and for both methods of DF, both TBR and TBF information show typical time-dependent forgetting with memory performance declining rapidly soon after study followed by a long, much slower decline in memory performance. For LMDF, this pattern also extended to memory performance of remember- and forget-cued participants for List 2. Importantly, for all types of information, this decline was well described by a power function of time. Second, both when using free recall and when using item recognition testing, forgetting rates differed between TBR and TBF information in IMDF, with a higher relative rate of forgetting for the TBF information. At the same time, both TBR and TBF information showed very similar rates of forgetting in LMDF, and similar rates of forgetting also for the two cue conditions regarding List 2 recall, as no significant differences in forgetting arose between conditions. Third, geometrically instead of arithmetically averaging the data of all three experiments led to a very similar pattern of results, alleviating the concerns that averaging artifacts might have contributed to the findings.

This contrasting finding regarding the forgetting rates of TBR and TBF information is generally in line with the assumption that different mechanisms contribute to DF effects using the item- vs. the list-method. Overall, processes working at or around encoding probably play a bigger role in IMDF, while in LMDF contributing mechanisms need to work at later processing stages (Bäuml, 2008; Rummel et al., 2016). The finding that DF effects in both methods seem to be long-lasting is consistent with most theoretical accounts of the two methods, apart from the context change account of LMDF that assumes DF effects to diminish for longer retention intervals (e.g. Abel & Bäuml, 2017; Divis & Benjamin, 2014; Sahakyan & Kelley, 2002). The current set of experiments was not designed to distinguish between existing accounts, but the present results of the power function analysis nonetheless provide new observations successful accounts should be able to incorporate. Combining the selective rehearsal accounts of IMDF and LMDF (B. H. Basden et al., 1993; R. A. Bjork, 1970) and the observation of reduced relative forgetting for strong vs. weak items (Wixted, 2022a), the current findings would be compatible with a selective rehearsal explanation of IMDF but not of LMDF. Assuming a contribution of inhibitory processes to DF effects (Fawcett & Taylor, 2008; Geiselman et al., 1983), the rate of forgetting for TBF information would need to be similar to that of TBR information or even smaller, if inhibition is transient in nature (e.g. R. A. Bjork; 1989; Storm et al., 2012). Such an assumption would be at odds with the results for IMDF, but could be aligned with the results for LMDF. If, however, inhibitory effects in IMDF are assumed to be long-lasting, the present results could also be compatible with such a variant of the inhibition account. Finally, for a mixed account of IMDF that assumes both selective rehearsal and inhibition to be contributing factors (e.g. Fellner et al., 2020), the present results would be in line with variants that assume selective rehearsal to play a bigger role than inhibition – if inhibition is assumed to be transient –, but also with variants that assume both mechanisms to contribute over longer delays.

There are several possible explanations why relative time-dependent forgetting would differ between TBR and TBF information in IMDF. One possibility is that forget- and remember-cues might serve as an indication of how future-relevant studied materials are and through this distinction influence further processing differently. This idea is in line with previous findings of selective sleep benefits for information that is cued as relevant before sleep, for example, by manipulating test expectancy for studied materials. Wilhelm et al. (2011) had their participants study a set of materials and then introduced retention intervals of 9 hours before a test that participants either spent awake or asleep. The authors informed some subjects before the delay that there would be a memory test at the end of the retention interval to examine how this information would influence recall after the delay. The resulting memory performance was higher after sleep compared to wakefulness, but only for participants that had been told to expect the test. This finding indicates that merely expecting that a memory will be of use for a future test may determine whether or not sleep benefits consolidation of this memory (see also van Dongen et al., 2012). However, such an explanation should also extend to TBR and TBF information in LMDF (see Cowan et al., 2021), and the results of Experiment 3 showed no measurable difference in forgetting for the two kinds of information. Also, the finding of Wilhelm et al. (2011) does not seem to be particularly robust, as several studies since have failed to replicate the described influence of test expectancy on sleep benefits (Ashton & Cairney, 2021; Reverberi et al., 2020; Wamsley et al., 2016).

It is also possible that the difference in forgetting for TBR and TBF materials in

IMDF is purely due to differences in strength. Wixted (2022a) had observed a difference in forgetting rates depending on degree of learning and speculated that degree of learning might serve as an indicator of how subjectively meaningful studied material is. In this vein, material of higher meaningfulness might, at least to some degree, be forgotten less over time, for instance, by prioritized consolidation of this kind of information (Cowan et al., 2021; Stickgold & Walker, 2013). If TBR items imitated items with a high degree of learning, TBR items might also be consolidated preferentially and thus, to some degree, be forgotten less over time. Studies that examined the role of sleep-associated memory consolidation for IMDF effects are roughly consistent with such an idea. Saletin et al. (2011) introduced a 6 hrs delay between study and recall in an IMDF task and varied whether the delay contained a 100 min nap or not. The authors found a larger difference – that is, a bigger DF effect – between TBR and TBF items after the nap than in the no-nap condition. Critically, this larger effect was due to a selective benefit of the nap for the recall of TBR items, in line with the idea of overall better consolidation for TBR than for TBF information (see also Rauchs et al., 2011). In contrast, sleep that occurred shortly after encoding was found to leave the LMDF effect unaffected (Hupbach, 2018) or to even reduce the effect by increasing recall of the TBF information (Abel & Bäuml, 2013), which points to similar, or even enhanced, consolidation of the TBF information. These findings on sleep-related consolidation processes in IMDF and LMDF therefore fit at least roughly with the present finding of a discrepancy between the two methods regarding forgetting of TBR and TBF information. Assuming that differences in strength determine, at least to some degree, further processing of studied materials, this would indicate that TBR and TBF items differ in strength in IMDF, but do not, or not to a large extent, in LMDF. Such an idea would not be readily compatible with the selective rehearsal account of LMDF (R. A. Bjork, 1970), and it would also underline that a mere performance difference between conditions should not be taken as a proxy for strength difference.

It is also possible that the discrepancy in results between IMDF and LMDF in regards to time-dependent forgetting of TBR and TBF information is due to the locus of processing differences in the two methods. In this case, the results would point to a crucial role of processes acting at encoding for subsequent long-term memory effects that determine how quickly studied materials are forgotten, for example by creating differences in the underlying memory traces. Such a difference could derive from degree of learning (e.g. Wixted, 2022a), from additional active processes during encoding that impair the memory representation (e.g. Fawcett & Taylor, 2008), or from the encoding of future-relevant tags (Cowan et al., 2021). If differences in encoding were crucial for the emergence of differential rates of time-dependent forgetting, such differences might also arise for early List 2 items when comparing remember- and forget-cued participants in an LMDF paradigm (Pastötter & Bäuml, 2010; Pastötter et al. 2012). For the present Experiment 3, preliminary curve fitting was conducted for the first four List 2 items using non-linear least squares and a two-parameter version of the power function of time. For the whole sample, no numerical difference in parameter b emerged between remember- and forget-cued participants, disagreeing with the idea that differences in encoding might underlie differences in time-dependent forgetting. However, when self-chosen list output order was taken into account, a higher rate of forgetting emerged for those participants who started recall with List 1 compared with those who started recall with List 2,[2] indicating that time-dependent forgetting for early List 2 items might have been reduced at least for a subgroup of the forget-cued participants. To investigate this issue more conclusively, fully balancing list output order and a substantial increase in sample size might be necessary in future research. Such an investigation might shed more light on the temporal dynamics and output dependence of List 2 effects, and also allow for an examination of possible encoding differences contributing to variations in time-dependent forgetting.

Taken together, the results of Experiments 1-3 further support the assumption that different mechanisms are responsible for DF effects in IMDF and LMDF, and add to the body of findings that successful theoretical accounts need to be able to explain. With the finding of differential time-dependent forgetting for TBR vs. TBF information in IMDF but not in LMDF, they offer some first insights into the circumstances under which rates of forgetting might differ between conditions: The absence of a significant difference in forgetting between TBR and TBF information in LMDF could indicate that a mere expectation that some information will not be relevant at a later test is not enough to significantly affect the rate of forgetting. Instead, differences in encoding strength might be necessary for differences in forgetting in DF paradigms. However, due to the novelty of these findings, this difference between IMDF and LMDF should be treated with caution, and future work should involve replication attempts with high statistical power.

---

[2] For the whole sample, TBR and TBF information differed numerically in parameter a (0.326 vs. 0.508), but showed very similar parameters b (0.057 vs. 0.053). Within forget-cued conditions, when the sample was split into those who started recall with List 1 vs. those who started with List 2, numerical differences arose both for parameter a (0.423 vs. 0.583) and parameter b (0.090 vs. 0.012).

Chapter 3

Time-dependent Forgetting and Retrieval Practice Effects

When tasked with the goal to reduce forgetting of important information, most people probably would not focus on consciously trying to forget less important information and instead focus on remembering the important information – most likely by employing some sort of practice strategy. A relevant question is then whether targeted practice can actually slow down relative time-dependent forgetting and whether different practice strategies do so to different extents. Trying to optimize retention over long retention intervals has been a goal of memory research for at least a century, evidenced by such early studies as Abbott (1909) and Spitzer (1939), and continues to be relevant for every day life and especially educational fields, such as school and university. The question of whether specific study or practice strategies are capable of reducing forgetting is however not only of interest from a practical perspective, but also from a theoretical perspective: While the results from Experiments 1-3 could be interpreted to indicate that differences at initial encoding are necessary to produce differences in relative forgetting, it would be interesting to investigate whether processes exist that can influence the rate of forgetting also after encoding has been completed, as the future relevance of materials is not always determined already at encoding. The results of Experiment 3 seem to indicate that a forget-cue in an LMDF paradigm does not measurably alter the rate of forgetting over time, but a likely candidate for lasting effects induced after initial encoding has been completed can be found in retrieval practice, where previously studied information is retrieved before a final test.

Based on the observation that many people who try to memorize a certain set of information do so by repeating the studied information (or at least parts of it) to themselves without any external memory aids, several researchers started to investigate whether recall is indeed a "desirable and helpful factor in the learning process“ (Abbott, 1909, p. 1). Other early research goals included comparisons between retrieval practice and restudy (where previously studied information is simply presented again), examinations of recall at different stages in the learning process (Gates, 1917) and the influence of recall on long-term retention in a classroom-like environment (Spitzer, 1939), questions that are still being investigated today. These studies converged on the finding that recall of previously studied materials can improve memory in comparison to control conditions. Some of these early studies contrasted a condition that contained some sort of retrieval practice of previously studied materials with a condition where no further exposition to the studied materials took place (e.g. Myers, 1914; Spitzer, 1939; see also Chan, 2010; Slamecka & Katsaiti, 1988), a design choice that confounds practice condition with total exposure time to the studied materials. Since then, retrieval practice has most often been compared to a restudy condition (for reviews, see e.g. Roediger & Butler, 2011; Roediger & Karpicke, 2006a; Rowland, 2014). Such a study set up typically results in a performance benefit for retrieval practiced material in a final recall test compared to restudied material – sometimes called the "testing effect“ (e.g. Roediger & Butler, 2011; Roediger & Karpicke, 2006a, 2006b; Toppino & Cohen, 2009) – demonstrating that introducing a recall test does not only measure memory performance but can also increase final memory performance (e.g. Roediger & Karpicke, 2006b). In these studies, retrieval practice usually involves all studied items, or unrelated sets of items, in contrast to the paradigm of retrieval-induced forgetting (e.g. M. C. Anderson et al., 1994), where retrieval practice is restricted to a subset of related materials. Such incomplete retrieval practice has been shown to impair retention of not retrieved, related materials.

3.1 Retrieval Practice Effects: Basic Findings and Theoretical Approaches

Studies on retrieval practice effects (see Roediger & Karpicke, 2006b; Rowland, 2014) usually include an initial study phase that is typically immediately followed by a practice phase, during which studied material is either retrieved or restudied (see Figure 10). After a time interval of a few minutes to several days, a final recall test takes place that measures the performance of all studied materials, regardless of practice type. The typical recall benefit for retrieval practiced materials becomes especially apparent after longer delays, as short intervals of only some minutes often (but not always) lead to a performance benefit for restudied material (for a meta-analysis, see Rowland, 2014). Overall, the beneficial effects of retrieval practice on memory performance have long been argued to increase in size over time (e.g. Roediger & Karpicke, 2006b), which could indicate differential forgetting depending on practice schedule. Aside from direct benefits for final memory performance, several other beneficial effects have been observed (Roediger et al., 2011b). For example, retrieval practice has also been found to improve recall of subsequently studied new materials (forward testing effect, Pastötter & Bäuml, 2014; Szpunar et al., 2008) and to increase transfer effects of studied concepts to new materials (Jacoby et al., 2010; Siler & Benjamin, 2020; for a meta-analysis see Pan & Rickard, 2018).

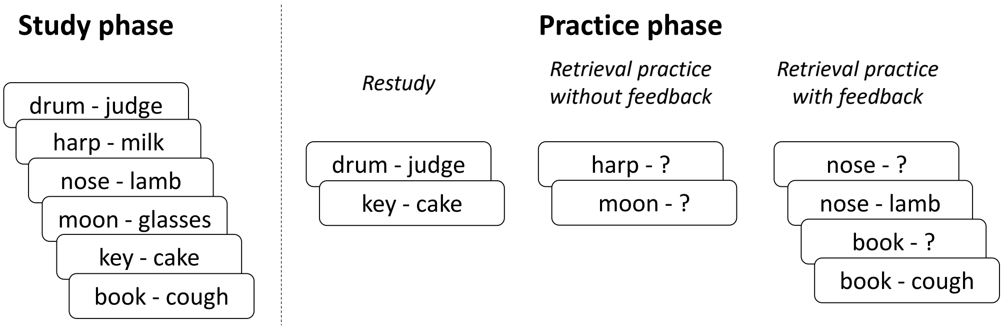

Figure 10: *Investigating Retrieval Practice Effects.* Displayed is the study phase and the practice phase of typical experiments investigating retrieval practice effects. After materials have been studied (in this case, paired associates), some items are restudied (i.e., presented again), while others are retrieval practiced (i.e., participants are asked to recall the studied target information) with or without feedback. In some studies, a subset of studied materials is not presented again after study (no-practice condition). In a subsequent test phase after a retention interval of varying length, all studied items are tested, regardless of practice condition. Especially after prolonged retention intervals, retrieval practice conditions typically outperform restudy conditions.

Final performance benefits after retrieval practice have been reported for a wide range of materials, samples and practice protocols (for an overview, see e.g. Karpicke, 2017), for example using lists of words (Carpenter & DeLosh, 2006; Wheeler et al., 2003), paired associates (Abel & Roediger, 2017), vocabulary pairs (Carpenter et al., 2008, Exp. 3; Carrier & Pashler, 1992), obscure facts (Carpenter et al., 2008, Exp. 1 & 2), prose materials (Glover, 1989; Roediger & Karpicke, 2006b) and visuo-spatial stimuli (Carpenter & Pashler, 2007), for various age groups (Glover, 1989; Meyer &

Logan, 2013; Spitzer, 1939), in classroom settings (McDaniel et al., 2007; Roediger et al., 2011a), and with different numbers of practice cycles (Roediger & Karpicke, 2006b; Toppino & Cohen, 2009; Wheeler et al., 2003). The influence of recall format, both for retrieval practice and for the final test, has also been experimentally studied. Typically, more difficult initial retrieval, such as free recall or recall with weak retrieval cues, results in bigger final recall benefits compared to control conditions (e.g. Glover, 1989; Carpenter & DeLosh, 2006; see also R. A. Bjork, 1975, 1994). At the same time, these performance benefits also depend on success during retrieval practice, with low success rates often failing to produce benefits compared to control conditions (Rowland, 2014). A benefit of retrieval practice over restudy has been observed in free recall (Roediger & Karpicke, 2006b; Carpenter & DeLosh, 2006), cued recall (Abel & Roediger, 2017; Carpenter et al., 2008; Carrier & Pashler, 1992), and item recognition (Mulligan & Peterson, 2015; Siler & Benjamin, 2020).

An important variation across studies is the presence or absence of corrective feedback during retrieval practice. Providing feedback after a retrieval attempt can be used to equate exposure to studied materials in comparison to a restudy condition (Kang et al., 2007; Karpicke, 2017). It also allows for a correction of errors and for a strengthening of correct answers (Pashler et al., 2005; Roediger & Butler, 2011). Conversely, in cases when no feedback is provided, the final recall benefits of retrieval practice mostly involve successfully practiced items, as unsuccessfully practiced items are rarely recalled on the final test (Glover, 1989; Halamish & Bjork, 2011; Pashler et al., 2005). The provision of feedback during retrieval practice can accordingly increase final memory performance compared to retrieval practice without feedback (Abel et al., 2019; Butler et al., 2008). However, in two meta-analyses, Adesope et al. (2017) found that the presence or absence of feedback during retrieval practice had no impact on the size of the testing effect, while Rowland (2014) reported higher effect sizes for studies that provided feedback during retrieval practice. Additionally, feedback after unsuccessful retrieval attempts has been shown to be more beneficial for final recall than restudy (Kornell et al., 2009), and delayed feedback seems to be even more beneficial for final recall than immediate feedback (Butler et al., 2007; Butler & Roediger, 2008).

The beneficial effects of feedback seem to be mostly restricted to previously not recalled materials: Karpicke and Roediger (2008) demonstrated that repeatedly restudying items that had already been recalled correctly once before led to lower final recall rates than repeated retrieval practice of these items and to comparable recall as a condition where once recalled items were not presented again (see also Soderstrom et al., 2016). Indeed, Pashler et al. (2005) administered two rounds of retrieval practice, with or without feedback, for previously studied vocabulary pairs and found substantial improvements in retention for initially incorrect responses when feedback was provided compared to a no-feedback condition, but no improvements when initial responses had been correct. Butler et al. (2008) also found large improvements on the final retention test through feedback for initially incorrect responses, but additionally reported improvements, if more modest in size, for initially correct responses. Further analyses revealed that providing feedback led to higher retention of initially correct answers that had been made with low confidence, while for correct answers that had been made with higher confidence, final retention was very similar, regardless of whether feedback had been provided or not. Providing feedback in a retention test after initial restudy or retrieval practice without feedback has also been shown to reverse the testing effect to a restudy benefit (Pastötter & Bäuml, 2016; Storm et al., 2014), but this reversal seems to emerge only in cases of relatively low success rates during retrieval practice (Racsmány et al., 2020).

Despite its long research history, the theoretical mechanisms responsible for retrieval practice effects remain under debate (for overviews of existing theoretical ideas see Karpicke et al., 2014; Rickard & Pan, 2018; Rowland, 2014). Based on the observation that benefits of retrieval practice are often bigger under more difficult initial retrieval conditions (e.g. Carpenter & DeLosh, 2006; Glover, 1989; Pyc & Rawson, 2009; see also R. A. Bjork, 1994), it was proposed that the increase in strength for practiced items is linked to initial *retrieval effort* (see also R. A. Bjork, 1975; Bjork & Bjork, 1992), but this idea does not explain exactly how this strengthening takes place. According to the concept of *transfer appropriate processing* (Morris et al., 1977), final retrieval success depends on the match in processes during final recall and initial practice. In this view, retrieval practice is more beneficial for retention than restudy because it is more similar to the final retention test (Karpicke, 2017). While this idea again proposes no exact mechanism for retrieval practice benefits, it has seen some support in the literature (e.g. Adesope et al., 2017; McDaniel & Fisher, 1991; Morris et al., 1977), but an exact match in test format for the initial and the final recall test does not always lead to the highest final recall (Carpenter & DeLosh, 2006; Glover, 1989).

The *bifurcation framework* (Halamish & Bjork, 2011; Kornell et al., 2011) also assumes that restudy and retrieval practice lead to lower vs. higher increases in memory strength for practiced items but adds the proposal that this leads to a bifurcated distribution in memory strength for retrieval practiced items: The framework assumes that restudy increases memory strength for all restudied items to a moderate, but uniform degree. Retrieval practice, on the other hand, increases memory strength to a higher degree, but only if it is successful, while unsuccessfully retrieval practiced items obtain no increase in memory strength. On a final memory test, successful recall depends on test difficulty, as only items that exceed a certain memory strength are retrievable. As successfully retrieval practiced items have a higher memory strength than restudied items, the former will be retrievable even on more difficult final tests. This framework can explain the observation that testing effects often only emerge at longer retention intervals, when test difficulty is high, with an initial benefit for restudied materials at short retention intervals, when test difficulty is low, and also the reversal of the testing effect by introducing feedback during a retention test (Racsmány et al., 2020; see Pastötter & Bäuml, 2016; Storm et al., 2014). Again, it is not specified exactly how retrieval practice confers greater increases in memory strength to items than restudy does.

Some other theoretical accounts do assume an explicit mechanism to explain the beneficial effects of retrieval practice. According to the *elaborative retrieval hypothesis* (Carpenter, 2009, 2011), memory search for a target item during a retrieval attempt leads to activation of elements that are also associated with the retrieval cue and form a more elaborate memory trace. These semantically related elements can then act as additional retrieval cues at later retrieval attempts, facilitating successful retrieval of target information (e.g. J. R. Anderson, 1983). This account can explain the benefit of retrieval practice over restudy, as restudy of target information should not lead to a memory search, and the observation of increased testing effects after more difficult retrieval practice, as a difficult memory search should lead to more activated related elements. However, the account has also been criticized, for example, because the association of additional related elements with the retrieval cue could lead to cue overload which would be detrimental to the retrieval of target information (e.g. Karpicke et al., 2014; Karpicke, 2017). Criticisms have also been based on findings such as that divided attention during the practice phase only impairs retention for restudied items but not for retrieval practiced items (Mulligan & Picklesimer, 2016), at odds with the conceptualization of elaboration as an effortful process. It is also mostly restricted to verbal study materials, leaving it open how testing effects e.g. for spatial materials can be explained (e.g. Carpenter & Pashler, 2007). A related account is the *mediator effectiveness hypothesis* (Pyc & Rawson, 2010), which postulates that retrieval practice leads to the establishment of more effective mediators that connect retrieval cue and target (see also Bouwmeester & Verkoeijen, 2011).

The *episodic context account* of retrieval practice benefits (Karpicke et al., 2014) is based on the assumption that items are encoded along with contextual information about the encoding context (Howard & Kahana, 2002; Raaijmakers & Shiffrin, 1981) and that retrieval often includes the attempt to reinstate the target's encoding context (Lehman & Malmberg, 2013). At each retrieval, and due to contextual drift (Howard & Kahana, 2002), new contextual information is associated with the target and can serve as additional retrieval cues at future retrieval attempts and also help to restrict the search set (Karpicke et al., 2014), facilitating retrieval success in comparison to restudy conditions, where no new contextual information is encoded. Again, this account is in line with the observation of greater performance benefits after retrieval practice than after restudy and with the general observation of bigger testing effects after difficult initial retrieval, as more contextual elements would need to be reinstated to ensure successful retrieval under such conditions. Support for this idea comes e.g. from the finding that recollection of contextual information connected to studied materials is enhanced when the practice phase contains retrieval practice (Akan et al., 2018).

Recently, Carpenter and Yeung (2017) argued that the elaborate retrieval account could be reconciled both with the assumptions of the bifurcation framework and the episodic context account. The authors posited that the increases in memory strength through retrieval practice might in part be due to the establishment of mediating memory elements and that such mediating information might also be viewed as a contextual cue, pointing out that "the factors underlying the testing effect are likely to be multifaceted, such that the effect is best accounted for by a number of potentially inter-related mechanisms" (p. 138). In more recent years, several other theories have also been proposed to explain testing effects, attributing benefits of retrieval practice to increased automatization of retrieval (Racsmány et al., 2018), fast memory consolidation (Antony et al., 2017) or the establishment of a secondary memory trace during retrieval practice (Rickard & Pan, 2018), adding to the group of potential explanatory mechanisms.

3.2 The Role of Delay For Retrieval Practice Effects

A common observation in the field of retrieval practice effects is that of a stabilization of memory performance after retrieval practice. For example, Potts and Shanks (2012) reported reduced retroactive interference when participants were allowed to practice retrieval of first-list items without feedback before learning a second, interfering list (see also Halamish & Bjork, 2011). In a related vein, Kliegl and Bäuml (2016) demonstrated a reduction in intra-list interference after retrieval practice (without feedback) compared to restudy conditions, such as eliminated retrieval-induced forgetting and output interference. Congleton and Rajaram (2011) found collaborative inhibition – the finding that memory performance is often impaired for participants who collaboratively recall in a group setting in comparison to participants who recall individually and form a group in name only – to be eliminated when participants performed retrieval practice (without feedback) of studied materials individually before a group recall. In a study by Abel and Bäuml (2016), list-method directed forgetting was absent when List 1 items were retrieved (without feedback) before the cue to forget or remember List 1 was presented, but not when List 1 items were restudied. Also, there is some tentative evidence of a protective effect against stress that takes place after study and a practice phase, but before the final test (A. M. Smith et al., 2016; A. M. Smith et al., 2018; see also Wolf & Kluge, 2017). However, Szőllősi et al. (2017) found no influence of stress induction on memory performance, neither in a restudy condition nor in a retrieval-practice-with-feedback condition. Altogether, many detrimental effects on memory apparently can be eliminated or reduced by introducing retrieval practice before the manipulation that usually impairs performance, resulting in a shielding or stabilizing effect. Accordingly, it seems worthwhile to investigate whether retrieval practice also has a stabilizing effect on long-term memory performance by reducing the rate of time-dependent forgetting. Such an idea is not new, as can be seen in Runquist (1986a), who stated that "the primary effect of recalling studied material is to reduce the rate at which that material is forgotten“ (p. 282). Retrieval practice should however not be expected to be a panacea against forgetting – as Bartlett (1932, see also Wheeler & Roediger, 1992) showed in his seminal work, previously retrieved material will still be forgotten and misremembered at prolonged retention intervals. The question of interest, then, is rather whether the observable forgetting is slowed down compared to other practice strategies, such as restudy.

To examine how retrieval practice effects vary depending on delay, the following literature review focuses on studies that varied the time interval between the end of the practice phase and the final test, and did so in the same way for all practice conditions, keeping the time between initial study and the end of the practice phase roughly constant for all conditions. In the related research field investigating spacing or lag effects, the latter interval – that between initial study and the end of the practice phase – is often varied as well, confounding the length of the retention interval between initial study and final test with the respective practice condition (see e.g. Cepeda et al., 2008, 2009).

3.2.1 Previous ANOVA-based Work

Most of the studies that examine the role of delay for retrieval practice effects compare final recall at two delay intervals after initial study and the intervening practice phase, usually one short delay of a few minutes and one longer delay of one or several days. The practice phase in these studies consists either of retrieval practice (with or without corrective feedback) or, as a control condition, of restudy (some studies also include a no-practice condition, e.g. Mulligan & Picklesimer, 2016; M. A. Smith et al., 2013; Runquist, 1983, Exp. 1; Wheeler & Roediger, 1992). In these studies, it is then of interest, whether, for both types of practice, memory performance decreases similarly from the short to the long delay, or whether the two conditions show different drops in performance across the retention interval. In the case of different decreases in performance, a significant test-delay interaction of the factors delay interval and practice type would arise in an ANOVA, which is assumed to imply different rates of forgetting.

Studies that compare retrieval practice without feedback with a restudy condition have produced both significant (e.g. Abel et al., 2019, Exp. 1a; Mulligan & Picklesimer, 2016; Thompson et al., 1978, Exp. 2; Toppino & Cohen, 2009; Wenger et al., 1980, Exp. 3; Wheeler et al., 2003) and non-significant test-delay interactions (e.g. Abel et al., 2019, Exp. 2b; Agarwal et al., 2017; Thompson et al., 1978, Exp. 3; Wenger et al., 1980, Exp. 2). In some of these studies, pronounced cross-over interactions emerged, with a performance benefit for restudy at the short delay that reversed to a performance benefit for retrieval practice at the long delay (e.g. Toppino & Cohen, 2009, Exp. 1; Wheeler et al., 2003). For studies employing retrieval practice with feedback, similarly both significant (Abel & Bäuml, 2020; Abel & Roediger, 2018;

Mulligan & Peterson, 2015) and non-significant interactions (Abel et al., 2019, Exp. 1a; Carrier & Pashler, 1992) have been reported in the literature. Overall, more significant than non-significant interactions have been reported for both kinds of retrieval practice, indicating that time-dependent forgetting might differ between restudy and retrieval practice conditions. Exact methods and designs differ widely in these studies – for example regarding the number of practice cycles, within- vs. between-subjects manipulations and length of retention intervals – making it harder to examine potential differences in time-dependent forgetting between practice strategies or potential mediating factors. Some studies also included additional practice conditions beside a retrieval practice condition and a restudy condition, for example a no-practice condition, or both retrieval practice with and without feedback, but reported only global test-delay interactions for all included conditions at once (e.g. Kornell et al., 2011, Exp. 1 & 2; M. A. Smith et al., 2013).

Only a few studies so far have examined retrieval practice effects across more than two retention intervals. Roediger and Karpicke (2006b) let their participants study prose materials, and then compared the effect of restudy or retrieval practice without feedback after initial study at three different retention intervals (5 min, 2 days and 7 days). In an ANOVA, a significant test-delay interaction emerged (see also Bertilsson et al., 2021), with a steeper drop in memory performance for materials that had been restudied. Similar to the findings of Toppino and Cohen (2009, Exp. 1) and Wheeler et al. (2003), restudied materials were recalled better than retrieval practiced materials after the short delay of 5 min, but the pattern was reversed at the longer retention intervals. Some other studies also examined memory performance after retrieval practice without feedback at a short and several longer delays. However, instead of using restudy as a control condition, they compared the effects of retrieval practice with those of a no-practice condition (Chan, 2010; Runquist, 1983, Exp. 2; Runquist, 1986b, 1987; Slamecka & Katsaiti, 1988; Spitzer, 1939). In such a study set-up, practice condition is confounded with the total amount of exposure to studied materials, making it difficult to pinpoint why retrieval practiced materials in these studies seem to be forgotten at a slower pace, as numerically, retention appears to be stable across delays after retrieval practice, but not after the no-practice condition.

Particularly noteworthy among these studies is the pioneering work by Spitzer (1939), who tested over 3000 sixth-graders on their retention of prose materials. All subjects practiced retrieval of studied materials before their final retention test that took place 1, 7, 14, 21, 28 or 63 days after study, depending on their practice schedule group. Spitzer included no groups without retrieval practice, but, at every retention interval, his experimental design allowed for a comparison of recall performance between one group that had been tested before and one that was tested for the first time. Memory performance during retrieval practice across all retention intervals followed the typical curvilinear shape of Ebbinghaus' (1885) forgetting curve, but within each group of participants, memory performance showed only small decreases from retrieval practice to the final test, creating performance benefits at each retention interval compared to the respective group that had not been tested before. This benefit was especially pronounced for groups that conducted retrieval practice within the first few days after initial study. Taken together, his results demonstrate stark changes to the typical forgetting curve by introducing a single test before final recall.

Overall, the majority of the literature on the effects of retrieval practice on long-term memory performance has used ANOVAs for data analysis, and indicates differences between restudy conditions and retrieval practice conditions, both with and without feedback. In these studies, predominantly significant test-delay interactions have been reported between the two types of retrieval practice and restudy, due to more pronounced absolute decreases in memory performance after restudy than after retrieval practice.

3.2.2 Previous Work Using Power Function Analysis

Unlike in the case of directed forgetting, relative time-dependent forgetting in the context of retrieval practice effects has been examined using power function analysis, if only by two studies so far: Carpenter et al. (2008) manipulated both type of practice and length of retention interval within-subjects and had their participants study obscure facts (Exp. 1 & 2) or Swahili-English vocabulary pairs (Exp. 3). After study, some materials were retrieval practiced with feedback, while others were restudied. All items were practiced once (Exp. 1) or three times (Exp. 2 & 3) in both practice conditions. After the practice phase, six retention tests were conducted at six different retention intervals (5 min or 1, 2, 7, 14 or 42 days), at each of which, only one of six different portions of the studied materials was tested. This way, each item was only tested one final time after the practice phase. The authors then estimated individual power curves of the form $r(t) = a(1 + ct)^{-b}$ for each subject and for restudied vs. retrieval practiced materials. Scaling parameter c was fixed across subjects and practice conditions. Results from the three experiments were mixed: In Experiments 1 and 2, Carpenter et al.

(2008) found forgetting rate b to be significantly reduced after retrieval practice with feedback compared to restudy, but Experiment 3 only produced a numerical trend in the same direction. In contrast, ANOVAs showed a significant test-delay interaction only in Experiment 1, underlining that the two methods of analysis do not necessarily support the same conclusions regarding differences in time-dependent forgetting between conditions (see also Wixted, 2022a).

Siler and Benjamin (2020, Exp. 2) modeled their study on the one by Carpenter et al. (2008), but introduced several important deviations from their experimental design. Siler and Benjamin also used a full within-subject manipulation, contrasted restudy and retrieval practice with feedback and examined retention after few minutes, 1, 7 and 25 days. As Carpenter et al. (2008), they also estimated individual fits of the three-parametric power function to their data. Participants were presented with pictures of different birds and studied which taxonomic families the birds belonged to. After initial presentation, some bird families were restudied and some were retrieval practiced with feedback (two times each). Importantly, Siler and Benjamin (2020) were interested both in long-term memory for studied pictures as well as in transfer effects to new pictures, which is why participants performed two tests at every retention interval: For both studied and new images of birds, they were first required to assign them to their taxonomic families (categorization test), and then to judge whether they had seen the image before during the course of the experiment (old/new-recognition test). Siler and Benjamin reported no difference in forgetting rates between retrieval practice and restudied items for the categorization performance of studied items, but there was a numerical trend of a lower forgetting rate b for restudied items. An ANOVA produced a significant test-delay interaction. Conversely, the recognition test showed a numerical trend of reduced time-dependent forgetting after retrieval practice with feedback, but this difference was not significant. This time, the test-delay interaction was not significant in an ANOVA. In consideration of these results it is important to note that while each individual bird image was present on only one of the four recall tests, taxonomic families were repeated across all four recall events. Accordingly, each recall event was also an additional learning opportunity for the taxonomic families, which produces an interdependence of the recall events. However, it is unclear to what extent this interdependence influenced the observed forgetting rates for the categorization data.

All in all, the two previous studies produced mixed results. When taking into account only the recognition test of Siler and Benjamin (2020) as well as the three experiments of Carpenter et al. (2008), a significant difference in relative forgetting rates between restudy and retrieval practice with feedback was present in only two out of four cases. At the same time, all four experiments showed a numerical pattern of reduced time-dependent forgetting after retrieval practice with feedback compared to restudied materials.

3.2.3 Theoretical Expectations

The more prominent recent theoretical accounts of retrieval practice effects are largely in line with the reported pattern across a large part of the testing effect literature of a relative stabilization in memory performance after retrieval practice: Carpenter (2011) referred to work by Bartlett (1932), indicating that long-term memory might rely to a larger degree on semantic organization, and concluded that time-dependent forgetting might be reduced after retrieval practice because the increased semantic elaboration might make those memory traces more stable across time. Similarly, the episodic context account (Karpicke et al., 2014; Karpicke, 2017) assumes that the benefit of retrieval practice over restudy should be increased when final recall is delayed and studied materials are therefore less accessible, which should induce a greater reliance on contextual cues, favoring recall after retrieval practice. However, it is unclear how the basic assumptions of both of these accounts would translate to relative forgetting rates for retrieval practiced and restudied materials. Additionally, the accounts remain silent on the role of feedback and whether differences in relative forgetting should emerge between retrieval practice conditions with and without feedback.

The bifurcation framework of retrieval practice effects on the other hand (Halamish & Bjork, 2011; Kornell et al., 2011) explicitly assumes that there are in fact no differences in forgetting rates for retrieval practiced vs. restudied materials. Test-delay interactions are instead attributed to selective overlearning of successfully retrieval practiced items. The framework also assumes that test-delay interactions should be reduced or absent in comparisons of restudy and retrieval-practice-with-feedback conditions, as the introduction of feedback should increase the memory strength of unsuccessfully retrieval practiced items, thereby reducing the gap in memory strength to successfully retrieval practiced items. The distribution of memory strength should accordingly be bifurcated to a greater extent after retrieval practice without feedback than after retrieval practice with feedback (see also Abel et al., 2019). The expectation of non-significant test-delay interactions for retrieval practice with feedback is at odds with at least some of the existing data on retrieval practice effects (e.g. Abel &

Bäuml, 2020; Mulligan & Peterson, 2015), and it is not clear whether the assumption of a constant rate of forgetting regardless of practice type within the bifurcation framework refers to absolute or relative losses in memory strength, making it hard to derive testable predictions.

3.3 Goals of Experiments 4 & 5

Based on the mixed results of Carpenter et al. (2008) and Siler and Benjamin (2020), Experiments 4 and 5 of this thesis were intended in a first step to address again the comparison of restudy and retrieval practice with feedback. The second goal was then to investigate whether the previously reported finding of a reduced forgetting rate after retrieval practice with feedback can also be observed after retrieval practice without feedback, or whether the two forms of retrieval practice behave differently across long retention intervals, when compared with restudy. As described above, previous studies had reported parallels (see e.g. Abel & Bäuml, 2020; Agarwal et al., 2017; Toppino & Cohen, 2009) between the two conditions, but also differences in (absolute) rates of forgetting (Abel et al., 2019; Kornell et al., 2011). A third goal was to examine whether any kind of repetition suffices to slow down the rate of forgetting or whether such effects are exclusive to retrieval practice conditions by comparing restudy with a no-practice control condition. Different kinds of practice could also serve as relevance-signals for the future (Cowan et al., 2021; J. R. Anderson & Schooler, 1991) that vary in strength and affect the rate of forgetting. As in Experiments 1-3, both ANOVAs and power function analysis were used to examine time-dependent forgetting.

Experiment 4 was designed as a follow-up on the results of Carpenter et al. (2008) and Siler and Benjamin (2020), contrasting relative forgetting rates of restudy and retrieval practice with feedback, of restudy and retrieval practice without feedback, and also of both types of retrieval practice. Experiment 5 was intended to replicate Experiment 4 by again comparing retrieval practice without feedback and restudy, and additionally included a control condition that contained no further exposure to studied materials after initial study. This condition was compared to restudy, in order to investigate the influence of practice-free repetition on time-dependent forgetting. In a deviation from both previous studies, pairs of concrete nouns were used as study materials for both experiments. Also, retention interval was manipulated between-subjects, in order to prevent a possible interdependence of multiple recall events for individual participants: (Partial) context reinstatement (e.g. Bäuml & Samenieh, 2010;

Bäuml & Trißl, 2022) could have reactivated as yet untested materials at each recall event in both Carpenter et al. (2008) and Siler and Benjamin (2020), thereby altering the rate of forgetting of these materials. For both Experiments 4 and 5, four retention intervals each were used: Experiment 4 used retention intervals of 3 min, 1, 3 and 7 days, while Experiment 5 used intervals of 3 min, 1, 2 and 3 days. This change was introduced to focus on the time period shortly after study and practice phase, as retention has been observed to stabilize at longer delays (e.g. Ebbinghaus, 1885). Overall, Experiments 4 and 5 were aimed at an investigation of whether relative time-dependent forgetting is actually reduced after retrieval practice and whether the two types of retrieval practice – with and without feedback – differ in this respect or not. The other main goal was to examine whether restudy, as a retrieval-free mode of practice, can also reduce forgetting.

3.4 Experiment 4: Replication of Previous Work and an Investigation of the Role of Feedback

3.4.1 Method

Participants

144 participants took part in the experiment ($M = 22.32$ years, $SD = 2.59$, range = 18-30 years, 102 female). As for Experiments 1-3, they were recruited mainly from Regensburg University, as well as by placing online advertisements in students' groups in Germany. 93.1 % of the participants were currently enrolled at university, while the remaining subjects reported to be employed or to undergo vocational training. Participants were distributed equally across the four between-subjects conditions, yielding $n = 36$ participants per delay condition. Sample size was again based on a power analysis conducted in G*Power (version 3.1.9.7, Faul et al., 2009) with $\alpha = .05$, $\beta = .20$, and $\eta_p^2 = 0.06$, as previous studies often had reported small to medium effect sizes ($\eta_p^2 s$ of 0.05 to 0.11) for the ANOVA interaction between delay and type of practice (e.g. Abel & Bäuml, 2020; Mulligan & Peterson, 2015; Toppino & Cohen, 2009) - as well as counterbalancing purposes. This way, sample size was also similar to studies from other research areas that compared relative forgetting across different experimental conditions (e.g., Bäuml & Trißl, 2022).

Materials

Study materials consisted of 24 unrelated word pairs of concrete nouns, which were drawn from van Overschelde et al. (2004) and had already been used by Bäuml et al. (2014). The nouns were chosen from different semantic categories, with 1-2 syllables each. Within each pair, one word was always used as the cue word while the other word served as the target word. The list of 24 pairs was divided into three sets of eight pairs each (A, B, and C). Assignment of sets to practice conditions and order of sets during the practice phase were counterbalanced across participants within each delay condition, resulting in 36 different combinations of type of practice and set sequence.

Design

The experiment followed a 3 (TYPE OF PRACTICE: restudy vs. retrieval practice with feedback vs. retrieval practice without feedback) × 4 (DELAY: 3 min vs. 1 day vs. 3 days vs. 7 days) mixed design. TYPE OF PRACTICE was manipulated within-subject, DELAY was manipulated between-subjects.

Procedure

Data collection took place via individual zoom meetings, following the same basic procedures as for Experiments 1-3. Both participants and experimenters were required to keep their cameras and microphones on during the meetings. For participants in the long delay conditions (1 day, 3 days, or 7 days), the experiment consisted of two sessions that were scheduled for the same time of day (± 2 hrs). The software OpenSesame (version 3.3, Mathôt et al., 2012) was used for stimulus presentation and balancing. During sessions, experimenters shared their screen for stimulus presentation and instructed participants orally.

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

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

Fitting Procedures

The fitting procedure was similar to the ones employed in Experiments 1-3. This time, before running more detailed analyses, a common scaling parameter c was estimated for the three practice conditions using the same procedure as in Experiments 1-3. As before, the restricted power function model with a common scaling parameter c for the three practice conditions described the recall rates equally as well as the more general power function model in which the parameter varied freely across conditions, $\chi^2(2) = 0.34$, with $c = 0.45$ as the best fitting parameter. Like in Experiments 1-3, this parameter estimate was used for all further analyses.

3.4.2 Results

Success Rates

Mean recall rates during the practice phase were .75 ($SD = .23$) for retrieval practice with feedback and .76 ($SD = .22$) for retrieval practice without feedback. A 4 × 2 mixed-factors ANOVA with the between-subjects factor of DELAY (3 min vs. 1 day vs. 3 days vs. 7 days) and the within-subject factor of TYPE OF PRACTICE (retrieval practice with feedback vs. retrieval practice without feedback) produced no significant effects, neither of DELAY, $F(1, 140) = 0.48$, $MSE = .08$, $p = .696$, $\eta_p^2 = 0.01$, nor of TYPE OF PRACTICE, $F(1, 140) = 0.46$, $MSE = .02$, $p = .498$, $\eta_p^2 < 0.01$, nor for the interaction between both factors, $F(3, 140) = 2.04$, $MSE = .02$, $p = .111$, $\eta_p^2 = 0.04$, indicating that these success rates did not vary significantly across practice conditions or delays.

Analysis of Variance

Figure 11 shows the mean recall rates for all three practice conditions across the four delay intervals. Using ANOVA, recall rates in the restudy condition were first compared to those in the retrieval-practice-with-feedback and the retrieval-practice-without-feedback conditions. Finally, recall rates between the two retrieval practice conditions were compared. Regarding the comparison between restudy and retrieval practice with feedback, a 2 × 4 mixed-factors ANOVA with the within-subject factor of TYPE OF PRACTICE (restudy vs. retrieval practice with feedback) and the between-subjects factor of DELAY (3 min vs. 1 day vs. 3 days vs. 7 days) revealed a main effect of TYPE OF PRACTICE, $F(1, 140) = 14.34$, $MSE = .02$, $p < .001$, $\eta_p^2 = 0.09$, with lower recall in the restudy condition, a main effect of DELAY, $F(3, 140) = 49.93$, $MSE = .10$, $p < .001$, $\eta_p^2 = .52$, with lower recall after longer than shorter delay, as well as a significant interaction between the two factors, $F(3, 140) = 3.06$, $MSE = .02$, $p = .030$, $\eta_p^2 = 0.06$, reflecting the fact that the detrimental effect of delay on recall performance was larger in the restudy than in the retrieval-practice-with-feedback condition.

The comparison between restudy and retrieval practice without feedback showed no main effect of TYPE OF PRACTICE, $F(1, 140) = 0.39$, $MSE = .02$, $p = .535$, $\eta_p^2 < 0.01$, a main effect of DELAY, $F(3, 140) = 46.88$, $MSE = .10$, $p < .001$, $\eta_p^2 = .50$, and a non-significant numerical interaction between the two factors, $F(3, 140) = 2.39$, $MSE = .02$, $p = .071$, $\eta_p^2 = 0.05$, indicating a trend towards a larger detrimental effect of delay in the restudy condition. Finally, the comparison between the two retrieval

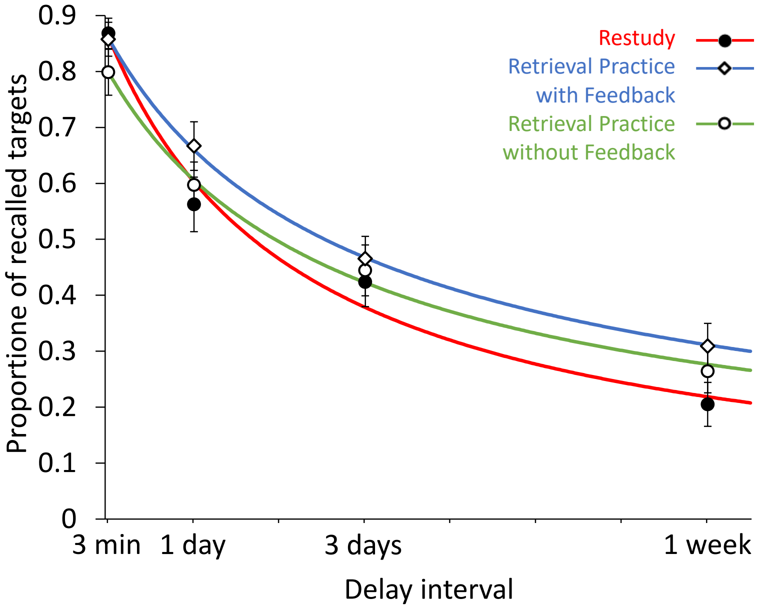

Figure 11: *Results of Experiment 4.* Recall rates for the three practice conditions are displayed along with the best-fitting power functions. The power function described the recall rates well, showing a larger forgetting rate parameter in the restudy condition than in the two retrieval practice conditions, which did not differ in forgetting rates (see Table 7). Error bars represent ± 1 standard error.

practice conditions revealed a main effect of TYPE OF PRACTICE, $F(1, 140) = 8.21$, $MSE = .02$, $p = .005$, $\eta_p^2 = 0.06$, with overall higher recall after retrieval practice with feedback, and a main effect of DELAY, $F(3, 140) = 40.83$, $MSE = .10$, $p < .001$, $\eta_p^2 = .47$, but no interaction between the two factors, $F(3, 140) = 0.38$, $MSE = .02$, $p = .765$, $\eta_p^2 < 0.01$, pointing to a similarly sized delay effect for the two retrieval practice formats.

Power Function Analysis

The power function described the recall rates of all three practice conditions well (Figure 11), as is reflected in $\chi^2(2)$ values of ≤ 4.82 for the single practice conditions (Table 7). Again, the power function explained most of the variance in the data, as is represented in r^2 values of $\geq .983$ in the three conditions. Parameters a and b of the function in the restudy condition were first compared with those in the two retrieval practice conditions. Finally parameters between the two retrieval practice formats were compared. Regarding the comparison between restudy and retrieval practice with feedback, estimates for parameter a did not differ significantly between conditions, $\chi^2(1) < 0.01$, suggesting similar initial recall levels in the two conditions. In contrast, the two conditions differed in parameter b, $\chi^2(1) = 10.15$, reflecting a higher forgetting rate for restudy than for retrieval practice with feedback.

For the comparison between restudy and retrieval practice without feedback, parameter a was found to be significantly smaller for retrieval practice than for restudy, $\chi^2(1) = 4.58$, indicating that initial recall was lower in the retrieval practice condition. Parameter b was also significantly smaller in the retrieval practice condition, $\chi^2(1) = 6.92$, reflecting a higher forgetting rate for restudy than for retrieval practice without feedback. Finally, when comparing the two retrieval practice conditions, parameter a was significantly smaller in the retrieval-practice-without-feedback condition, $\chi^2(1) = 4.40$, indicating lower initial recall in this condition, whereas forgetting rate parameter b was comparable for the two retrieval practice formats, $\chi^2(1) = 0.17$.

Table 7

Best Power Function Fits and Explained Variance for Experiment 4

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

3.4.3 Discussion

In all three practice conditions, and mirroring the findings of Experiments 1-3, time-dependent forgetting was well described by the three-parametric power function of time, which is also in line with the previous studies by Carpenter et al. (2008) and Siler and Benjamin (2020). For the comparison between restudy and retrieval practice with feedback, a lower forgetting rate parameter (b) emerged in the retrieval-practice-with-feedback condition than in the restudy condition, replicating the findings from two of the three experiments reported in Carpenter et al. (2008). Extending these previous findings, the results of Experiment 4 also show that the reduction in relative time-dependent forgetting generalizes from retrieval practice with feedback to retrieval practice without feedback, with forgetting rates even comparable in size for the two retrieval practice formats. Taken together, these results seem to indicate that retrieval practice - both with and without feedback - slows down time-dependent forgetting.

It has been pointed out before that, due to the focus on absolute (ANOVA) versus relative forgetting (power function analysis), the conclusions drawn from the two methods do not necessarily match in comparisons of time-dependent forgetting across experimental conditions (see e.g. Carpenter et al., 2008; Wixted, 2022a). Nevertheless, for Experiment 4, the results from the ANOVAs largely mimicked those from the power function analysis: ANOVAs revealed a larger amount of (absolute) time-dependent forgetting for restudy compared to retrieval practice with feedback, at least a numerical trend towards a larger amount of forgetting for restudy compared to retrieval practice without feedback, and no difference in amount of forgetting between the two retrieval practice formats.

As the finding of a reduced (relative) forgetting rate after retrieval practice without feedback relative to restudy had not been reported before, Experiment 5 was designed to replicate this result. Experiment 5 differed from Experiment 4 in three ways: First, no initial letter cues were provided for the target words during final recall and recall was instead conducted without presenting any target-specific retrieval cues in an effort to mimic the standard procedure of free recall more closely (e.g. Roediger & Karpicke, 2006b; see also Kornell et al., 2011; Toppino & Cohen, 2009). Second, instead of a retrieval-practice-with-feedback condition a no-practice condition was included to look at whether restudy, as a retrieval-free method of practice, also decreases the relative rate of time-dependent forgetting. Third, the 7-days delay condition was replaced with a 2-days delay condition, to concentrate on the time period shortly after initial study that is defining for the overall shape of the forgetting curve.

3.5 Experiment 5: Replication of Experiment 4 and an Investigation of the Role of Restudy

3.5.1 Method

Participants

Another 144 participants were recruited for the experiment (M = 22.43 years, SD = 2.91, range = 18-30 years, 86 female), mainly from Regensburg University but also by placing online advertisements in students' groups in Germany. 73.6 % of the participants were currently enrolled at university, while the remaining subjects reported to be employed or to undergo vocational training. Participants were again distributed equally across the four between-subjects conditions, yielding $n = 36$ participants per delay condition. Sample size followed Experiment 4.

Materials

24 new unrelated word pairs of concrete nouns from different semantic categories (1-2 syllables) were used, drawn from van Overschelde et al.'s (2004) norms. Pairs were partially sampled from Bäuml et al. (2014). Like in Experiment 4, the pairs were divided into three sets of 8 pairs each (A, B, and C). Assignment of sets to practice conditions and order of sets during the practice phase were again counterbalanced across participants within each delay condition. Experiment 5 had only 12 different combinations of type of practice and set sequence, as only two (instead of all three) sets were presented again during practice.

Design

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

Procedure

Experiment 5 closely followed Experiment 4 with the following exceptions: a) Between the two study cycles and after the second cycle, a different 1 min distractor task was used: subjects were presented with two two-digit numbers and were then asked to first add up the digits for both numbers and then either add or subtract the two separate results; b) The retrieval-practice-with-feedback condition was replaced with a no-practice condition and the set of items assigned to this condition (A, B, or C) was not presented again after the two initial study cycles; c) during the final test, subjects were asked to recall the targets without any item-specific retrieval cues.

Fitting Procedures

The fitting procedure was identical to the one employed in Experiment 4. Again, before running more detailed analyses, a common scaling parameter c was estimated for the three practice conditions using the same procedure as in Experiment 4. Like in Experiment 4, the restricted power function model with a common scaling parameter c for the three practice conditions described the recall rates equally as well as the more general power function model in which the parameter varied freely across conditions, $\chi^2(2) = 1.70$, with $c = 27.78$ as the best fitting parameter. Like in Experiments 1-4, this parameter estimate was used for all further analyses.

3.5.2 Results

Success Rates

Mean recall rates during the practice phase were .83 ($SD = .20$) for retrieval practice without feedback. According to a univariate ANOVA, these success rates did not differ significantly across DELAY conditions, $F(3, 140) = 0.37$, $MSE = .04$, $p = .778$, $\eta_p^2 < 0.01$.

Analysis of Variance

Figure 12 shows the mean recall rates for all three practice conditions across the four delay intervals. Using ANOVA, recall rates in the restudy condition were compared to those in the retrieval-practice-without-feedback condition as well as to those in the no-practice condition. For the comparison between restudy and retrieval practice without feedback, a 2 × 4 mixed-factors ANOVA with the within-subject factor TYPE OF PRACTICE (restudy vs. retrieval practice without feedback) and the between-subjects factor DELAY (3 min vs. 1 day vs. 2 days vs. 3 days) revealed no main effect of TYPE OF PRACTICE, $F(1, 140) = 2.37$, $MSE = .02$, $p = .126$, $\eta_p^2 = 0.02$, a main effect of DELAY, $F(3, 140) = 10.61$, $MSE = .13$, $p < .001$, $\eta_p^2 = .19$, with lower recall after longer delays, and a significant interaction between the two factors, $F(3, 140) = 3.06$, $MSE = .02$, $p = .030$, $\eta_p^2 = 0.06$, indicating that the detrimental effect of delay on recall rates was larger in the restudy condition. Comparing restudy and the no-practice condition, a main effect of TYPE OF PRACTICE emerged, $F(1, 140) = 101.31$, $MSE = .03$, $p < .001$, $\eta_p^2 = 0.42$, with higher recall in the restudy condition, as well as a main effect of DELAY, $F(3, 140) = 14.61$, $MSE = .11$, $p < .001$, $\eta_p^2 = .24$, but no significant interaction between the two factors, $F(3, 140) = .94$, $MSE = .03$, $p = .422$, $\eta_p^2 = 0.02$.

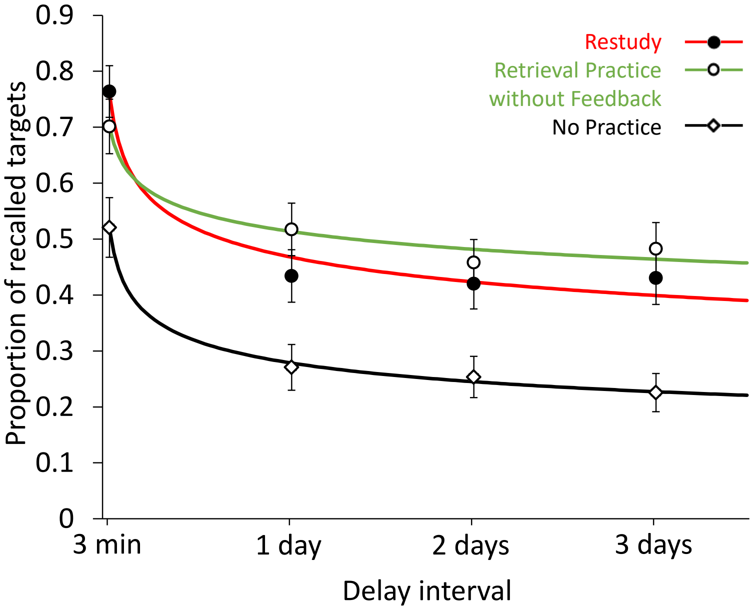

Figure 12: *Results of Experiment 5.* Recall rates for the three practice conditions are displayed along with the best-fitting power functions. The power function described the recall rates well, showing a larger forgetting rate parameter in the restudy than the retrieval practice without feedback condition. Forgetting rates did not differ between the restudy and the no-practice conditions (see Table 8). Error bars represent ± 1 standard error.

Power Function Analysis

Again, the power function described the recall rates in the three practice conditions well (Figure 12), as is reflected in $\chi^2(2)$ values of ≤ 2.52 for the single conditions (Table 8). Consistent with this finding, in all three conditions the power function also explained most of the variance in the data, as is reflected in r^2 values $\geq .975$ in the three conditions. Comparing parameters a and b between restudy and the retrieval-practice-without-feedback condition, parameter a was found not to differ between the two conditions, $\chi^2(1) = 2.70$, suggesting similar initial recall, whereas forgetting rate parameter b was found to be higher in the restudy than in the retrieval-practice-without-feedback condition, $\chi^2(1) = 7.96$. Comparing parameters a and b between restudy and the no-practice condition indicated that parameter a was significantly smaller in the no-practice condition, $\chi^2(1) = 35.15$, suggesting lower initial recall in this condition, whereas no difference was found in parameter b, $\chi^2(1) = 2.83$, pointing to similar forgetting rates in the two conditions.

It should be noted that the parameter estimates for b had much higher values in Experiment 4 than Experiment 5, which is connected to the fact that the best fitting

Table 8

Best Power Function Fits and Explained Variance for Experiment 5

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

scaling parameter c also differed considerably for the two experiments. The recall rates in Experiment 4 were well described by a relatively low parameter c with relatively high parameters b, whereas the recall rates in Experiment 5 were fitted better by a relatively high parameter c with relatively low parameters b. When the recall rates of the two experiments were fit with the two-parametric power function, $r(t) = at^{-b}$, which contains no scaling parameter, the estimated values of parameter b were much more similar for the two experiments. In Experiment 4, estimates of b were 0.113 for restudy, 0.084 for retrieval practice with feedback, and 0.089 for retrieval practice without feedback; in Experiment 5, estimates of b were 0.085 for restudy, 0.054 for retrieval practice without feedback, and 0.110 for the no-practice condition.

3.5.3 Discussion

As in Experiments 1-4, the power function of time described time-dependent forgetting for all three practice conditions well. In addition, Experiment 5 replicated the new finding of Experiment 4 of a smaller forgetting rate parameter after retrieval practice without feedback than after restudy, corroborating the finding that retrieval practice both with and without feedback can reduce time-dependent forgetting. The results also showed that restudy in comparison to a no-practice condition did not lead to a significant reduction in forgetting rate. However, the parameter value was numerically smaller after restudy, leaving it open whether a significant reduction relative to a no-practice condition might emerge in a study with increased statistical power.

Similar to Experiment 4, the ANOVA results turned out to run parallel to those of the power function analyses: for the comparison between restudy and retrieval practice without feedback, a significant interaction between delay and type of practice emerged, implying less absolute forgetting after retrieval practice than after restudy. For the comparison between restudy and the no-practice condition, the interaction was not significant, indicating a similar absolute amount of forgetting in the two conditions.

3.6 Additional Analyses for Experiments 4 & 5

As for Experiments 1-3, the data for both Experiments 4 & 5 was reanalyzed using geometric averaging (e.g. R. B. Anderson & Tweney, 1997; Wixted & Ebbesen, 1997). Again, geometric averaging for Experiments 4 and 5 led to the same pattern of results as arithmetical averaging (see Table 9). The same fitting procedures as for arithmetically averaged data were used. Estimates for parameter c were numerically very similar to those derived for arithmetic averages for both experiments. For all practice conditions, the power function described the recall rates equally as well as the statistical baseline models. Importantly, estimates of forgetting rate parameter b again were smaller for both retrieval practice conditions compared to restudy in Experiment 4, while b was very similar for the two retrieval practice conditions. For Experiment 5, estimated parameters b closely resembled those for arithmetically averaged data, and again, retrieval practice without feedback produced the lowest rate of forgetting, followed by restudy and the no-practice condition, which differed to a smaller extent. Again, these findings indicate that the results presented above were not exclusively produced by averaging artifacts.

Previous research has demonstrated the impact of initial retrieval success on final recall (Glover, 1989; Pashler et al., 2005), indicating that forgetting rates might also be influenced by this factor. Data for all three retrieval practice conditions of Experiments 4 and 5 were therefore re-analyzed, by determining for each participants which items were correctly recalled during retrieval practice (successfully practiced) and which were not. In a next step it was then examined how many out of these two item-groups were correctly recalled at the final test. Success rates during retrieval practice were however relatively high in the present experiments, with .75 (with feedback) and .76 (without feedback) in Experiment 4, and .83 in Experiment 5 (without feedback), meaning that only relatively few items were practiced unsuccessfully. Out of these, even in the shortest delay condition (3 min), less than a quarter of the items were recalled correctly at test when no feedback was given, and only about half when feedback was provided, meaning that, on average, less than one unsuccessfully practiced item was recalled at test per participant. These numbers decreased even further with length of

Table 9

Best Power Function Fits and Explained Variance for Geometrically Averaged Data of Experiments 4 & 5

Condition	**a**	**b**	**c**	**$\chi^2(2)$**	**r^2**
Experiment 4					
Restudy	0.854	0.993		5.62	.982
Retrieval Practice with Feedback	0.851	0.727	0.48	0.07	.999
Retrieval Practice without Feedback	0.782	0.751		0.77	.996
Experiment 5					
Restudy	0.743	0.155		2.49	.976
Retrieval Practice without Feedback	0.677	0.097	28.13	0.70	.983
No Practice	0.492	0.191		0.26	.997

delay (see Table 10). Fitting these data points with power functions of time would have produced unreliable parameter estimations, and, for the retrieval practice conditions without feedback, would have been confounded with floor effects.[3]

A numerical examination of mean final recall rates for successfully and unsuccessfully practiced items in the three retrieval practice conditions of Experiments 4 and 5 showed overall reduced memory performance of unsuccessfully practiced items. As an approximation of forgetting rates, relative loss rates were computed separately for successfully and unsuccessfully practiced items, by comparing recall performance at the shortest and at the longest delay interval in each condition (see Table 10). Relative losses were consistently higher for unsuccessfully practiced items, indicating a more drastic decline in memory performance across the retention interval compared to successfully practiced items.

---

[3] Accordingly, only final recall rates for successfully practiced items were fit with power functions, using the same fitting procedures as described above, with the previously estimated parameter c for arithmetically averaged data. For Experiment 4, mean final recall rates of the successfully practiced items were well described by the power function of time, both with feedback, $a = 0.942$, $b = 0.614$, $\chi^2(2) = 0.84$, $r^2 = .997$, and without, $a = 0.939$, $b = 0.646$, $\chi^2(2) = 2.10$, $r^2 = .987$. The same was true for Experiment 5, where the final recall rates of the successfully practiced items were again well described by the power function, $a = 0.846$, $b = 0.101$, $\chi^2(2) = 1.98$, $r^2 = .966$.

Table 10

Mean Final Recall for Successfully and Unsuccessfully Practiced Items with Relative Loss Rates

Experiment 4	**3 min**	**1 day**	**3 days**	**7 days**	**relative loss**
Retrieval Practice with Feedback					
Successfully Practiced	.94	.77	.56	.38	.60
Unsuccessfully Practiced	.59	.36	.18	.10	.83
Retrieval Practice without Feedback					
Successfully Practiced	.94	.73	.58	.35	.63
Unsuccessfully Practiced	.22	.07	.09	.01	.95
Experiment 5	**3 min**	**1 day**	**2 days**	**3 days**	**relative loss**
Retrieval Practice without Feedback					
Successfully Practiced	.84	.62	.53	.56	.33
Unsuccessfully Practiced	.14	.02	.09	.04	.71

Note. Relative loss was computed as retention at time 1 minus retention at time 4, relative to the amount retained at time 1.

3.7 Discussion of Experiments 4 & 5

Across Experiments 4 and 5, recall rates in all practice conditions were well described with the power function of time, in line with previous studies (e.g., Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991, 1997). Additionally, several noteworthy findings emerged by comparing forgetting rate parameter b between conditions: After retrieval practice with feedback, the rate of forgetting was lower compared to the restudy condition (Exp. 4) and this was also the case for retrieval practice without feedback relative to restudy (Exp. 4 and 5). At the same time, forgetting rates did not differ between retrieval practice with feedback and retrieval practice without feedback (Exp. 4). Finally, restudy did not significantly reduce the rate of forgetting compared with a no-practice condition, even though forgetting was numerically higher in the latter condition (Exp. 5).

To start, the present experiments add to those of Carpenter et al. (2008) and Siler and Benjamin (2020) who had previously used power function analysis to investigate the impact of retrieval practice with feedback vs. restudy on relative time-dependent forgetting. The results replicate those of Experiments 2 and 3 of Carpenter et al.

(2008), who also reported a significantly reduced rate of relative forgetting after retrieval practice with feedback in these cases. At least numerically, the results are also in line with the findings of Experiment 1 by Carpenter et al. and with the recognition data reported in Experiment 2 of Siler and Benjamin (2020), that did not find a significant reduction in forgetting but reported numerically lower rates of forgetting after retrieval practice with feedback. Taken together, these experiments demonstrate at least a numerical reduction in forgetting rate parameter b for different kinds of materials – unrelated paired associates in the present experiment, vocabulary pairs and obscure facts in Carpenter et al. (2008) and visual stimuli and their category labels in Siler and Benjamin (2020) –, as well as for different final test formats – cued recall in the present experiment and in Carpenter et al. and item recognition in Siler and Benjamin –, indicating that this finding might generalize to many instances of retrieval practice with feedback.

Experiments 4 and 5 of this thesis also extend the two previous studies that used power function analysis (Carpenter et al., 2008; Siler & Benjamin, 2020) by demonstrating that the findings for retrieval practice without feedback mirror those for retrieval practice with feedback: For both cases of retrieval practice, the rate of time-dependent forgetting was reduced in comparison to restudy. Additionally, rates of forgetting did not differ significantly between the two retrieval practice conditions. Even though final recall was higher after retrieval practice with compared to without feedback (see also Abel et al., 2019; Butler et al., 2008), evidenced by a higher parameter a, this increase apparently did not translate into a decrease in forgetting rate. The observed decrease in time-dependent forgetting after retrieval practice is in line with the implications drawn from previous ANOVA-based findings (e.g. Roediger & Karpicke, 2006b; Toppino & Cohen, 2009; Wheeler et al., 2003) that were based on differences in absolute forgetting. The parallel behavior of retrieval practice with and without feedback in regards to time-dependent forgetting is at odds with some of the earlier ANOVA-based findings (e.g. Abel et al., 2019; Kornell et al., 2011) that instead reported parallels between restudy and retrieval practice with feedback across longer retention intervals. At least in terms of absolute drops in performance from 3 min to 7 days, the present results are therefore also at odds with the assumptions of the bifurcation framework (Halamish & Bjork, 2011; Kornell et al., 2011), according to which test-delay interactions should be reduced or even eliminated for retrieval practice with feedback. Instead, a significant test-delay interaction emerged for retrieval practice with feedback compared to restudy in Experiment 4 (see also Abel & Bäuml, 2020; Abel & Roediger, 2018; Mulligan &

Peterson, 2015). However, success rates during retrieval practice were relatively high in the present Experiments 4 and 5. Differences between the two retrieval practice conditions regarding time-dependent forgetting might emerge under conditions that leave more room for the effects for feedback, for example by increasing initial retrieval difficulty.

In contrast with the retrieval practice conditions, restudy had no significant retarding effect on time-dependent forgetting compared to a no-practice control condition (Exp. 5). This result is at odds with the findings of Wixted (2022a) who reported reduced forgetting for conditions with high degree of learning. As in Experiment 5, degree of learning was manipulated by varying the number of (re)study cycles. In line with Wixted's findings, restudy did however lead to a lower numerical rate of forgetting. It is possible that an increased sample size or a bigger difference in the number of study cycles would have been necessary to yield a significant difference in forgetting. Some of the studies analyzed in Wixted (2022a) did indeed contrast conditions that differed to a greater extent in their numbers of study cycles, for example one vs. three cycles (Slamecka & McElree, 1983, Exp. 1; see also Krueger, 1929; Rivera-Lares et al., 2022), while in others, the same number of cycles was used as in the present Experiment 5 (two vs. three cycles; Slamecka & McElree, 1983, Exp. 2). Sample sizes for each data point in these studies seemed to be comparable to or even smaller than the one in the present experiment. At the same time it should be noted that Wixted (2022a) did not conduct any statistical tests to examine whether or not the numerical differences in forgetting rate he reported for high vs. low degree of learning were significant, leaving the correspondence between the present results and his findings open, as well as the issue of whether restudy itself can reduce time-dependent forgetting. It is well possible that restudy exerts a more subtle but nonetheless relevant influence on time-dependent forgetting than retrieval practice does, and this should be examined in future research.

The present findings of significantly reduced time-dependent forgetting after retrieval practice add to the observation of stabilizing effects of practicing retrieval on memory performance, such as reduced retroactive interference (Potts & Shanks, 2012), interlist interference (Kliegl & Bäuml, 2016) and list-method directed forgetting (Abel & Bäuml, 2016). While these observations are vaguely in line with the more prominent theoretical ideas, semantic elaboration (Carpenter, 2009, 2011) and use of contextual cues (Karpicke et al., 2014), as both assume a greater reliance on semantic or contextual cues in the face of increased final test difficulty, they do not favor one of the accounts over the other. Reduced time-dependent forgetting after retrieval practice might also be explained with other theoretical proposals. For example, retrieval practice might be assumed to induce memory consolidation in itself (e.g., Antony et al. 2017), thereby attenuating time-dependent forgetting. According to Wixted (2022a), the rate of forgetting might be slowed for information of subjective relevance, for which retrieval practice might be an indicator. Following J. R. Anderson and Schooler's (1991) investigation of the effects of previous use of items on accessibility (see also R. B. Anderson et al., 1997; Cowan et al., 2021), it could also be argued that information that has been retrieved before might have a higher chance of being retrieved in the future and might therefore be worth retaining.

Common to many ideas regarding the origin of retrieval practice effects is the assumption of some sort of change to the underlying memory trace that emerges only through retrieval practice but not through restudy, even though the exact mechanisms differ: Assumptions include additional encoding of a test memory trace (Rickard & Pan, 2018), an elaboration of the memory trace by integration of semantic mediators (Carpenter, 2009, 2011), encoding of contextual elements (Karpicke et al., 2014) or differential processing that makes retrieval practiced memory traces more procedural in nature (Racsmány et al., 2018). Perhaps it is this common theme of re-encoding processes, in whatever shape, that form the central difference between retrieval practice and retrieval-free practice conditions and that lead to differences in time-dependent forgetting. Some sort of structural changes to a memory trace might be necessary to change the course of time-dependent forgetting, while a mere reactivation of the initially encoded memory trace, i.e. restudy, might be insufficient for such a change, or induce a smaller change in forgetting.

Previous work has demonstrated that the beneficial effects of feedback mostly pertain to items that were not successfully retrieved during practice (Butler et al., 2008; Pashler et al., 2005) and that those effects can vary for example depending on type (Pashler et al., 2005) and timing of feedback (Butler et al., 2007; Butler & Roediger, 2008). Another set of studies has shown that, at least under some conditions, feedback during retrieval practice can reverse initial retrieval practice benefits over restudy into restudy benefits (Pastötter & Bäuml, 2016; Storm et al., 2014; Racsmány et al., 2020). These findings indicate that success rate during retrieval practice as well as feedback might have moderating effects on long-term retention. In the present Experiments 4 and 5, success rates during retrieval practice were relatively high in all conditions, leaving only little room for positive effects of feedback on overall recall performance. Following from this, only few unsuccessfully practiced items were recalled correctly on the final test, suggesting that retrieval practice mostly facilitated recall of successfully practiced items. In studies that produce lower success rates during retrieval practice and thus leave more room for the effects of feedback on final recall performance, forgetting rates might however differ for successfully and unsuccessfully practiced items. For successfully practiced items, forgetting rates might be lower than for the whole set of items, and at least in Experiment 4 (where initial success rates were slightly lower than in Exp. 5), forgetting rates were indeed numerically lower when only successfully practiced items were analyzed. For unsuccessfully practiced items forgetting rates might generally be higher, and an examination of relative loss rates for the present experiments supported this idea, but differences might emerge depending on whether feedback is provided or not. Unsuccessfully practiced items that are followed by feedback might mimic the rate of forgetting observable for restudied items, while items that are unsuccessfully practiced in the absence of feedback might be forgotten at an even higher rate. Future studies may examine the exact influence of feedback and initial retrieval success on forgetting rates in more detail, but also of other factors such as type or timing of feedback.

The high success rates during retrieval practice in the present experiments can be attributed to the relatively easy test format during retrieval practice, during which the cue item of each paired associate and the target item's first two initial letters served as retrieval cues. Previous work suggests that the effects of retrieval practice are generally more pronounced in cases of high initial retrieval difficulty (Carpenter & DeLosh, 2006; Glover, 1989; see also R. A. Bjork, 1994). The test format was chosen to avoid floor effects for the long retention intervals, but probably decreased the size of the observed retrieval practice effects (Rowland, 2014). More difficult retrieval practice tasks could not only positively influence the size of the retrieval practice effect but may potentially also reduce forgetting rates relative to restudy even further. The impact of initial retrieval difficulty on forgetting rates should be examined in future research and could be compared directly.

In contrast with the full within-subject designs of Carpenter et al. (2008) and Siler and Benjamin (2020), the present Experiments 4 and 5 manipulated delay as a between-subjects factor. There are potential advantages and disadvantages to both designs: A within-subject manipulation allows for the estimation of individual power function parameters for each participant, making averaging across subjects superfluous and circumventing potential averaging artifacts (e.g. Estes, 1956). However, individual fits can be poor, a problem that is reported in both Carpenter et al. (2008) and Siler and

Benjamin (2020), were some subjects yielded ”extreme“ or ”uninterpretable“ results for the parameter estimates. Such outliers could bias the overall results. Additionally, recalling a subset of studied items can improve memory performance of other items at later retention intervals (see Bäuml & Trißl, 2022), which could potentially obscure the actual forgetting rates by underestimating them. A between-subjects manipulation of retention interval can avoid such a potential interdependence of recall events and also outliers in individual fits. However, this choice of design makes averaging across participants necessary to fit the recall rates with the function, which could result in averaging artifacts. This problem was addressed by performing an additional analysis with geometrically averaged data that produced a similar pattern of results as that performed with arithmetically averaged data, which can alleviate the concerns about artifacts. As no design is obviously superior, it would be ideal if both designs generally produced similar patterns of results. Taking the present results and those of Carpenter et al. (2008) together, the findings do point in such a direction. Still, more research is needed that addresses the relevance of choice of design more directly in investigations of time-dependent forgetting after retrieval practice.

All in all, the results of Experiments 4 and 5 demonstrate that retrieval practice can slow down time-dependent forgetting, regardless of whether feedback is presented or not and thereby adds to the existing literature on retrieval practice effects. Moving forward, these results need to be replicated, especially as the findings using power function analysis so far have been mixed (Carpenter et al., 2008; Siler & Benjamin, 2020), and possible moderating factors such as choice of design, test formats and materials need to be examined. Also, more work into the exact circumstances under which reduced time-dependent forgetting after retrieval practice can be observed might help provide more information on the mechanisms responsible for retrieval practice effects. Another worthwhile field for future research might be related practice strategies that also involve more active processing of materials, such as generation from semantic memory (Jacoby, 1978; Mulligan & Peterson, 2015; Slamecka & Graf, 1978) or semantic elaboration (Karpicke & Blunt, 2011; Karpicke & Smith, 2012). If previous use can act as a relevance-signal and indicate the probability of future use (J. R. Anderson & Schooler, 1991; see also Cowan et al., 2021), it would be interesting to examine what such previous use needs to look like. The results from Experiment 4 and 5 additionally show that processes that take place after initial study can in fact also influence the rate of forgetting, but such effects might depend on very specific characteristics of those processes, for example the involvement of encoding-like operations.

Chapter 4

General Discussion

4.1 Summary of Findings

Despite the long history of memory research, results on whether relative time-dependent forgetting varies systematically between different experimental manipulations have been scarce. In the present thesis, five experiments were therefore conducted to examine this issue for directed forgetting and for retrieval practice effects in an effort to extend existing research on factors that influence how quickly studied materials are forgotten. In all conditions, retention across several days was well described by power functions of time that were estimated using maximum likelihood methods. Experiments 1-3 were concerned with directed forgetting and investigated the influence of a cue to forget individual items (IMDF; Exp. 1 & 2) or a whole list of items (LMDF; Exp. 3) compared to a cue to remember a different set of materials. In the two IMDF-experiments, the rate of forgetting was significantly increased for items that were cued to be forgotten (TBF) compared to items that were cued to be remembered (TBR), both for free recall (Exp. 1) and for item recognition (Exp. 2). At the same time, no significant difference in forgetting rate arose between the two cuing conditions in an LMDF-experiment (Exp. 3), and this was true both for the first list that served as TBR or TBF information, as well as for the second list that all participants were asked to remember. The results of Experiment 1-3 show that while a cue to forget can have lasting effects on retention, as DF was present after one week in all three experiments, it does not necessarily increase the rate of forgetting for TBF materials, or at least not to a significant extent.

In Experiments 4 and 5, different kinds of practice of materials that had been studied twice before were compared in their effects on relative time-dependent forgetting. Retrieval practice (with and without feedback) was found to decrease the rate of forgetting compared to a restudy condition, and importantly, the two types of retrieval practice did not differ significantly in their rates of forgetting. This replicates and extends previous findings on the effects of retrieval practice on relative forgetting (Carpenter et al., 2008; Siler & Benjamin, 2020; see also Bäuml & Trißl, 2022), and corroborates the idea that retrieval practice can induce a protective effect against time-dependent forgetting. In contrast, restudy itself did not significantly reduce forgetting in comparison to a no-practice control condition, even though a sizable numerical difference was observed, indicating that such a protective effect might be exclusive to retrieval practice, or that restudy has a comparatively smaller effect on time-dependent forgetting.

Taken together, the findings add to the few existing studies on variations in relative time-dependent forgetting (Bäuml & Trißl, 2022; Carpenter et al., 2008; Siler & Benjamin, 2020; Wixted, 2022a) which mostly investigated the effects of additional practice after study on forgetting. In line with these studies, the beneficial effect of retrieval practice on time-dependent forgetting seems to be relatively well-supported so far. New is the finding that intentionally trying to forget materials can also have an influence on the rate of forgetting, in that, at least under some circumstances, TBF information seems to be forgotten at an increased rate.

The results of all five experiments also emphasize the importance of investigating long-term effects of experimental manipulations, as performance differences between conditions at a short delay do not always translate into a difference in relative forgetting. Instead, very different patterns of results can emerge: Both for TBR and TBF information in IMDF (Exp. 1 & 2) as well as for retrieval-practice-without-feedback and restudy (Exp. 4), memory performance differed significantly after 3 min and additionally, rates of relative forgetting also differed. However, rate of forgetting was higher for TBF in the one case, which initially showed lower retention, and for restudy in the other case, which initially showed higher retention. Conversely, for both lists in LMDF (Exp. 3), and for the restudy and no-practice conditions (Exp. 5), memory performance differed significantly after 3 min, but no significantly different rate of forgetting arose. Finally, for retrieval-practice-with-feedback and restudy, no difference in memory performance was present after 3 min, but the two conditions differed significantly in their rates of forgetting. Except for the latter case, the wide-spread approach in the literature to look at whether two conditions differ in performance at a short and at a long delay (e.g. Abel & Bäuml, 2017; MacLeod, 1975; MacLeod & Macrae, 2001;

Toppino & Cohen, 2009) in order to investigate long-term effects of experimental manipulations would not have sufficed to uncover the patterns of results illustrated above. The methodical set-up used in this thesis can accordingly unveil additional information on how forgetting proceeds in different conditions.

4.2 Theoretical Implications for Variations in Time-dependent Forgetting

As the present experiments were not designed to test predictions of theoretical accounts of either of the experimental paradigms examined in this thesis, the results provide no clear support for any of the mentioned accounts (see also Sections 2.9 and 3.7). They do, however, add to the results successful long-term accounts of IMDF, LMDF and retrieval practice effects need to be able to explain. Aside from illustrating that DF effects in both standard paradigms can be long-lasting, the comparison of forgetting rates might indicate that further processing of TBR vs. TBF information diverges in IMDF but is similar in LMDF. This might mean that the difference between both kinds of information is a quantitative one in LMDF, with TBR and TBF information differing mostly regarding how much they are processed, resulting e.g. in different amounts of memory strength, but a qualitative one in IMDF, with a difference in how both kinds of information are processed, resulting e.g. in structural differences. Alternatively, it could mean that with both methods, the difference between TBR and TBF information is a quantitative one, but that this difference is simply greater in IMDF than in LMDF. For retrieval practice effects, the current findings emphasize the need for an explanation why retrieval practice but not restudy significantly changes the rate of long-term time-dependent forgetting of previously studied materials, in that memory traces strengthened by retrieval practice but not by restudy seem to be less vulnerable to forgetting. Additionally, theoretical accounts should incorporate the effects of feedback on retrieval practice, as the present results found no difference in forgetting rates regardless of whether feedback was absent or present during retrieval practice.

Overall, the results of the present experiments indicate clearly that, at least under certain conditions, the rate of time-dependent forgetting can vary between different memory conditions. One potential process that might progress differently depending on condition could be memory consolidation, which is assumed to transform newly encoded information into a stable long-term memory representation and protects them from decay (Dudai, 2004; Dudai et al., 2015; Müller & Pilzecker, 1900). Newer work has argued for a certain selectivity of consolidation, with prioritization of more salient or future-relevant information (Cowan et al., 2021; Stickgold & Walker, 2013), which could result in differences in the rate of forgetting over time. Such variations might be captured by the power function of time's characteristic of a decreasing proportional rate of forgetting, which has been suggested to be a manifestation of memory consolidation (Wixted, 2004a, 2004b). If so, power function forgetting rates might indicate the amount or pervasiveness of consolidation the material has received. Applied to the current findings, this would imply an attenuation of time-dependent forgetting through consolidation primarily for (stronger or more relevant) TBR information in IMDF (Exp. 1 and 2), as well as for retrieval practiced information (Exp. 4 and 5), but less differences in consolidation for TBR and TBF information in LMDF (Exp. 3), as well as for restudy and no-practice control (Exp. 5). It is also possible that encoded information is subjected not only to consolidation processes but to active erasure processes as well (e.g. Davis & Zhong, 2017; Hardt et al., 2013), with varying contributions of both types of processes depending on the strength or assumed relevance of information. This balance could influence how quickly memories are perceived to be forgotten. Active forgetting itself might also vary in its extent depending e.g. on the kind or amount of further processing that memory traces receive (Zhang et al., 2016). Regarding the present findings, TBR in IMDF as well as retrieval practiced information might undergo more consolidation vs. less erasure, while TBF in IMDF as well as restudied or not-practiced information might experience less consolidation vs. more erasure, resulting overall in enhanced time-dependent forgetting for these kinds of information, while again, the balance between both processes might be more similar for TBR and TBF information in LMDF as well as for restudy and no-practice control, resulting in similar rates of time-dependent forgetting.

Such selectivity would hinge on specific characteristics of the encoded memories that signify relevance, or more generally, whether a memory is worth retaining (see also Cowan et al., 2021). In all probability, this would not be a matter of either/or but rather a graded characterization, resulting in many different extents of consolidation or erasure. Such a view would fit in with the present results where studied materials seemed to be forgotten at relatively moderate rates overall, instead of either very quickly or very slowly. It is possible that there are many contributing factors that decide the long-term fate of a memory trace, that some of them are more effective than others and that they can maybe even cancel each other out. They could also work at different time frames, as for a lot of information that we encounter in a given day it will be unclear at the moment of encoding whether this information will be relevant at a later point and thus worth retaining or not (Cowan et al., 2021; Hardt et al., 2013). Cases like that of IMDF, where the importance of an item becomes apparent at a very early stage of processing probably make up a minority of the information we consciously encode. To balance processes such as consolidation and active forgetting (Davis & Zhong, 2017) adaptively, additional clues that surface at later stages after initial encoding need to be made use of. However, such factors should adhere to some common rules if they act on memory traces in a similar way, and could for example all be guided by the probability of future use of a given memory trace (J. R. Anderson & Schooler, 1991; R. B. Anderson et al., 1997; see also Cowan et al., 2021).

Largely unclear as of yet is which factors might determine differential processing of studied materials, and research into differential consolidation so far has mostly concentrated on affective and motivational factors (see Cowan et al., 2021; Stickgold & Walker, 2013). Some additional insights into relevance-signals can come from the current results. In the present experiments, the rate of relative forgetting was significantly reduced in the case of TBR items in Experiments 1 and 2, and in the case of retrieval practiced items (with and without feedback) in Experiments 4 and 5. This could either be due to the same factor in both cases or be due to different factors that both contribute to the rate of forgetting. As arguably, initial encoding differs between TBR and TBF information in IMDF (but not in LMDF), and many explanations of retrieval practice benefits over restudy assume some sort of additional encoding that is exclusive to retrieval (e.g. of semantic associations or contextual information; Carpenter, 2009; Karpicke et al., 2014), this could indicate the importance of encoding, or in the case of retrieval practice, re-encoding processes, for the rate of forgetting. Better encoded traces might be expected to be used again in the future and this could e.g. involve the additional encoding of future-relevant tags or the establishment of especially elaborate memory traces that make those traces less prone to erasure. For retrieval practice such a view could be aligned with the process of reconsolidation (e.g. Hardt et al., 2013; Lee et al., 2017), a process by which already consolidated memories become fragile again when reactivated and need to be re-stabilized (e.g. Dudai, 2004; Nader & Hardt, 2009; Lee et al., 2017), and which can lead both to degradation and enhancement of these memories. Such memory adaptation could be induced by retrieval of previously encoded memories (e.g. Hardt et al., 2013) and could also influence the observed rate of forgetting.

It is also possible that it is not the involvement of (re)encoding processes but the generation of sufficiently strong memory traces that explain the cases of reduced forgetting in the present experiments. Following Wixted (2022a), subjective meaningfulness could indicate the degree of future relevance, thereby influencing further processing. Based on his findings of numerically lower forgetting rates for materials with a higher degree of learning, he proposed that degree of learning might in turn be an indicator of meaningfulness. If a selective rehearsal explanation of IMDF is adopted (B. H. Basden et al., 1993; R. A. Bjork, 1970), which could result in a higher degree of learning for TBR information, the finding of reduced forgetting of TBR items in Experiments 1 and 2 would be in line with this idea. Following the bifurcation framework (Halamish & Bjork, 2011; Kornell et al., 2011), retrieval practice also results in a higher memory strength of (successfully retrieved) items than restudy does, and again, forgetting was reduced in those conditions (Exp. 4 & 5). In contrast, if following a selective rehearsal explanation of LMDF (R. A. Bjork, 1970; Sheard & MacLeod, 2005), reduced forgetting might also have been expected for remember-cued participants in List 1 and for forget-cued participants in List 2, but in both cases, very similar rates of forgetting were observed. This might mean that selective rehearsal and degree of learning should not be equated, or that the strength differences between conditions in LMDF were too small to translate into observable differences in forgetting. Similarly, the finding of no significantly reduced forgetting after restudy compared with no-practice control in Experiment 5 seems at odds with the findings of Wixted's (2022a) re-analyis of degree of learning studies, where higher amounts of study cycles were associated with lower rates of forgetting. Again, it is possible that the attenuating effect of restudy in Experiment 5 was too subtle to gain significance and a substantial increase in statistical power in future studies investigating degree of learning or memory strength could shed more light on this issue.

Other factors could also be used to explain the findings for directed forgetting and retrieval practice effects separately. For example, prior use could indicate future relevance, and when taking into account the findings from Experiments 4 and 5, prior use in the form of retrieval might confer a bigger change in processing than prior use in the form of restudy, which is reminiscent of the bifurcation framework of retrieval practice effects (Halamish & Bjork, 2011; Kornell et al., 2011), where retrieval practice is assumed to induce bigger increases in memory strength than restudy. For DF, test expectancy could also have impacted results (see Stickgold & Walker, 2013; Saletin et al., 2011; Wilhelm et al., 2011): In the case of IMDF (Exp. 1 & 2), where forgetting rates were smaller for TBR than for TBF information, it is possible that this was due not to reduced forgetting of TBR information but due to increased forgetting of TBF information. TBF information might have been tagged for active erasure at encoding, and similar tags in LMDF could have been less effective than the item-specific tags in IMDF. Alternatively, TBR information could have been tagged as future-relevant for preferred consolidation in both methods (Cowan et al., 2021), and again, those tags could have been less effective in LMDF. In future research it could accordingly be worthwhile to include a neutral or uncued condition in DF paradigms (see e.g. Schindler & Kissler, 2018; Zwissler et al., 2015) to investigate how forgetting would progress for this kind of information.

In returning to some of the factors described in the introduction, interference and context-dependency could also have played a role for the present results. Retrieval practice has been shown to reduce susceptibility to interference (e.g. Kliegl & Bäuml, 2016; Potts & Shanks, 2012; see also Karpicke et al., 2014), and this could have positively influenced the rate of forgetting for these conditions. However, if a context-change (Sahakyan & Kelley, 2002) or inhibition explanation (Geiselman et al., 1983) of LMDF is adopted, this should have resulted in reduced proactive interference for forget-cued participants during List 2, and here no (significant or numerical) reduction in rate of forgetting could be observed. Retrieval practice could also have reduced the dependency on study-context cues (Karpicke et al., 2014), which could have facilitated retrieval at longer delays. As study-context and retrieval-practice-context were very similar in the present experiments, it is also possible that this factor had only a small effect on the forgetting rates, if any.

The rate of forgetting could also be influenced by factors not at play in the current experiments. For example, Hardt et al. (2013) proposed that stress or emotional responses could induce differential processing of some memory traces but not others. Emotional content seems to be consolidated differently than neutral content and motivational dynamics could also play a role (see e.g. Cowan et al., 2021; Stickgold & Walker, 2013). It is also possible that pre-existing associations or familiarity with the studied materials impact forgetting. Another question altogether is whether all of these possible factors affect one main underlying attribute of the memory trace that signals for enhanced consolidation (or decreased erasure) or whether there are different pathways for such processing. If it boiled down to differences in memory strength, stronger traces could for example be intrinsically more resistant to early trace degradation, which would make them available for systems consolidation (e.g. Dudai, 2004), be more firmly anchored in extensive associative networks, or be more elaborate in general, by consisting of additional contextual, emotional or semantic information. Overall, many questions are still open regarding how and why some memories seem to be forgotten at faster rates than others.

4.3 Methodological Implications

4.3.1 Using Curve Fitting to Analyze Time-dependent Forgetting

Deviating from most of the existing literature that examines the role of delay for different experimental manipulations (e.g. Abel & Bäuml, 2017; MacLeod & Macrae, 2001; Roediger & Karpicke, 2006b), this thesis used not only ANOVAs to examine variations in forgetting, but also curve fitting techniques in the form of power function analysis, in the hope of gleaning additional information on how time-dependent forgetting proceeds for different kinds of experimental manipulations. The use of curve fitting and especially power function analysis has several key advantages: It allows for the comparison of relative forgetting, which has been argued to be of greater theoretical interest due to its ties to various memory models (Wixted, 2022a) and can describe the actual form of the forgetting curve more closely (Wixted, 2004a, 2004b). Which functions provide good fits and how those functions differ between conditions might accordingly provide us with a more complete picture of the time course of forgetting.

For the present analyses, power functions of times were estimated, but the last decades of memory research have seen the use of many other functions as well. Some authors for example are proponents of the exponential function (White, 2001), or exponential-power function (Rubin & Wenzel, 1996; Wickelgren, 1972), while more recently, Fisher and Radvansky (2018) suggested that long-term forgetting is described best by a linear function. The logarithmic function has also been found to describe various data sets well (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991). Out of the many candidate functions, the power function of time has the advantages of being able to fit very different kinds of data (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991), having parameters that are easy to interpret and being in line with prominent theoretical assumptions about forgetting (Wixted, 2004a, 2004b), making this function a, but by no means the only, practicable choice. It has also been argued that it would be more profitable to examine how experimental manipulations influence forgetting (Kahana & Adler, 2012) instead of focusing on the exact form of the function used (Brown et al., 2007). Nonetheless, it is important to know whether systematic differences that can be found between conditions hinge on the use of a specific function.

In the case of power functions, the version of the function used during curve fitting should also be considered. For the present set of experiments, all conditions were fit with the three-parameter version of the power function, $r(t) = a(1 + ct)^{-b}$, as reported above, but also with the two-parameter version, $r(t) = at^{-b}$, and with an adjusted version of the three-parameter version, $r(t) = a(1+t)^{-b}$, where $c = 1$. While the three-parameter version was, according to the maximum likelihood methods, able to explain the data of all 14 conditions well, the two other versions failed for some of the conditions (3 conditions in the case of the two-parameter version, and 4 conditions in the case of the adjusted three-parameter version). This seemed to be mainly due to the shape of the initial drop in performance soon after study, which varied across conditions (see e.g. the difference in parameter c estimates in Exp. 4 & 5). It is important to note that this failure to describe some conditions was based on the results of the approximate χ^2-test, but the r^2 values for some of these conditions were not much below some of the r^2 values of conditions the functions were able to describe well[4], illustrating that r^2 might not always be a suitable criterion for goodness-of-fit. The three-parameter version used in the present experiments not only was able to explain all 14 conditions well, but also overcomes a characteristic of the two-parameter version, being undefined for $t = 0$, which has been viewed as problematic (Wickens, 1998; White, 2001). Instead, for $t = 0$, the three-parameter version takes on the value of a. Introducing a third parameter is however at odds with the principle of parsimony (e.g. Lazar, 2010), but by pre-estimating c for all conditions within one experiment, the three-parameter version becomes a de-facto two-parameter version with c as a constant. As c influences the shape of the initial drop in performance soon after study, this version of the power function is better able to accommodate differences between data sets.

Another methodical question is that of how curve fitting is carried out and which statistics are reported. Using maximum likelihood methods as in this thesis allows for statistical tests for goodness-of-fit of the estimated models and also allows for a direct comparison of models and their parameters. A lot of the previous literature

---

[4] Overall, r^2 values for rejected models ranged from .77 to .84 for the two-parameter version, and from .76 to .93 for the adjusted three-parameter version. For several accepted models, and for all three versions of the power function models, some r^2 values were similar in size, ranging from .83 to .93.

that employed curve fitting techniques to describe time-dependent forgetting relied on estimating r^2 to illustrate the amount of explained variance (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991) as well as on describing numerical differences between parameter estimates (Wixted, 2022a). As demonstrated above, relying on criteria such as r^2 for model or function selection might sometimes be misleading, and the comparison of forgetting rates between restudy and no-practice control in Experiment 5 shows that a sizable numerical difference need not always be significant.

A more specific issue is that of how to best test for differences in rate of forgetting parameter b. In this thesis, parameter comparisons between conditions were carried out by allowing one parameter to vary and at the same time restricting the parameter of interest to be constant for both conditions. However, as one parameter is free, this leaves room for the flexibility of the power function and might reduce the explanatory power of the individual parameter comparisons. Another way, especially for the comparison of b, could be to fix a at the values estimated in the original power function fits for both conditions, with b restricted to be the same for both conditions.

4.3.2 Within- vs. Between-subject Manipulation of Retention Interval

Some research into time-dependent forgetting has favored a within-subject manipulation of retention interval that allows for an estimation of individual forgetting curves for each participant (Averell & Heathcote, 2011; Carpenter et al., 2008; Rubin et al., 1999; Siler & Benjamin, 2020). Arguably, this process could lead to a cleaner estimation of the underlying processes, as inter-individual differences can be big in memory research and accordingly, rates of forgetting might differ substantially between subjects (Carpenter et al., 2008; Wixted & Ebbesen, 1997). Individual subject data also circumvents the concern of averaging artifacts that can arise for between-subjects data and that obscure the form of the underlying individual forgetting curves (Estes, 1956; Sidman, 1952). For example, it has been demonstrated that (arithmetically) averaging exponential functions can lead to an aggregate curve that mimics the form of a power function (R. B. Anderson & Tweney, 1997; Murre & Chessa, 2011). This seems to especially be the case for simulations where individual curves vary wildly in their estimated slope parameters (R. B. Anderson & Tweney, 1997; Wixted & Ebbesen, 1997).

However, a within-subject manipulation of retention interval may not always be possible or feasible due to specific aspects of the manipulation of interest. A within-subject manipulation also requires a relatively large stimulus array to generate enough data points at each retention interval for each subject (e.g. 60 items in each experiment of Carpenter et al., 2008; 80 bird images in Exp. 2 of Siler & Benjamin, 2020), as well as individually cuing studied items during recall, making the use of free recall as test format impossible. As an example of a case where a within-subject manipulation would have created undesirable effects due to the manipulation of interest, the present experiments on directed forgetting (Exp. 1-3) can be considered: Testing subjects repeatedly on portions of both TBR and TBF information would have made it necessary to debrief them regarding the actual relevance of TBF information at the first retention interval. This would very likely have changed further processing of TBF information, for example via increased rehearsal attempts for these items, and thus the rate of time-dependent forgetting – the measure of interest – observable for TBF information, while further processing of TBR information would likely not have been influenced, thereby skewing any implications drawn from the findings.

In the case of retrieval practice effects, both design-choices are possible, making it possible to compare findings from both and to evaluate the impact of the advantages and disadvantages inherent to within- and between-subject manipulations. The replication of previous results of within-subject designs (Carpenter et al., 2008; Siler & Benjamin, 2020) in the present Experiment 4 and the fact that the pattern of results was preserved when geometric instead of arithmetic averaging was used (R. B. Anderson & Tweney, 1997; Estes, 1956; Wixted, 2022a; Wixted & Ebbesen, 1997) indicate that, at least in this field of interest, both design choices seem to be valid. Should this be replicated in future work and should such parallel results also be found for other fields where both design-choices are feasible, this could open up the application of curve fitting to other research questions where within-subject manipulations of retention interval are undesirable or simply not possible.

4.4 Conclusions and Future Directions

4.4.1 Methodological Limitations

Several issues might be seen as limitations to the implications of the present results, such as the fact that data collection took place via Zoom and not in a more controlled lab environment. This, however, probably led to study and recall conditions that were closer to real-life conditions. At the same time, some of these issues could be interesting avenues for future research. One worthwhile endeavor for future investigations into variations of time-dependent forgetting might be extending the retention intervals to include longer delays than the one week examined in the present experiments, as the form and the rate of longer-term forgetting might very well change at very long delays. This important issue has recently been raised and described by Fisher and Radvansky (2018; see also Radvansky et al., 2022).

When examining long-term forgetting of studied materials, it is also important to consider the avoidance of floor effects as those might skew and underestimate the observable rate of forgetting. A concern for floor effects informed several methodical decisions in the experiments presented here, such as the number of study cycles and relatively easy initial test format in Experiments 4 and 5, and the duration of stimulus presentation in Experiments 1-3. Nonetheless, performance for TBF information was near or at floor in Experiment 1, and due to the peculiarities of IMDF experiments, it is not immediately clear how such a finding might be circumvented in other research: Eliciting higher memory performance for TBF items without changing the central aspects of the paradigm is a difficult balance in IMDF. One possibility might be introducing an additional task, such as a pleasantness rating, to ensure better encoding of TBF items.

Another important possible limitation concerns the collected recognition data in Experiment 2: Across delay conditions, the response criterion changed, illustrated by an increase in false alarm rates, suggesting an increased uncertainty which items had been studied before. As this might be expected to be a relatively normal side effect of time-dependent forgetting, this finding raises the question of how such data should be analyzed. In order to attain a binomially distributed performance measure for the item recognition data, corrected hits were calculated for the present Experiment 2. This might not have been the ideal method, as some differences regarding the outcome of the ANOVA emerged when comparing the results for free recall (Exp. 1) and both corrected hits and d' as measures for item recognition (Exp. 2). Additionally, no individual false alarm rates for TBR vs. TBF information could be analyzed, making the corrected hits only an approximation of recognition performance. To circumvent this, a different set-up of the recognition test would have been required, for example using lures that were matched to the studied materials, or asking participants to indicate the assumed study item type (TBR or TBF) for each item classified as old during the recognition test. Simply proceeding with d' for the remaining analyses was however also not possible, as the use of maximum likelihood methods is reliant on the use of binomially distributed performance measures, and without the use of these methods, no statistical testing of different models and comparison of parameter estimates between conditions would have been possible. This could be an important issue to consider when long-term recognition data is to be examined.

4.4.2 Areas of Interest for Future Research

As relative time-dependent forgetting had not been investigated before for the case of directed forgetting and literature on retrieval practice effects had so far been mixed in this regard, the findings of the present set of experiments need to be replicated in future research. This is especially important for the non-significant differences regarding forgetting rates in LMDF (Exp. 3) and for restudy vs. no-practice (Exp. 5), as these results could also have been due to insufficient statistical power. In the case of LMDF, the findings could also have been impacted by self-chosen output order effects. As pointed out above, it would also be interesting to investigate the rate of forgetting for neutral or uncued information in DF research (e.g. Schindler & Kissler, 2018; Zwissler et al., 2015), to see how exactly forget- and remember-cues influence forgetting: The difference in forgetting rates in IMDF (Exp. 1 & 2) could be due to increased forgetting following forget-cues, decreased forgetting following remember-cues, or a mixture of both.

For any application of experimental findings that show that time-dependent forgetting can be reduced under certain conditions to every-day situations, such variations would need to be replicated for more complex materials. If results generalized to complex materials, retrieval-practice could for example be viewed as a useful method to decrease long-term forgetting of studied materials. It would also need to be established whether forgetting differs systematically, e.g. in the shape of the curve, for more complex real-world materials. Recently Radvansky et al. (2022) argued that, overall, the forgetting curve has been under-researched and that forgetting for complex materials may well be better described by a linear function than by a negatively accelerated, curvilinear function (such as the power function; see also Fisher & Radvansky, 2018).

Additionally, several other areas of interest emerge for further experiments to investigate potential systematic differences in forgetting rates. Among these should be inter-individual differences, such as aptitude to learning and study preferences, as well as developmental aspects and examinations of clinical groups. For example, there have been studies on how differences in working memory capacity relate to differences in retrieval practice effects, but so far, the results remain inconclusive (e.g. Agarwal et al., 2017; Bertilsson et al., 2021). Differences in working memory capacity could also generally influence the rate of forgetting (see Unsworth et al., 2011). DF effects have been shown to be absent in some clinical groups, e.g. with psychiatric disorders (for a review on LMDF, see Sahakyan et al., 2013) or brain lesions (Conway & Fthenaki, 2003), and to be absent or reduced in very young or in older people (Aslan & Bäuml, 2013; Aslan et al., 2010; Sego et al., 2006; Zacks et al., 1996). Given these differences in the effects of forget cues, it would be interesting to examine whether differences regarding rates of forgetting would also emerge. More generally, a useful extension of existing research on how different populations or clinical groups forget (e.g. Davis et al., 2003; Rodini et al., 2022; Squire, 1981) could be how relative time-dependent forgetting varies for these groups, as most research so far has been focused on absolute forgetting. It should however be taken care to equate encoding strength between groups, as the current research indicates that such differences could influence the rate of forgetting which in turn could obscure comparisons between groups. Following Hardt et al. (2013), who proposed that stress or emotional responses to encoded information could function as relevance signals, an investigation of if and how such post-hoc manipulations influence the rate of forgetting would be worthwhile. Other contributing factors might be the emotional valence of encoded content and motivational factors (see e.g. Cowan et al., 2021; Stickgold & Walker, 2013) as well as the degree to which participants have pre-existing knowledge of or personal connections to study materials.

It would also be interesting to examine whether and how rates of forgetting differ when other common experimental manipulations are used. It could for example be worthwhile to compare rates of forgetting for materials that are studied exclusively within one context or in several differing contexts to investigate the influence of contextual cues. Materials that are associated with several instead of just one context might have a higher chance of future use and be therefore processed as more relevant. Previous research has suggested that memories can decontextualize if associated with several different contexts (e.g. Saufley et al., 1985), which seems to improve final retention (e.g. S. M. Smith & Handy, 2014, 2016). Another field of interest could be retrieval-induced forgetting (e.g. M. C. Anderson et al., 1994) where only a subset of studied materials is subsequently retrieval practiced. As the effects of such a manipulation have been observed to dissipate over time (Abel & Bäuml, 2014; MacLeod & Macrae, 2001), unpracticed but related materials might be forgotten at a similar rate as practiced related materials, but at a lower rate than unrelated unpracticed materials. A different, but important question altogether is whether the factors that influence the rate of time-dependent forgetting differ for different kinds of memories, or whether findings on episodic memories can be generalized to semantic or procedural memories (see e.g. Dudai, 2004; Rasch et al., 2007; Tulving, 1972).

4.4.3 Conclusions

The five experiments conducted for this thesis add to the so far sparse literature concerned with whether rates of relative time-dependent forgetting can vary between conditions. This was done by examining the effects of cuing some but not all studied information as to be forgotten, as well as those of different kinds of practice after initial study. In two experiments on item-method directed forgetting, where cues are given for individual items, information that was cued to be forgotten was found to show significantly higher rates of forgetting compared to information cued to be remembered, while in an experiment on list-method directed forgetting, where cues are given for lists of items, no such difference emerged. In two experiments on the effects of practice on previously studied materials, retrieval practiced information (both with and without feedback) was forgotten at a lower rate than restudied information, but restudy did not show a significant retarding effect compared to a no-practice condition. The results could indicate that factors that are specific to item-method directed forgetting and to retrieval practice can change the rate of time-dependent forgetting, or alternatively, that different experimental manipulations differ in how pronounced their effects on forgetting are.

Taken together, the results point to systemic variations in how quickly studied information is forgotten in the context of well-known experimental paradigms, thereby extending existing research. Overall, memories might be forgotten depending on whether they are deemed to be relevant in the future or not. If the differences in forgetting reported in this thesis are due to the same factor both for directed forgetting and retrieval practice effects, such a distinction might arise based on differences in item strength, or on differences that occur at (re)encoding. Alternatively, different factors could have contributed to forgetting in both fields of research, such as prior use or a tagging as future-relevant. Investigations such as the ones presented here can further enrich our understanding of which memories can be remembered for a long time and which are forgotten more quickly, thereby getting closer to an answer to an age-old question.

Literature

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

Abel, M., & Bäuml, K.-H. T. (2013). Sleep can eliminate list-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 946–952.

Abel, M., & Bäuml, K.-H. T. (2014). The roles of delay and retroactive interference in retrieval-induced forgetting. *Memory & Cognition, 42*(1), 141-150.

Abel, M., & Bäuml, K.-H. T. (2016). Retrieval practice can eliminate list method directed forgetting. *Memory & Cognition, 44*(1), 15–23.

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

Abel, M., & Bäuml, K.-H. T. (2019). List-method directed forgetting after prolonged retention interval: Further challenges to contemporary accounts. *Journal of Memory and Language, 106*, 18–28.

Abel, M., & Bäuml, K.-H. T. (2020). Would you like to learn more? Retrieval practice plus feedback can increase motivation to keep on studying. *Cognition, 201*, Article 104316.

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

Abel, M., Kuchler, B., Meier, E., & Bäuml, K.-H. T. (2021). List-method directed forgetting: Do critical findings generalize from short to long retention intervals? *Memory & Cognition, 49*, 1677-1689.

Abel, M., & Roediger, H. L. (2017). Comparing the testing effect under blocked and mixed practice: The mnemonic benefits of retrieval practice are not affected by practice format. *Memory & Cognition, 45*(1), 81–92.

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

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659-701.

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory, 25*(6), 764-771.

Akan, M., Stanley, S. E., & Benjamin, A. S. (2018). Testing enhances memory for context. *Journal of Memory and Language, 103*, 19–27.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261–295.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*(6), 396-408.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*(4), 415-445.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063-1087.

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature, 410*(6826), 366–369.

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences, 18*(6), 279–292.

Anderson, M. C., & Hulbert, J. C. (2021). Active forgetting: Adaptation of memory by prefrontal control. *Annual Review of Psychology, 72*(1), 1–36.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory. Handbook of Perception and Cognition* (p. 237–313). Academic Press.

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., Glover, G. H., & Gabrieli, J. D. E. (2004). Neural systems underlying the suppression of unwanted memories. *Science, 303*(5655), 232–235.

Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition, 25*(5), 724-730.

Anderson, R. B., Tweney, R. D., Rivardo, M., & Duncan, S. (1997). Need probability affects retention: A direct demonstration. *Memory & Cognition, 25*(6), 867–872.

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences, 21*(8), 573–576.

Ashton, J. E., & Cairney, S. A. (2021). Future-relevant memories are not selectively strengthened during sleep. *PLOS ONE, 16*(11), Article e0258110.

Aslan, A., & Bäuml, K.-H. T. (2013). Listwise directed forgetting is present in young-old adults, but is absent in old-old adults. *Psychology and Aging, 28*(1), 213–218.

Aslan, A., Staudigl, T., Samenieh, A., & Bäuml, K.-H. T. (2010). Directed forgetting in young children: Evidence for a production deficiency. *Psychonomic Bulletin & Review, 17*(6), 784–789.

Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology, 55*(1), 25-35.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*(3), 296-308.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*(1), 1–29.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge University Press.

Basden, B. H., & Basden, D. R. (1996). Directed forgetting: Further comparisons of the item and list methods. *Memory, 4*(6), 633–654.

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: A contrast of methods and interpretations. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139–172). Lawrence Erlbaum Associates.

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.

Bäuml, K.-H. T. (2008). Inhibitory processes. In H. L. Roediger III (Ed.), *Learning and memory: A comprehensive reference: Vol. 2. Cognitive psychology of memory*, (pp. 195–220). Elsevier.

Bäuml, K.-H. T., Abel, M., & Kliegl, O. (2020). Inhibitory processes in episodic memory. In M. Eysenck & D. Groome (Eds.), *Forgetting: Explaining Memory Failure*, (pp. 125–146). Sage Publishing.

Bäuml, K.-H., Hanslmayr, S., Pastötter, B., & Klimesch, W. (2008). Oscillatory correlates of intentional updating in episodic memory. *NeuroImage, 41*(2), 596–604.

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

Bäuml, K.-H. T., & Kliegl, O. (2017). Retrieval-induced remembering and forgetting. In J. T. Wixted (Ed.), *Learning and Memory: A comprehensive reference: Vol. 2. Cognitive psychology of memory* (2nd ed., pp. 27–51). Academic Press.

Bäuml, K.-H. T., & Samenieh, A. (2010). The two faces of memory retrieval. *Psychological Science, 21*(6), 793-795.

Bäuml, K.-H. T., & Trißl, L. (2022). Selective memory retrieval can revive forgotten memories. *Proceedings of the National Academy of Sciences, 119*(8), Article e2114377119.

Benjamin, A. S. (2006). The effects of list-method directed forgetting on recognition memory. *Psychonomic Bulletin & Review, 13*(5), 831–836.

Bertilsson, F., Stenlund, T., Wiklund-Hörnqvist, C., & Jonsson, B. (2021). Retrieval practice: Beneficial for all students or moderated by individual differences? *Psychology Learning & Teaching, 20*(1), 21–39.

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9*(3), 255–268.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency & related phenomena. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 123–144). Erlbaum.

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

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

Bjork, R. A. (2011). On the symbiosis of remembering, forgetting, and learning. In A. S. Benjamin (Ed.), *Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork* (pp. 1–22). Psychology Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes* (Vol. 2, pp. 35–67). Erlbaum.

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

Bogartz, R. S. (1990). Evaluating forgetting curves psychologically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 138–148.

Bolhuis, J. J., & van Kampen, H. S. (1988). Serial position curves in spatial memory of rats: Primacy and recency effects. *The Quarterly Journal of Experimental Psychology, 40*(2), 135–149.

Bouwmeester, S., & Verkoeijen, P. P. J. L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language, 65*(1), 32–41.

Brown, G. D. A., & Lewandowsky, S. (2010). Forgetting in memory models: Arguments against trace decay and consolidation failure. In S. Della Sala (Ed.), *Forgetting* (pp. 49–75). Psychology Press.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*(3), 539–576.

Brown, J. (1954). The nature of set-to-learn and of intra-material interference in immediate memory. *Quarterly Journal of Experimental Psychology, 6*(4), 141–148.

Bulevich, J. B., Roediger, H. L., Balota, D. A., & Butler, A. C. (2006). Failures to find suppression of episodic memories in the think/no- think paradigm. *Memory & Cognition, 34*(8), 1569–1577.

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

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

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

Carlesimo, G. A., Fadda, L., Bonci, A., & Caltagirone, C. (1993). Differential rates of forgetting from long-term memory in Alzheimer's and multi-infarct dementia. *International Journal of Neuroscience, 73*(1–2), 1–11.

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.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1547-1552.

Carpenter, S. K., & Delosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268-276.

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review, 14*(3), 474-478.

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

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language, 92*, 128–141.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*(6), 633-642.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology, 56*(4), 236-246.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19*(11), 1095–1102.

Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory, 18*(1), 49-57.

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*, Article 117983.

Clark, R. E., Broadbent, N. J., Zola, S. M., & Squire, L. R. (2002). Anterograde amnesia and temporally graded retrograde amnesia for a nonspatial memory task after lesions of hippocampus and subiculum. *The Journal of Neuroscience, 22*(11), 4663–4669.

Congleton, A. R., & Rajaram, S. (2011). The influence of learning methods on collaboration: Prior repeated retrieval enhances retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory. *Journal of Experimental Psychology: General, 140*(4), 535–551.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language, 53*(4), 594–628.

Conway, M., & Fthenaki, A. (2003). Disruption of inhibitory control of memory following lesions to the frontal and temporal lobes. *Cortex, 39*(4–5), 667-686.

Conway, M. A., Harries, K., Noyes, J., Racsmány, M., & Frankish, C. R. (2000). The disruption and dissolution of directed forgetting: Inhibitory control of memory. *Journal of Memory and Language, 43*(3), 409–430.

Cowan, E. T., Schapiro, A. C., Dunsmoor, J. E., & Murty, V. P. (2021). Memory consolidation as an adaptive process. *Psychonomic Bulletin & Review, 28*, 1796–1810.

Davidson, P., Hellerstedt, R., Jönsson, P., & Johansson, M. (2020). Suppression-induced forgetting diminishes following a delay of either sleep or wake. *Journal of Cognitive Psychology, 32*(1), 4–26.

Davis, J. C., & Okada, R. (1971). Recognition and recall of positively forgotten items. *Journal of Experimental Psychology, 89*(1), 181–186.

Davis, H., Small, S., Stern, Y., Mayeux, R., Feldstein, S., & Keller, F. (2003). Acquisition, recall, and forgetting of verbal information in long-term memory by young, middle-aged, and elderly individuals. *Cortex, 39*(4–5), 1063–1091.

Davis, R. L., & Zhong, Y. (2017). The biology of forgetting—A perspective. *Neuron, 95*(3), 490–503.

Delaney, P. F., Nghiem, K. N., & Waldum, E. R. (2009). The selective directed forgetting effect: Can people forget only part of a text? *Quarterly Journal of Experimental Psychology, 62*(8), 1542–1550.

Dewar, M. T., Cowan, N., & Della Sala, S. (2007). Forgetting due to retroactive interference: A fusion of Müller and Pilzecker's (1900) early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex, 43*(5), 616–634.

Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition, 42*(7), 1049-1062.

Donkin, C., & Nosofsky, R. M. (2012). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological Science, 23*(6), 625-634.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology, 55*(1), 51–86.

Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron, 88*(1), 20–32.

Duyck, W., Desmet, T., Verbeke, L. P. C., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers, 36*(3), 488–499.

Ebbinghaus, H. (1885). *Über das Gedächtnis - Untersuchungen zur experimentellen Psychologie* [About memory - studies in experimental psychology]. Duncker & Humblot.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53*(2), 134–140.

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.

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.

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.

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.

Fischer, S., Diekelmann, S., & Born, J. (2011). Sleep's role in the processing of unwanted memories: Sleep and processing of unwanted memories. *Journal of Sleep Research, 20*(2), 267–274.

Fisher, J. S., & Radvansky, G. A. (2018). Patterns of forgetting. *Journal of Memory and Language, 102*, 130–141.

Fjell, A. M., Walhovd, K. B., Reinvang, I., Lundervold, A., Dale, A. M., Quinn, B. T., Makris, N., & Fischl, B. (2005). Age does not increase rate of forgetting over weeks— Neuroanatomical volumes and visual memory across the adult life-span. *Journal of the International Neuropsychological Society, 11*(1), 2–15.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6*(40).

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.

Giambra, L. M., & Arenberg, D. (1993). Adult age differences in forgetting sentences. *Psychology and Aging, 8*(3), 451–462.

Glover, J. A. (1989). The "testing“ phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392–399.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*(3), 325–331.

Golding, J. M., & Gottlob, L. R. (2005). Recall order determines the magnitude of directed forgetting in the within-participants list method. *Memory & Cognition, 33*(4), 588–594.

Haist, F., Gore, J. B., & Mao, H. (2001). Consolidation of human memory over decades revealed by functional magnetic resonance imaging. *Nature Neuroscience, 4*(11), 1139–1145.

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

Hanslmayr, S., Volberg, G., Wimber, M., Oehler, N., Staudigl, T., Hartmann, T., Raabe, M., Greenlee, M. W., & Bäuml, K.-H. T. (2012). Prefrontally driven downregulation of neural synchrony mediates goal-directed forgetting. *The Journal of Neuroscience, 32*(42), 14742–14751.

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences, 17*(3), 111–120.

Hart, R. P., Kwentus, J. A., Taylor, J. R., & Harkins, S. W. (1987). Rate of forgetting in dementia and depression. *Journal of Consulting and Clinical Psychology, 55*(1), 101–105.

Hayne, H., Garry, M., & Loftus, E. F. (2006). On the continuing lack of scientific evidence for repression. *Behavioral and Brain Sciences, 29*(5), 521–522.

Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review, 12*(3), 484–489.

Hertel, P. T., & Gerstle, M. (2003). Depressive deficits in forgetting. *Psychological Science, 14*(6), 573–578.

Hesiod (1914). *The Homeric Hymns And Homerica* (H. G. Evelyn-White, Trans.). London Heinemann. Retrieved online from: https://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0130 (Original work published ca. 700 B.C.E.)

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10*(5), 562–567.

Hotta, C., & Kawaguchi, J. (2009). Self-initiated use of thought substitution can lead to long term forgetting. *Psychologia, 52*(1), 41–49.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*(3), 269-299.

Hupbach, A. (2018). Long-term effects of directed forgetting. *Memory, 26*(3), 321–329.

Huppert, F. A., & Piercy, M. (1979). Normal and abnormal forgetting in organic amnesia: Effect of locus of lesion. *Cortex, 15*(3), 385–390.

Isarida, T., Sakai, T., Kubota, T., Koga, M., Katayama, Y., & Isarida, T. K. (2014). Odor-context effects in free recall after a short retention interval: A new methodology for controlling adaptation. *Memory & Cognition, 42*(3), 421–433.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*(6), 649–667.

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1441–1451.

James, W. (1952). *The principles of psychology.* Encyclopaedia Britannica. (Original work published 1890)

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology, 35*(4), 605–612.

Johnson, H. M. (1994). Processes of successful intentional forgetting. *Psychological Bulletin, 116*(2), 274–292.

Joslyn, S. L., & Oakes, M. A. (2005). Directed forgetting of autobiographical events. *Memory & Cognition, 33*(4), 577–587.

Jost, A. (1897). Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen [The strength of associations in their dependence on the distribution of repetitions]. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 14*, 436–472.

Kahana, M. J., & Adler, M. (2002). *Note on the power law of forgetting.* PsyArXiv. https://doi.org/10.1101/173765

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4–5), 528–558.

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. H. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (2nd ed., pp. 487–514). Elsevier.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772–775.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 61, pp. 237–284). Elsevier Academic Press.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language, 67*(1), 17–29.

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966–968.

Klein, S. B., & Nichols, S. (2012). Memory and the sense of personal identity. *Mind, 121*(483), 677–702.

Klein, K. A., Shiffrin, R. M., & Criss, A. H. (2007). Putting context in context. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 171–189). Psychology Press.

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

Kliegl, O., Carls, T., & Bäuml, K.-H. T. (2019). How delay influences search processes at test. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(12), 2174–2187.

Kliegl, O., Wallner, L., & Bäuml, K.-H. T. (2018). Selective directed forgetting in children. *Journal of Experimental Child Psychology, 167*, 433–440.

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

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989–998.

Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12*(1), 71–78.

Lazar, N. (2010). Ockham's Razor. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(2), 243–246.

Lee, J. L. C., Nader, K., & Schiller, D. (2017). An update on memory reconsolidation updating. *Trends in Cognitive Sciences, 21*(7), 531–545.

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 970–988.

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review, 120*(1), 155–189.

Liu, X. (2001). On the dynamics of directed forgetting: Facilitation and interference in the updating of human memory. *Dissertation Abstracts International, 61*, 6459B.

Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(2), 397–406.

Luria, A. R. (1968). *The mind of a mnemonist: A little book about a vast memory* (L. Solotaroff, Trans.). Basic Books.

MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning & Memory, 104*(3), 271–279.

MacLeod, C. M. (1989). Directed forgetting affects both direct and indirect tests of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(1), 13–21.

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.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 43, pp. 163–214). Elsevier.

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

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates.

Marevic, I., & Rummel, J. (2020). Retrieval-mediated directed forgetting in the item-method paradigm: The effect of semantic cues. *Psychological Research, 84*(3), 685–705.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*(4–5), 494–513.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*(2), 192–201.

McGaugh, J. L. (2000). Memory—A century of consolidation. *Science, 287*(5451), 248–251.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39*(4), 352–370.

Meier, B., König, A., Parak, S., & Henke, K. (2011). Suppressed, but not forgotten. *Swiss Journal of Psychology, 70*(1), 5–11.

Mensink, G.-J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95*(4), 434–455.

Mensink, G.-J. M., & Raaijmakers, J. G. W. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology, 33*(2), 172–186.

Meyer, A. N. D., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging, 28*(1), 142–147.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*(5), 519–533.

Mulji, R., & Bodner, G. E. (2010). Wiping out memories: New support for a mental context change account of directed forgetting. *Memory, 18*(7), 763–773.

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

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

Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis [Experimental contributions to the study of memory]. *Zeitschrift Für Psychologie. Ergänzungsband, 1*, 1–300.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General, 128*(4), 403–415.

Murre, J. M. J. (2021). The Godden and Baddeley (1975) experiment on context-dependent memory on land and underwater: A replication. *Royal Society Open Science, 8*(11), Article 200724.

Murre, J. M. J., & Chessa, A. G. (2011). Power laws from individual differences in learning and forgetting: Mathematical analyses. *Psychonomic Bulletin & Review, 18*(3), 592–597.

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLOS ONE, 10*(7), Article e0120644.

Murty, V. P., Tompary, A., Adcock, R. A., & Davachi, L. (2017). Selectivity in postencoding connectivity with high-level visual cortex is associated with reward-motivated memory. *The Journal of Neuroscience, 37*(3), 537–545.

Muther, W. S. (1965). Erasure or partitioning in short-term memory. *Psychonomic Science, 3*(1–12), 429–430.

Myers, G. C. (1914). Recall in relation to retention. *Journal of Educational Psychology, 5*(3), 119-130.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*(1), 90–100.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology, 7*, 217–227.

Nader, K., & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nature Reviews Neuroscience, 10*(3), 224–234.

Neisser, U. (1984). Interpreting Harry Bahrick's discovery: What confers immunity against forgetting? *Journal of Experimental Psychology: General, 113*(1), 32–35.

Noreen, S., & MacLeod, M. D. (2014). To think or not to think, that is the question: Individual differences in suppression and rebound effects in autobiographical memory. *Acta Psychologica, 145*, 84–97.

Nowicka, A., Marchewka, A., Jednorog, K., Tacikowski, P., & Brechmann, A. (2011). Forgetting of emotional information is hard: An fMRI study of directed forgetting. *Cerebral Cortex, 21*(3), 539–549.

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin, 144*(7), 710–756.

Parker, E. S., Cahill, L., & McGaugh, J. L. (2006). A case of unusual autobiographical remembering. *Neurocase, 12*(1), 35–49.

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

Pastötter, B., & Bäuml, K.-H. T. (2007). The crucial role of postcue encoding in directed forgetting and context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(5), 977–982.

Pastötter, B., & Bäuml, K.-H. T. (2010). Amount of postcue encoding predicts amount of directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 54–65.

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology, 5*, Article 286.

Pastötter, B., & Bäuml, K.-H. T. (2016). Reversing the testing effect by feedback: Behavioral and electrophysiological evidence. C *ognitive, Affective, & Behavioral Neuroscience, 16*(3), 473–488.

Pastötter, B., Bäuml, K.-H., & Hanslmayr, S. (2008). Oscillatory brain activity before and after an internal context change—Evidence for a reset of encoding processes. *NeuroImage, 43*(1), 173–181.

Pastötter, B., Kliegl, O., & Bäuml, K.-H. T. (2012). List-method directed forgetting: The forget cue improves both encoding and retrieval of postcue information. *Memory & Cognition, 40*(6), 861–873.

Pastötter, B., Tempel, T., & Bäuml, K.-H. T. (2017). Long-term memory updating: The reset-of-encoding hypothesis in list-method directed forgetting. *Frontiers in Psychology, 8*, Article 2076.

Potts, R., & Shanks, D. R. (2012). Can testing immunize memories against interference? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(6), 1780–1785.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*(4), 437–447.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*(6002), 335–335.

R Core Team (2021). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*(2), 93–134.

Racsmány, M., & Keresztes, A. (2015). Initial retrieval shields against retrieval-induced forgetting. *Frontiers in Psychology, 6*, Article 657.

Racsmány, M., Szőllősi, Á., & Bencze, D. (2018). Retrieval practice makes procedure from remembering: An automatization account of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(1), 157–166.

Racsmány, M., Szőllősi, Á., & Marián, M. (2020). Reversing the testing effect by feedback is a matter of performance criterion at practice. *Memory & Cognition, 48*(7), 1161–1170.

Radvansky, G. A., Doolen, A. C., Pettijohn, K. A., & Ritchey, M. (2022). A new look at memory retention and forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(11), 1698-1723.

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science, 315*(5817), 1426–1429.

Rauchs, G., Feyers, D., Landeau, B., Bastin, C., Luxen, A., Maquet, P., & Collette, F. (2011). Sleep contributes to the strengthening of some memories over others, depending on hippocampal activity at learning. *Journal of Neuroscience, 31*(7), 2563–2568.

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

Reed, P. (1996). Rats' memory for serially presented novel flavours: Evidence for non-spatial primacy effects. *The Quarterly Journal of Experimental Psychology, 49*(2), 174–187.

Reverberi, S., Kohn, N., & Fernández, G. (2020). No evidence for an effect of explicit relevance instruction on consolidation of associative memories. *Neuropsychologia, 143*, Article 107491.

Ribot, T. (1882). *The diseases of memory: An essay in the positive psychology* (W. Huntington Smith, Trans.). D. Appleton and Company. (Original work published 1881)

Rickard, T. C., & Pan, S. C. (2018). A dual memory theory of the testing effect. *Psychonomic Bulletin & Review, 25*(3), 847–869.

Rickles, W. H., Cohen, M. J., Whitaker, C. A., & McIntyre, K. E. (1973). Marijuana induced state-dependent verbal learning. *Psychopharmacologia, 30*(4), 349–354.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95*(3), 318–339.

Rivera-Lares, K., Logie, R., Baddeley, A., & Della Sala, S. (2022). Rate of forgetting is independent of initial degree of learning. *Memory & Cognition, 50*, 1706-1718.

Rizio, A. A., & Dennis, N. A. (2013). The neural correlates of cognitive control: Successful remembering and intentional forgetting. *Journal of Cognitive Neuroscience, 25*(2), 297–312.

Rodini, M., De Simone, M. S., Caltagirone, C., & Carlesimo, G. A. (2022). Accelerated long-term forgetting in neurodegenerative disorders: A systematic review of the literature. *Neuroscience & Biobehavioral Reviews, 141*, Article 104815.

Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition, 2*(2), 261–269.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27.

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210.

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011a). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17*(4), 382–395.

Roediger III, H. L., Putnam, A. L., & Smith, M. A. (2011b). Ten benefits of testing and their applications to educational practice. In J. P. Mestre & B. H. Ross (Eds.) *Psychology of Learning and Motivation* (Vol. 55, pp. 1–36). Elsevier.

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.

RStudio Team (2020). RStudio: Integrated development for R [Computer software]. RStudio, PBC, Boston, MA.

Rubin, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 21–38.

Rubin, D. C., Hinton, S., & Wenzel, A. (1999). The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(5), 1162–1176.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*(4), 734–760.

Rummel, J., Marevic, I., & Kuhlmann, B. G. (2016). Investigating storage and retrieval processes of directed forgetting: A model-based approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(10), 1526–1543.

Runquist, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition, 11*(6), 641–650.

Runquist, W. N.. (1986a). Changes in the rate of forgetting produced by recall tests. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 40*(3), 282–289.

Runquist, W. N. (1986b). The effect of testing on the forgetting of related and unrelated associates. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 40*(1), 65–76.

Runquist, W. N. (1987). Retrieval specificity and the attenuation of forgetting by testing. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 41*(1), 84–90.

Sahakyan, L., & Delaney, P. F. (2003). Can encoding differences explain the benefits of directed forgetting in the list method paradigm? *Journal of Memory and Language, 48*(1), 195–206.

Sahakyan, L., & Delaney, P. F. (2005). Directed forgetting in incidental learning and recognition testing: Support for a two-factor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 789–801.

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. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 59, pp. 131–190). Elsevier.

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(6), 1064-1072.

Saletin, J. M., Goldstein, A. N., & Walker, M. P. (2011). The role of sleep in directed forgetting and remembering of human memories. *Cerebral Cortex, 21*(11), 2534–2541.

Saufley, W. H., Otaka, S. R., & Bavaresco, J. L. (1985). Context effects: Classroom tests and context independence. *Memory & Cognition, 13*(6), 522–528.

Schindler, S., & Kissler, J. (2018). Too hard to forget? ERPs to remember, forget, and uninformative cues in the encoding phase of item-method directed forgetting. *Psychophysiology, 55*(10), Article e13207.

Scullin, M. K., Fairley, J., Decker, M. J., & Bliwise, D. L. (2017). The effects of an afternoon nap on episodic memory in young and older adults. *Sleep, 40*(5), Article zsx035.

Scully, I. D., & Hupbach, A. (2020). Directed forgetting affects how we remember and judge other people. *Journal of Applied Research in Memory and Cognition, 9*(3), 336–344.

Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy effect in episodic memory. *NeuroImage, 32*(3), 1422–1431.

Sego, S. A., Golding, J. M., & Gottlob, L. R. (2006). Directed forgetting in older adults using the item and list methods. *Aging, Neuropsychology, and Cognition, 13*(1), 95–114.

Shapiro, S., Lindsey, C., & Krishnan, H. S. (2006). Intentional forgetting as a facilitator for recalling new product attributes. *Journal of Experimental Psychology: Applied, 12*(4), 251–263.

Sharot, T., & Phelps, E. A. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience, 4*(3), 294–306.

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

Shin, Y. S., Masís-Obando, R., Keshavarzian, N., Dáve, R., & Norman, K. A. (2021). Context-dependent memory effects in two immersive virtual reality environments: On Mars and underwater. *Psychonomic Bulletin & Review, 28*(2), 574–582.

Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin, 49*(3), 263–269.

Siler, J., & Benjamin, A. S. (2020). Long-term inference and memory following retrieval practice. *Memory & Cognition, 48*(4), 645–654.

Slamecka, N. J. (1985). On comparing rates of forgetting: Comment on Loftus (1985). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 812–816.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592–604.

Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 716–727.

Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(3), 384–397.

Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior, 10*(4), 400–408.

Smith, A. M., Floerke, V. A., & Thomas, A. K. (2016). Retrieval practice protects memory against acute stress. *Science, 354*(6315), 1046–1048.

Smith, A. M., Davis, F. C., & Thomas, A. K. (2018). Criterial learning is not enough: Retrieval practice is necessary for improving post-stress memory accessibility. *Behavioral Neuroscience, 132*(3), 161–170.

Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1712–1725.

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(6), 1582–1593.

Smith, S. M., & Handy, J. D. (2016). The crutch of context-dependency: Effects of contextual support and constancy on acquisition and retention. *Memory, 24*(8), 1134–1141.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34–50.

Soderstrom, N. C., Kerr, T. K., & Bjork, R. A. (2016). The critical importance of retrieval—and spacing—for learning. *Psychological Science, 27*(2), 223–230.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30*(9), 641–656.

Squire, L. R. (1981). Two forms of human amnesia: An analysis of forgetting. *The Journal of Neuroscience, 1*(6), 635–640.

Squire, L. R. (1989). On the course of forgetting in very long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 241–245.

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology, 7*(8), Article a021766.

Stamate, A., Logie, R. H., Baddeley, A. D., & Della Sala, S. (2020). Forgetting in Alzheimer's disease: Is it fast? Is it affected by repeated retrieval? *Neuropsychologia, 138*, Article 107351.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature, 437*(7063), 1272–1278.

Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience, 16*(2), 139–145.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2012). On the durability of retrieval-induced forgetting. *Journal of Cognitive Psychology, 24*(5), 617–629.

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

Szőllősi, Á., Keresztes, A., Novák, B., Szászi, B., Kéri, S., & Racsmány, M. (2017). The testing effect is preserved in stressful final testing environment. *Applied Cognitive Psychology, 31*, 615–622.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1392–1399.

Takashima, A., Nieuwenhuis, I. L. C., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *The Journal of Neuroscience, 29*(32), 10087–10093.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning & Memory, 4*(3), 210–221.

Thorndike, E. L. (1913). *The psychology of learning: Vol. II.* Teachers College, Columbia University.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology, 56*(4), 252–257.

Trißl, L., & Bäuml, K.-H. T. (2022). Selective restudy can reset recall of forgotten information. *Psychonomic Bulletin & Review, 29*(6), 2202–2210.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science, 2*(3), 67–70.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 381–391.

Ullsperger, M., Mecklinger, A., & Müller, U. (2000). An electrophysiological test of directed forgetting: The role of retrieval inhibition. *Journal of Cognitive Neuroscience, 12*(6), 924–940.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64*(1), 49–60.

Underwood, B. J., & Keppel, G. (1963). Retention as a function of degree of learning and letter-sequence interference. *Psychological Monographs: General and Applied, 77*(4), 1–16.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Variation in working memory capacity and forgetting over both the short and the long term: An application of the Population Dilution model. *Journal of Cognitive Psychology, 23*(2), 243–255.

Van Dongen, E. V., Thielen, J.-W., Takashima, A., Barth, M., & Fernández, G. (2012). Sleep supports selective retention of associative memories based on relevance for future utilization. *PLoS ONE, 7*(8), Article e43426.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*(3), 289–335.

Wallner, L., & Bäuml, K.-H. T. (2017). Beneficial effects of selective item repetition on the recall of other items. *Journal of Memory and Language, 95*, 159–172.

Wamsley, E. J., Hamilton, K., Graveline, Y., Manceor, S., & Parr, E. (2016). Test expectation enhances memory consolidation across both sleep and wake. *PLOS ONE, 11*(10), Article e0165141.

Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology, 53*(1), 5–13.

Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning & Memory, 6*(2), 135–144.

Werner, P., Ulitsa, N., Shephet, D., Abojabel, H., Alpinar-Sencan, Z., & Schicktanz, S. (2021). Fear about Alzheimer's disease among Israeli and German laypersons, persons with Mild Neurocognitive Disorder and their relatives: A qualitative study. *International Psychogeriatrics, 33*(10), 1019–1034.

Wessel, I., & Merckelbach, H. (2006). Forgetting "murder“ is not harder than forgetting "circle“: Listwise-directed forgetting of emotional words. *Cognition & Emotion, 20*(1), 129–137.

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*(6), 571–580.

Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3*(4), 240–246.

White, K. G. (2001). Forgetting functions. *Animal Learning & Behavior, 29*(3), 193–207.

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 9*(4), 418–455.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition, 2*(4), 775–780.

Wickelgren, W. A. (1975). Age and storage dynamics in continuous recognition memory. *Developmental Psychology, 11*(2), 165–169.

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

Wickens, T. D. (1998). On the form of the retention function: Comment on Rubin and Wenzel (1996): A quantitative description of retention. *Psychological Review, 105*(2), 379–386.

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., & Born, J. (2011). Sleep selectively enhances memory expected to be of future relevance. *The Journal of Neuroscience, 31*(5), 1563–1569.

Winocur, G., Sekeres, M. J., Binns, M. A., & Moscovitch, M. (2013). Hippocampal lesions produce both nongraded and temporally graded retrograde amnesia in the same rat. *Hippocampus, 23*(5), 330–341.

Wixted, J. T. (1990). Analyzing the empirical course of forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 927–935.

Wixted, J. T. (2004a). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review, 111*(4), 864–879.

Wixted, J. T. (2004b). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55*(1), 235–269.

Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science, 14*(1), 6–9.

Wixted, J. T. (2022a). Absolute versus relative forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(12), 1775-1786.

Wixted, J. T. (2022b). The enigma of forgetting. *Proceedings of the National Academy of Sciences, 119*(12), Article e2201332119.

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science, 2*(6), 409–415.

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

Wixted, J. T., Ghadisha, H., & Vera, R. (1997). Recall latency following pure- and mixed-strength lists: A direct test of the relative strength model of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 523–538.

Wolf, O. T., & Kluge, A. (2017). Commentary: Retrieval practice protects memory against acute stress. *Frontiers in Behavioral Neuroscience, 11*, Article 48.

Woodward, A. E., Bjork, R. A., & Jongeward, R. H. (1973). Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 608–617.

Wright, A. A. (1994). Primacy effects in animal memory and human nonverbal memory. *Animal Learning & Behavior, 22*(2), 219–223.

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Cook, R. G. (1985). Memory processing of serial lists by pigeons, monkeys, and people. *Science, 229*(4710), 287–289.

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.

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience, 20*(6), 364–375.

Zacks, R. T., Hasher, L., & Radvansky, G. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 143–156.

Zellner, M., & Bäuml, K.-H. T. (2006). Inhibitory deficits in older adults: List-method directed forgetting revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 290–300.

Zhang, X., Li, Q., Wang, L., Liu, Z.-J., & Zhong, Y. (2016). Cdc42-dependent forgetting regulates repetition effect in prolonging memory retention. *Cell Reports, 16*(3), 817–825.

Zwissler, B., Schindler, S., Fischer, H., Plewnia, C., & Kissler, J. M. (2015). 'Forget me (not)?' – Remembering forget-items versus un-cued items in directed forgetting. *Frontiers in Psychology, 6*, Article 1741.

Appendix

Appendix A

Maximum likelihood models

The following shows how the likelihood-functions for the statistical baseline models and the separate power function models for the individual conditions were set up for Experiment 1.

Statistical baseline model

$$L(B; D) = (p_1^{n_1} \cdot (1 - p_1)^{1-n_1}) + ... + (p_m^{n_m} \cdot (1 - p_m)^{1-n_m})$$

with
n = number of subjects $\cdot$ number of words per condition
p_i = mean recall rate at time i
$n_1 = p_i \cdot n$
m = number of delay intervals

For the separate power function models for each condition, each p_i was replaced with the power function of time, using the parameters to be estimated (a and b), the pre-estimated parameter c and the corresponding value of t (either 0.002083, 1, 3 or 7 days).

$p_1 = a(1 + c \cdot 0.002083)^{-b}$
$p_2 = a(1 + c \cdot 1)^{-b}$
$\cdots$

Power function model

$$L(M;D) = ((a(1+c\cdot 0.002083)^{-b})^{n_1} \cdot (1-(a(1+c\cdot 0.002083)^{-b}))^{1-n_1})+$$
$$+((a(1+c\cdot 1)^{-b})^{n_2} \cdot (1-(a(1+c\cdot 1)^{-b}))^{1-n_2})+...$$

Common parameters c were estimated by including data at all m delay intervals for both conditions (TBR and TBF) into one model, with separate parameters a, b, and c ($a1, a2, b1, b2, c1, c2$) for each condition in the more general model ($L(B;D)$) and separate a and b, but a common parameter c ($a1, a2, b1, b2, c$) in the more restricted model ($L(M;D)$).

For a comparison of the models an approximate χ^2 test was used of the form

$$\chi^2 = -2\ln\frac{L(M;D)}{L(B;D}$$

with degrees of freedom determined by the difference in number of parameters between the two models.

Appendix B

Serial Position Analysis for Experiment 3

The next pages show additional descriptive data for the serial position analysis for Experiment 3 (LMDF). Included are figures for both lists and all delay intervals (3 min, 1, 2, 3, and 7 days). Recall rates for all 16 items of both lists depending on initial output order during study were calculated for $n = 296$ subjects. Due to experimenter error, the information required for this analysis was missing for four subjects (see Results section for Experiment 3). The recall rates were then averaged across participants in all CUE × DELAY conditions, first for all 16 item positions, and then, for data reduction purposes, for four bins, spanning four items each. Following Pastötter and Bäuml (2010), the recall rates were also smoothed to make the data patterns more visible. Figures B1 and B2 show the serial position results (for all positions and for the four bins, as well as for smoothed recall rates) for both lists of Experiment 3, for both cue conditions and all five delay intervals. In all conditions, primacy effects can be observed, with higher recall for early compared to late list items, and these effects differ in size between the two cuing-conditions. For List 1 (Figure B1), a primacy benefit can be observed for remember-cued participants that increases in size at longer retention intervals. For List 2 (Figure B2), the primacy benefit for forget-cued participants remains stable across retention intervals. Table B1 shows the proportion recalled of the list participants started their free recall with (List 1 or List 2) before they switched to the other list the first time. Data is displayed for all delay conditions and both cue conditions, depending on self-chosen list-output order. The table shows cell sizes and illustrates that this proportion remained numerically stable over delays and conditions.

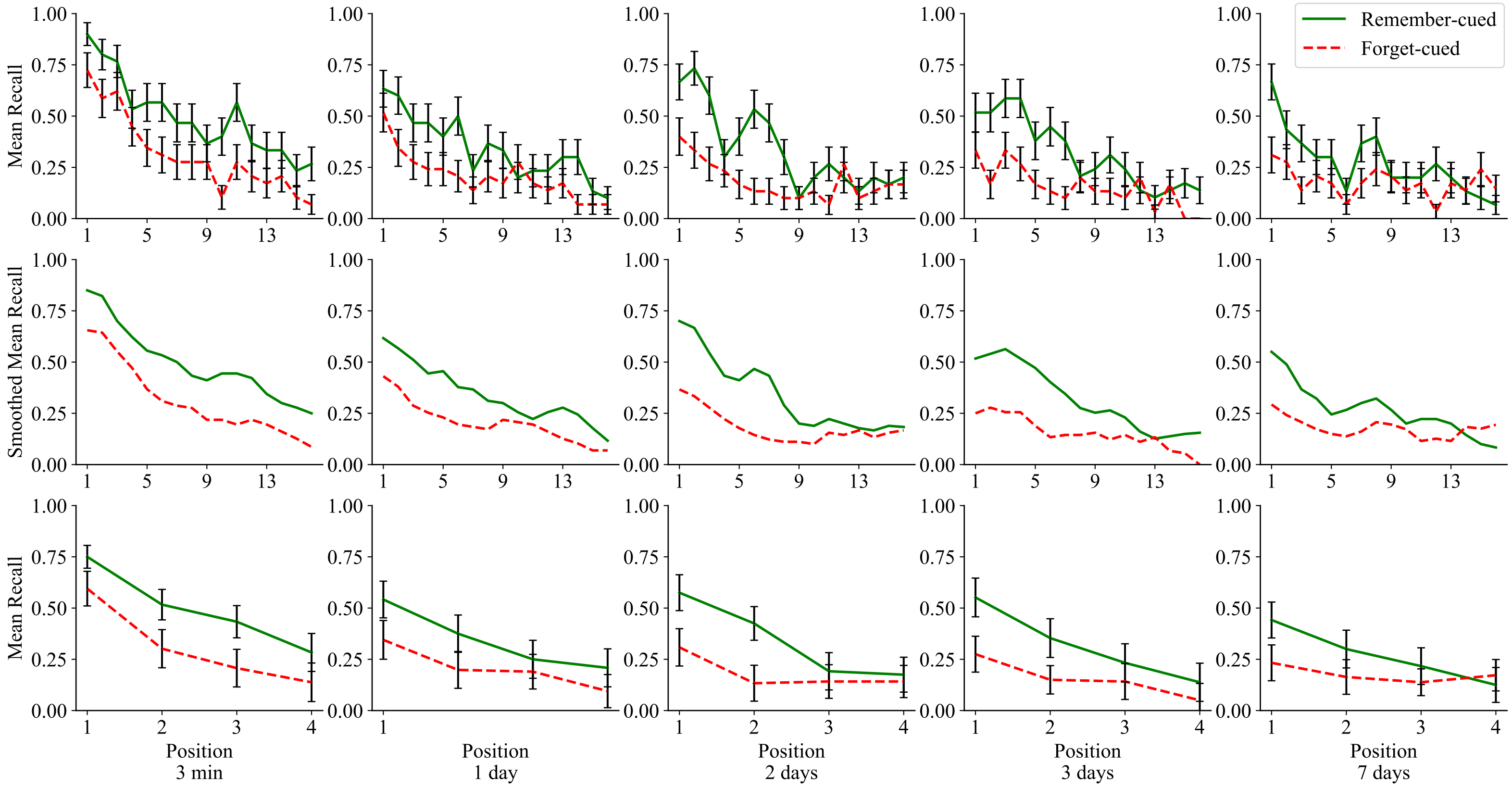

Figure B1: *Serial Position Effects for Experiment 3: List 1 Results.* Recall rates for to-be-remembered (solid green) and to-be-forgotten (dashed red) information depending on their serial position during initial study are displayed for all five delay intervals. The upper row shows the results for all 16 positions during study, the lower row shows the same information divided into four bins including four item positions each: Bin 1 (items 1-4), Bin 2 (items 5-8), Bin 3 (items 9-12), Bin 4 (items 13-16). The middle row shows the smoothed recall rates (following Pastötter & Bäuml, 2010) for all 16 item positions. Error bars represent ± 1 standard error.

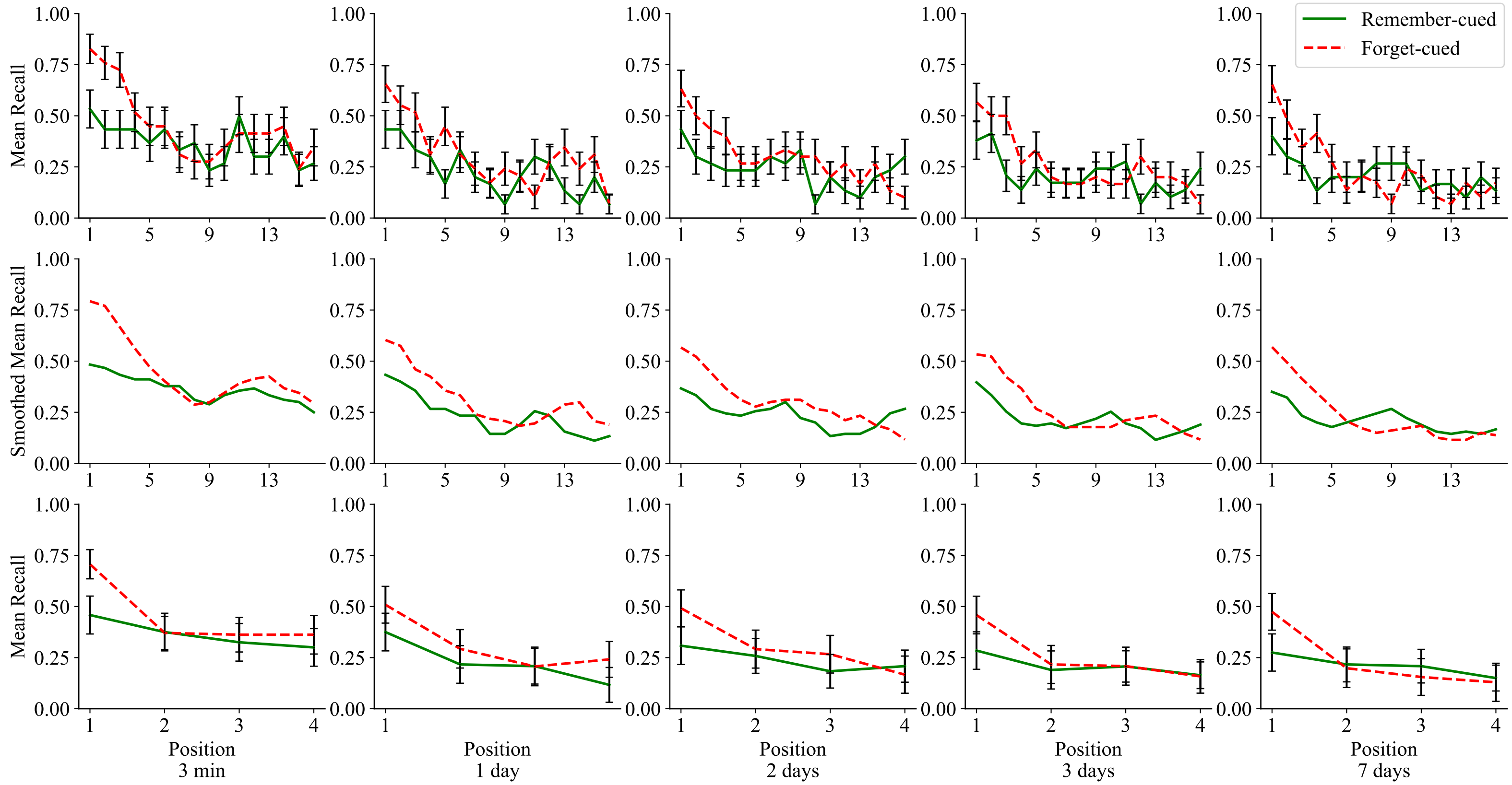

Figure B2: *Serial Position Effects for Experiment 3: List 2 Results.* Recall rates for items learned in the remember-cued (solid green) and the forget-cued conditions (dashed red) depending on their serial position during initial study are displayed for all five delay intervals. The upper row shows the results for all 16 positions during study, the lower row shows the same information divided into four bins including four item positions each: Bin 1 (items 1-4), Bin 2 (items 5-8), Bin 3 (items 9-12), Bin 4 (items 13-16). The middle row shows the smoothed recall rates (following Pastötter & Bäuml, 2010) for all 16 item positions. Error bars represent ± 1 standard error.

Table B1

Mean Proportion of List Recalled Before First List-Switch for Experiment 3, Depending on Cue- and Delay-Condition

	3 min		**1 day**		**2 days**		**3 days**		**7 days**	
Condition	*M (SD)*	*N*	*M (SD)*	*N*	*M (SD)*	*N*	*M (SD)*	*N*	*M (SD)*	*N*
Start with List 1										
Remember-Cued	.59 (.31)	25	.74 (.25)	21	.83 (.21)	22	.77 (.28)	20	.86 (.22)	17
Forget-Cued	.58 (.29)	21	.59 (.30)	15	.59 (.28)	8	.68 (.30)	10	.68 (.26)	11
Start with List 2										
Remember-Cued	.30 (.19)	5	.75 (.29)	6	.79 (.23)	3	.58 (.29)	4	.65 (.29)	7
Forget-Cued	.78 (.23)	6	.67 (.30)	11	.63 (.31)	16	.72 (.25)	11	.81 (.29)	12

Note. The table includes only data for subjects who correctly recalled items from both lists ($N = 251$).