Retrieval-Induced Remembering of Episodic Memories

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Philosophischen Fakultät II
(Psychologie, Pädagogik und Sportwissenschaft)
der Universität Regensburg

vorgelegt von

LISA ANNE WALLNER

aus Regensburg

Regensburg 2018

Erstgutachter: Prof. Dr. Karl-Heinz T. Bäuml

Zweitgutachter: Prof. Dr. Klaus W. Lange

Acknowledgement

This dissertation would never have come to fruition without the support of many individuals, and it is with pleasure and gratitude that I acknowledge their efforts.

First and most of all, I want to convey my genuine gratitude to my supervisor and the initiator of this dissertation, Prof. Dr. Karl-Heinz Bäuml, who was very generous with his time for questions and concerns, his scientific knowledge and experience. I am profoundly indebted for assisting me in each step with his helpful suggestions regarding memory research and teaching. In particular, I want to thank him for supporting me in pursuing not only an academic but also a clinical career.

I am grateful to my colleagues for their intellectual support, for always lending me an ear, for sharing their thoughts, but even more so for sharing lunch breaks, for diverting and entertaining chats and having good times. My (current and former) colleagues Magdalena Abel, Oliver Kliegl, Eva-Maria Lehmer, Bernhard Pastötter, Julia Rupprecht, Michael Wirth and Petra Witzmann definitively enriched my workaday life with their sympathy, humor and patience. Moreover, I want to allocate credit to all individuals who have contributed to the data collections this thesis is built on, the lab's research assistants (e. g., Christian Ebner, Armin Ederer, Elisabeth Meier, Jessica Völkl) and the scores of students that participated in my experiments.

In a completely different way, I want to thank my partner Mi for the endless support, encouragement, patience and understanding, and for not just tolerating my occasionally bad moods but rather cheering me up whenever I need it.

And finally, but certainly not least, thanks go to my friends and my family, who bring so much joy and fun into my life. I am greatly indebted to my sister and my parents. It was their unconditional love, care, and tolerance which afforded me writing this work. I would like to dedicate this dissertation to my family. You taught me the meaning of life and made me the person I am today, I owe everything to you.

Preface

"With one singular exception, time's arrow is straight. [...] The singular exception is provided by the human ability to remember past happenings." With this proposition, Tulving (2002) introduced his review of episodic memory, emphasizing its singularity and meaning. Episodic memory forms an integral part of our declarative long-term memory and refers to the collection of memories that are associated with a specific episode in time and space (Baddeley, 2001; Tulving, 1972, 1985). Particularly, such memories range from autobiographical events, i. e., past personal experiences like what we did yesterday or where we spent our last holiday trip, up to remembering single previously encoded items like sentences or words. Due to this wide range of memories that we have to deal with, episodic memory plays a key role in our everyday lives. Indeed, retrieving past personal experiences and episodes we experienced in our lives helps us to define who we are.

The three stages of our episodic memory contain encoding new information, consolidating it and accessing it through retrieval. Although these stages fulfill different tasks, they still interact with each other. Indeed, the process of encoding does not simply determine if a piece of information enters the memory system but also impacts the way information is stored. For instance, it has been shown that deeper and more elaborate processing (semantic processing compared to more shallow forms of processing like visual or phonological processing) leads to better retention (e. g., Craik & Lockhart, 1972; Slamecka & McElree, 1983). In turn, consolidation, the time-dependent process whereby the encoded information is more firmly established and interconnected with other stored memories, can be actively enhanced by recovery phases. Especially sleep has been shown to benefit the retention of memories (e. g., Craik & Lockhart, 1972; Walker, Stickgold, Alsop, Gaab, Schlaug, 2005). Notably, there are millions of memories that are stored in our episodic memory system, whereas during retrieval we are usually seeking for one particular memory. But even if a memory is stored very firmly, without appropriate retrieval cues, this particular memory cannot be retrieved precisely. Thus, available retrieval cues are essential pieces of information that enable us to access our memories (e. g., Tulving & Thomson, 1973).

However, the human ability to retrieve information from episodic memory is not only about successfully reactivating and remembering particular memories; retrieval also crucially changes our episodic memories and can improve the ability to retrieve and use these memories again in the future. Experimental studies on the power of retrieval date back at least to Abbott (1909), who showed impressively that retrieval is an aid in the learning process. In particular, she demonstrated that the opportunity to retrieve the encoded information, during or immediately after learning, is of great benefit for later recall considering both nonsense and sense material. In doing so, she paved the way for Gates (1917), who introduced the so-called testing effect. Employing scores of grammar school students, he revealed that repetition generally increases long-term memory and that repetition format also influences the amount of enhancement. In fact, a great deal of studies on the testing effect in the past decade have agreed on his findings that long-term retention benefits more from retrieving previously studied information than from restudying it (e. g., Karpicke & Roediger, 2008; Roediger & Karpicke, 2006a).

Another eminent example for the fact that retrieval changes episodic memories is the effect of hypermnesia. It refers to the finding that recall rates increase across repeated tests within varying delays and without intervening opportunities to restudy the material (Ballard, 1913; for a review, see Payne, 1987). The research on hypermnesia especially accentuates the impact of retrieval on our episodic memories because the finding that new material is recovered on later tests that could not be recalled on earlier tests seems as calling into question the generality of one of the very first and most replicated experimental findings by Ebbinghaus in 1885, namely that we forget gradually over time.

However, retrieval of some information does not only affect the retrieved information, but also bears on related though not directly retrieved information. While the process of retrieval is generally beneficial for the retrieved information, there are two, at first sight contradictory branches of memory research on selective retrieval, which suggest "two faces of selective memory retrieval" (for a review, see Bäuml, Aslan, & Abel, 2017): On the one hand, it is assumed that recall of some previously studied information impairs recall of the related information and on the other hand, it is supposed that it enhances recall of the related information.

Evidence for the assumption that selective retrieval impairs the recall of related information has mainly arisen from studies on two different experimental designs: the older output-order task, and the more recent retrieval-practice task (for a review, see Bäuml & Kliegl, 2017). The general finding of experiments on the output-order task is that recall chances of studied items decline as a function of the items' serial position in the testing sequence, suggesting that the preceding recall of some items from a list can impair the recall of the remaining items (e. g., Roediger, 1974; A. D. Smith, 1971). Likewise, the typical result of experiments on the retrieval practice task is that, relative to an appropriate control condition, retrieval practice on a subset of previously studied items can cause forgetting of related unpracticed items in a subsequent memory test, suggesting that the repeated retrieval of some list items can impair the later recall of other related items (e .g., Anderson, Bjork, & Bjork, 1994).

In contrast to these studies, everyday experiences rather suggest that selective retrieval of some information has a positive effect on related material. For example, when talking to a family member about earlier life, remembering a particular event from childhood often leads to the recollection of more and more apparently forgotten memories. Context-retrieval theories (Greene, 1989; Thios & D'Agostino, 1976) are in line with the idea of a positive effect of selective retrieval on related material. Indeed, these theories assume that, when a previously studied item is repeated at a later point in time, be it by virtue of re-exposure or successful recall, the context in which it was originally exposed is retrieved. Such repetition is supposed to update the current state of context, which in turn may then serve as a retrieval cue for the recall of the remaining information. Empirical evidence from eyewitness memory research confirms the assumption that the active retrieval of some previously encoded episodes can benefit the memory of related episodes. Particularly, Geiselman, Fisher, MacKinnon, and Holland (1985) developed a special interrogation technique, the so-called Cognitive Interview, in which participants repeatedly recall and report every detail of a witnessed event they remember. The authors showed that the repeated recall of some details of an event can activate other (probably more important) details of the witnessed event.

Remarkably, in more recent studies, Bäuml and Samenieh (2010; 2012) were able to show the two faces of selective memory retrieval within one experimental setting. They suggested that whether selective memory retrieval impairs or enhances the recall of other material depends on access to the original study context. Indeed, when study context access was (largely) maintained, prior selective retrieval of some items from a list reduced the subsequent recall of the remaining items. In contrast, when study context access was impaired (e. g., after a prolonged delay or an implicit context change task), prior selective retrieval of some items improved the subsequent recall of the other items. To explain these two faces of selective memory retrieval, they proposed a two-factor account and suggested that the detrimental effect of selective retrieval is caused by inhibition and blocking processes (e. g., Anderson, 2003), whereas the beneficial effect of selective retrieval is called forth by context reactivation processes (e. g., Greene, 1989; Thios & D'Agostino, 1976).

Whereas there is well-founded empirical support in favor of the view that the detrimental effect of selective retrieval is mediated by inhibition and blocking processes (e. g., Storm & Levy, 2012), to date there is no direct evidence for the assumption that the beneficial effect of selective retrieval is mediated by context reactivation processes. Additionally, it is not clear if these presumably underlying context reactivation processes are modulated by repetition format. The first part of this thesis deals with these open questions on the beneficial effect of selective memory retrieval, thus providing a more detailed insight into the mechanisms modulating the effect. Experiment 1 aims at yielding more direct evidence for context reactivation processes as the underlying mechanism. Furthermore, Experiments 2A, 2B and 3 are designed to examine whether selective restudy, easy selective retrieval and difficult selective retrieval result in differently powerful beneficial effects on related material and, consequently, influence context reactivation processes.

The second part of this thesis is dedicated to open questions on the beneficial effects of memory retrieval in hypermnesia. Across tests, some information is recalled that was not recalled in prior tests (gains), whereas other information, recalled on prior tests, is not recalled in following tests (losses). If gains exceed losses across tests, hypermnesia arises. Thus, findings on hypermnesia indicate that retrieval can be beneficial for both, the already recalled information by reducing losses between tests, and the not yet recalled information by enhancing gains. To explain hypermnesia, many accounts have been suggested but none of them can account for the whole range of findings. One factor that may speak to at least some of these accounts is the delay between study and test. Also, because existing data on this issue do not provide consistent results, the possible role of delay between study and test for hypermnesia is addressed in Experiments 4-7. Hence, these experiments may provide a more detailed insight into the mechanisms that underlie hypermnesia.

Finally, the third part of this thesis stresses how the present findings on beneficial effects of memory retrieval contribute to the fact that retrieval is an active process that changes our episodic memories. Moreover, it is emphasized how the present findings broaden this view, as they provide important implications on the beneficial effect of selective memory retrieval on related material, as well as on the beneficial effect of repeated recall in hypermnesia. Notably, the present experiments may offer a new perspective for investigating and understanding the impact of memory retrieval in a more general way by shedding further light on how selective and nonselective retrieval benefit our episodic memories. Beyond that, the findings may allow interesting implications for educational and psycho-legal research considering their objective to improve students' and eyewitnesses' memory performance.

Contents

Parts of the present thesis are published as:

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.

Wallner, L., & Bäuml, K.-H. T. (2018). Hypermnesia and the role of delay between study and test. *Memory and Cognition*, 1-17. doi.org/10.3758/s13421-018-0809-5

Abstract

Numerous studies in the past decade have shown that active retrieval from episodic memory is able to boost retention. Retrieval of some studied information has been shown to improve both the recall of retrieved information and the recall of related though not retrieved information. This thesis investigates two effects of retrieval-induced remembering in more depth: the beneficial effect of selective memory retrieval of some item on other items (chapter 1) and the beneficial effect of repeated recall on later recall performance, i. e., hypermnesia (chapter 2). The beneficial effect of selective retrieval of some items from a list on the other items from the list has been attributed to context reactivation processes. However, to date there has been no direct evidence for this proposal. By showing the effect after impaired access to the study context but not when study context was reinstated, Experiments 1-3 provide the first direct evidence; they also indicate that the format of selective item repetition influences these context reactivation processes. Also, the mechanisms that underlie hypermnesia have not been clearly identified to date. Thus, Experiments 4-7 investigate the role of delay between study and test, which is predicted differently by the single accounts of hypermnesia. The results suggest to favor a retrieval practice explanation of hypermnesia by showing that hypermnesia increases with longer delays, at least after repeated free recall tests. Notably, the present results support the assumption of retrieval as an active process that increases retention and, above all, broaden this view as they deepen our understanding of the way how retrieval benefits our memories.

Chapter 1

Beneficial Effects of Selective Memory Retrieval

1.1 Effects of Nonselective and Selective Memory Retrieval

Beneficial Effects of Nonselective Retrieval

Everyday experiences show that our memory benefits from repetition. The more often we listen to a song, the better we know it by heart. The more often we use a foreign word, the better we memorize it. Supplementary, research on episodic memory has repeatedly shown that repetition of previously studied memories can aid the later retrieval of the same information and enhance the recall performance on a final test compared to a condition without any repetition opportunity (e. g., Bjork, 1975; Hogan & Kintsch, 1971).

The two prior mentioned examples suggest that different forms of repetition can improve our memory. When we repeatedly listen to a song, it is a form of restudy: We are presented an intact version of a previously studied information. In contrast, when we repeatedly use a prior encoded foreign word, it is a form of retrieval: We recall a previously studied information with or without the aid of retrieval cues. Research on beneficial effects of nonselective retrieval has shown that different repetition formats can vary in the amount of their beneficial effects. Of particular importance is the vast number of studies on the testing effect, i. e., the finding that retrieval practice on previously studied material can increase its long-term retention more than restudy of the information does (e. g., Karpicke & Roediger, 2008; for reviews, see Roediger & Butler, 2011; Rowland, 2014). The testing effect is a robust phenomenon that was demonstrated in lab-based studies and classroom settings, whereby there are important implications for both the study of memory and its application to educational practice. It was shown in a variety of list-learning experiments, for instance, employing unrelated single words (e. g., Carpenter & DeLosh, 2006; Hogan & Kintsch, 1971; McDaniel & Masson, 1985; Rowland & DeLosh, 2014; Zaromb & Roediger, 2010), associated word pairs (e. g., Allen, Mahler, & Estes, 1969; Carrier & Pashler, 1992; Pyc & Rawson, 2010; Toppino & Cohen, 2009), pictures (e. g., Wheeler & Roediger, 1992), foreign language vocabulary (e. g., Pyc & Rawson, 2010), and nonverbal materials (e. g., Carpenter & Pashler, 2007; Kang, 2010). It arose with prose passages (e. g., Glover, 1989; Roediger & Karpicke, 2006a) and films (e. g., Bornstein, Liebel, & Scarberry, 1998).

In addition to this direct effect, there is further research that is in line with the view of beneficial effects of memory retrieval by showing that retrieval can indirectly enhance memory (in comparison to restudy; for a review see Roediger et al., 2011). For example, it has repeatedly been revealed that retrieval can generate better transfer of the study material, and thus increase organization of newly acquired knowledge compared to restudy (see Congleton & Rajaram, 2011; Masson & McDaniel, 1981). Another benefit of retrieval is the effect of test-potentiated learning (Izawa, 1966), i. e., the finding that attempting to retrieve items improves later encoding of those items. Supplementary, Szpunar, McDermott, and Roediger (2008) showed that retrieval of previously studied material increases long-term retention of subsequently studied material, an effect termed as the forward effect of retrieval practice (see Pastötter & Bäuml, 2014).

This impressive body of evidence demonstrates that repetition, and especially retrieval practice, enhances long-term retention. Largely, the referred studies on beneficial effects of retrieval practice employed conditions in which *all* of the previously studied material should be repeated. However, if we envision our everyday-life, we rather will encounter situations in which only a fraction of the originally experienced episode should be retrieved. For example, imagine you are talking about your working day at dinner. Do you tell every detail you experienced at work, or do you rather extract some experiences? Alternatively, conceive of an interrogation at school. Does the teacher ask every detail of the previous school lesson, or does he rather ask a subset of the learning matter? It seems likely that selective retrieval creates similar beneficial effects on this retrieved subset of the prior encoded information as nonselective retrieval does on the entire prior encoded information. But it is not that much obvious whether and how selective retrieval impacts later retention of the nonretrieved information, i. e., the experiences of a working day not mentioned at dinner or the details of the learning matter not interrogated by the teacher. As described below, on the basis of numerous studies there are good reasons to expect that, under certain circumstances, selective retrieval leads to forgetting of the nonretrieved information while, under other circumstances, it leads to facilitation of the nonretrieved information.

Detrimental Effects of Selective Retrieval

Research in the past five decades has shown that selective retrieval of a subset of the previously studied information can impair recall of related information. Such detrimental effects of selective retrieval have been observed in numerous studies using both the output-interference and the retrieval-practice task (for a review, see Bäuml & Kliegl, 2017). The first studies employing the output-interference task arose in the 1960s and 1970s, examining whether recall of studied items varies as a function of the items's serial position in the output sequence at test. In the first study on this issue of Tulving and Arbuckle (1963), participants had to learn a list of paired associates consisting of a cue item (a single digit between 0 and 9) and a target item (a common word) in a counterbalanced order. In the following cued recall test, participants had to remember the target words, again in a counterbalanced order. Recall performance of a given item declined steadily with its output position. Thus, they were the first to show that prior selective recall of some list items can impair the subsequent recall of other list items. The finding of output-interference, i. e., decreased recall performance of to be recalled information conditioned by interference with prior recalled information, was replicated by many studies and has been shown employing different experimental settings and study materials (for a review, see Roediger, 1974).

A more recent task to study the effects of selective retrieval is the retrieval-practice task, which was introduced into the literature by Anderson et al. (1994). A typical retrieval-practice task consists of three phases: (i) a study phase, in which subjects are asked to study a list of items, (ii) a retrieval practice phase, in which participants retrieve only a subset of the studied items, and (iii) a final test phase, in which all studied items are to be retrieved. In the pioneer study of Anderson et al. (1994), participants had to study items from different semantic categories before performing selective retrieval practice on half of the items of half of the categories. At retrieval practice, they were given the superordinated category (that also had been presented in the study phase) plus the word stems as cues. After a delay, participants were asked to recall all previously studied items. As a result, retrieval practice improved recall of the practiced items but impaired recall of the unpracticed items from the practiced items' categories relative to the control items from the unpracticed categories. Hence, it was demonstrated that selective retrieval of some items can induce forgetting of related items. Like output-interference, retrieval-induced forgetting is a very robust and general finding that was reported over a wide range of materials and settings, and a variety of testing formats (for reviews, see Bäuml & Kliegl, 2017; Storm & Levy, 2012; for a recent meta analysis, see Murayama, Miyatsu, Buchli, & Storm, 2014).

Beneficial Effects of Selective Retrieval

However, more recent work has demonstrated that selective memory retrieval can not only impair but also improve the recall of other items. First corresponding evidence has come from studies examining the effects of selective retrieval in listwise directed forgetting and context-dependent forgetting. In the studies on listwise directed forgetting, subjects studied an item list and after study received a cue to either remember or forget the list (e. g., Bjork, 1972). After study of a second list, they recalled some predefined first list target items, either first or after prior selective retrieval of the list's remaining items. As expected from the literature on retrieval-induced forgetting, selective retrieval impaired recall of the target items in the remember condition. In the forget condition, however, selective retrieval improved target recall (Bäuml & Samenieh, 2010, 2012). The same pattern of results arose in context-dependent forgetting, when between study of two item lists, subjects either participated in a neutral counting task or engaged in an imagination task to change their internal context (e. g., Sahakyan & Kelley, 2002). Again, at test, subjects selectively retrieved some of the first list items before they recalled the list's target items, or recalled the target items first. Selective retrieval impaired recall of the target items after the counting task, but improved target recall after the imagination task (Bäuml & Samenieh, 2012; Schlichting, Aslan, Holterman, & Bäuml, 2015).

Two faces of selective retrieval have also been found in studies on time-dependent forgetting. In these studies, participants studied a list of items and, after a short retention interval of few minutes or a prolonged retention interval of 48 hrs, were again asked to recall predefined target items of the list. These target items were recalled first or after prior selective retrieval of the list's remaining items. Consistent with the literature on retrieval-induced forgetting, selective retrieval impaired recall of the target items after the short retention interval. In contrast, in the prolonged retention interval conditions, selective retrieval improved recall of the target items (Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014). These findings fit with the results from the studies on context-dependent forgetting mentioned above, because prolonged retention intervals typically include a considerable amount of contextual change between study and test (e. g., Bower, 1972; Estes, 1955; Mensink & Raaijmakers, 1988). Besides, like the detrimental effects of selective retrieval, also the beneficial effects of selective retrieval were found with both the output-interference task (e. g., B äuml & Samenieh, 2010, 2012) and the retrieval- practice task (e. g., Bäuml & Dobler, 2015; Dobler & Bäuml, 2012).

Together, all of these results demonstrate that retrieval dynamics can depend critically on situation and selective retrieval can both impair and improve recall of other items. More precisely, it is assumed that whether selective retrieval of some list items generates detrimental or beneficial effects for the recall of related list items depends on access to the list's original study context during selective retrieval (for a recent review on these findings, see Bäuml et al., 2017).

1.2 Theoretical Accounts of the Effects of Retrieval

Theoretical Accounts of the Effects of Nonselective Retrieval

Like described above, numerous studies have confirmed the assumption that recall performance on a final test is higher when individuals beforehand get a chance to study or retrieve the entire encoded material, compared to a control condition without any repetition opportunity (e. g., Bjork, 1975; Hogan & Kintsch, 1971). A great deal of those studies in the last decade have focused on the testing effect (e. g., Karpicke & Roediger, 2008; Pastötter et al., 2011; Szpunar et al., 2008).

There are numerous accounts for the testing effect (for a review see Delaney, Verkoijen, & Spirgel, 2010; Roediger et al., 2011; Roediger & Karpicke, 2006a). The probably most common one is the elaboration account proposed by Carpenter (2009), which generally is committed to the idea that retrieval represents an effective opportunity for elaborative processing. Carpenter assumes that the active search process during retrieval of a target information activates information semantically related to the target information and that, in turn, this additional information leads to more cues which facilitate later retrieval. While the elaboration account of the testing effect is mainly based on semantic retrieval processes, there is another important account that is primarily based on episodic retrieval processes.

Recently, Karpicke, Lehman, and Aue (2014) proposed the episodic context account as an alternative explanation for the testing effect. Importantly, this account is based on some central ideas descending from formal memory models (e. g., Howard & Kahana, 2002; Mensik & Raaijmakers, 1989; Raaijmakers & Shiffrin, 1981). According to these formal memory models, events occur within a slowly changing representation of episodic context, and during study, these temporal, spatial and situational context features can get linked to the encoded information. Besides, during the process of repetition of the prior encoded information, be it by virtue of reexposure or its successful recall, contextual cues that are stored disposable in the present are used to aid remembering information encoded in the past. Reinstating a prior episodic context thus is an essential process during retrieval.

Accordingly, Karpicke et al. (2014) suggest that context reinstatement during repetition creates a unique set of context features for each restudied or successfully retrieved information. Hence, at repetition the past study context is retrieved and associated with the present repetition context so that an extended composite of features from both contexts emerges. On a later test, reinstatement of either context can serve as an effective cue for the to be remembered information, increasing successful retrieval. Importantly, retrieval can be more effective than restudy to reactivate the study context. This may be the case because with intentional recall instructions, i. e., during retrieval trials, context retrieval may be obligatory, whereas in the absence of such instructions, i. e., during restudy trials, it may not.

A Two-Factor Account of the Effects of Selective Retrieval

Also on the basis of the above described formal memory models (e. g., Howard & Kahana, 2002; Mensik & Raaijmakers, 1989; Raaijmakers & Shiffrin, 1981) and the assumption arising from the so-called *encoding specificity principle* (e. g., Thomson & Tulving, 1970; Tulving & Thomson, 1973), that recall is most effective when the context at the time of encoding matches the context at the time of retrieval, Bäuml and Samenieh (2012) suggested a two-factor account to explain why selective retrieval is sometimes beneficial and sometimes detrimental for other memories. As the two faces of selective retrieval have been found in studies on listwise directed forgetting (Bäuml & Samenieh, 2010, 2012), on context-dependent forgetting (Bäuml & Samenieh, 2012; Schlichting et al., 2015), as well as on time-dependent forgetting (Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014), it has to be alluded that in all of these forms of forgetting, context change between study and test may play an important role, be it in terms of a cue to forget previously studied items, an external or internal context change task, or a prolonged retention interval (e. g., Estes, 1955; Geiselman, Bjork, & Fishman, 1983; McGeoch, 1932; Mensink & Raaijmakers, 1988; Sahakyan & Kelley, 2002; Smith & Handy, 2014).

According to Bäum and Samenieh's (2012) account, selective retrieval generally triggers two types of processes, inhibition or blocking of interfering memories (e. g., Anderson, 2003; Roediger & Neely, 1982) and context reactivation (e. g., Howard & Kahana, 2002; Raaijmakers & Shiffrin, 1981). Critically, the relative contribution of the two types of processes in an experimental situation is supposed to depend on access to study context at test. When access to the study context is (largely) maintained - as may occur after a remember cue or a short retention interval filled with a neutral distractor task - then interference between items may be high enough to trigger inhibition or blocking processes, whereas there is little or no need to reactivate study context during retrieval. As a net result, selective retrieval may reduce recall of the remaining items. In contrast, when access to the study context is impaired and the interference level of the items is low - as may occur after a forget cue, an imagination task, or a prolonged retention interval - then access to the study context may benefit from retrieval-induced context reactivation processes, with inhibition or blocking processes hardly operating. The reactivated study context may then serve as a retrieval cue for recall of the remaining items and thus improve recall performance.

The empirical support in favor of the view that the detrimental effect of selective retrieval is mediated by inhibition and blocking processes is currently much stronger than is the evidence for the view that the beneficial effect is mediated by context reactivation. Indeed, findings on retrieval-induced forgetting strongly indicate that the detrimental effect is mediated by inhibition and blocking processes. While neither inhibition nor blocking seem to be able to explain the whole range of findings on the detrimental effect in its own, the assumption that inhibition and blocking conjointly contribute to the effect may explain the main findings (e. g., Bäuml & Kliegl, 2017; Storm & Levy, 2012; but see Jonker, Seli, & MacLeod, 2013). The proposal that context reactivation processes mediate the beneficial effect of selective retrieval is less well supported by data. Rather, current evidence for the proposal is fairly indirect, for instance, revealing a developmental trajectory of the beneficial effect that fits with the suggested development of context reactivation processes in children and older adults (e. g., Aslan & Bäuml, 2014; Aslan, Schlichting, John, & Bäuml, 2015). Therefore, this chapter aimed to fill this gap and come up with more direct evidence that context reactivation processes mediate the beneficial effect of selective retrieval (Experiment 1, see below).

An Extended Version of the Two-Factor Account

The two-factor account can explain a relatively wide range of findings on the beneficial and the detrimental effects of selective memory retrieval. Nevertheless, a so far unacknowledged question on the beneficial effect of selective retrieval is whether it is retrieval specific, that is, whether it is restricted to selective retrieval trials or alternatively generalizes to selective restudy trials. If both selective retrieval and selective restudy, induced beneficial effects for nonrepeated items, the two-factor account would be broadened to account for the effects of selective item repetition in general. Results from numerous studies on retrieval-induced forgetting indicate that the detrimental effect of selective retrieval is largely retrieval specific. Comparing the effects of selective retrieval and selective restudy on later recall of related unpracticed items, these studies typically found retrieval practice, but not restudy, to impair recall of the unpracticed items (e. g., Bäuml, 2002; Ciranni & Shimamura, 1999; Hulbert, Shivde, & Anderson, 2012; for exceptions, see Raaijmakers & Jakab, 2012, or Verde, 2013). Retrieval specificity of the detrimental effect of selective retrieval is consistent with the view that inhibition critically contributes to the effect. According to this view, the not-to-be practiced items interfere during selective retrieval, but not during selective restudy, and are inhibited to reduce the interference (Anderson, 2003; for a more detailed discussion of retrieval specificity of the detrimental effect, see Rupprecht & Bäuml, 2016, 2017).

The question of whether the beneficial effect of selective retrieval is also retrieval specific has hardly been investigated yet. Bäuml and Dobler (2015) addressed the issue in two experiments, in which they compared the effects of selective retrieval and selective restudy on the recall of other items when access to study context was (largely) maintained and when access to study context was impaired. Experiment 1 employed listwise directed forgetting to manipulate study context access and asked subjects to either remember or forget a previously studied list; Experiment 2 employed time-dependent forgetting to manipulate context access and varied the retention interval after study (4 min vs. 48 hrs). In both experiments, subjects selectively retrieved or selectively restudied some of the studied items before they recalled the list's target items, or they recalled the target items in the absence of any prior selective item repetition. Consistent with the previous studies on retrieval specificity of retrieval-induced forgetting, the results of both experiments showed that selective retrieval, but not selective restudy, impaired recall of the other items when access to study context at test was maintained. In contrast, when context access was impaired, both selective retrieval and selective restudy enhanced the recall of the other items, indicating that, unlike the detrimental effect, the beneficial effect of selective retrieval is not retrieval specific, which is consistent with both the context-retrieval theory and the two-factor account of selective retrieval.

The findings by Bäuml and Dobler (2015) fit with the two-factor account and the comprised view that the beneficial effect is driven by reactivation of the retrieved items' study context (e. g.,Howard & Kahana, 2002; Greene, 1989; Polyn, Norman, & Kahana, 2009; Thios & D'Agostino, 1976). While the two-factor account together with context retrieval theory can thus explain the finding that both selective retrieval and selective restudy can improve recall of other information, the question arises of whether selective repetition by virtue of retrieval and selective repetition by virtue of restudy are really equivalent. While context retrieval theory claims that both retrieval and restudy can trigger context reactivation, the theory is largely silent on whether the two forms of item repetition differ in degree of the induced reactivation. In their episodic-context account of the testing effect, Karpicke et al. (2014) suggested a variant of context retrieval theory, which assumes that different forms of item repetition can differ in context reactivation.

In this variant of the theory, Karpicke et al. (2014) made two core assumptions. The one assumption is that retrieval can be more effective than restudy to reactivate the study context (for detailed reasons, see above). The second assumption is that retrieval difficulty can influence context reactivation, with more difficult retrieval (i. e., retrieval in the presence of weak item-specific cues) creating more context reactivation than easy retrieval (i. e., retrieval in the presence of strong item-specific cues). In fact, difficult retrieval may require subjects to reinstate a prior context with minimal cues, and such effortful reconstruction of the study context may drive the gains in learning compared to easy retrieval. In some cases, easy retrieval may even be more semantic than episodic in nature, thus inducing hardly any context reactivation at all.

The suggested difference in context reactivation between retrieval and restudy conditions can explain the basic testing effect finding that retrieval leads to better final recall than restudy does, attributing the effect to a difference in the creation of unique context cues between the two repetition formats. In a similar way, the suggested difference in context reactivation between difficult and easy retrieval can explain the finding that difficult retrieval often creates a larger testing effect than easy retrieval (Carpenter & DeLosh, 2006; Halamish & Bjork, 2011; Pyc & Rawson, 2009; for alternative explanations of the finding, see also Halamish & Bjork, 2011). If context reactivation processes do not only contribute to the beneficial effects of retrieval and restudy on repeated items (Karpicke et al., 2014), but also mediate the beneficial effects of selective retrieval and selective restudy on recall of nonrepeated items (e. g., Bäuml & Dobler, 2015), then, following Karpicke et al., repetition format may not only influence the beneficial effect of item repetition on recall of the repeated items but may also influence this beneficial effect of selective item repetition on recall of the nonrepeated items as well.

Whether repetition format influences the beneficial effect of selective item repetition on nonrepeated items is unclear to date. To the best of my knowledge, there is no study yet that examined whether the beneficial effect of selective retrieval depends on how demanding retrieval is. In fact, all previous studies employed word stems as retrieval cues in the selective retrieval trials, examining how this affects recall of the nonrepeated items (see Bäuml et al., 2017). Supplementary, there is only a single study in the literature that compared the size of the beneficial effects of selective retrieval and selective restudy on nonrepeated items (Bäuml & Dobler, 2015). Employing listwise directed forgetting and time-dependent forgetting to impair study context access at test (see above), this study reported equivalent beneficial effects of selective retrieval and selective restudy after a forget cue, but a larger beneficial effect of selective retrieval than selective restudy after a prolonged retention interval. These results leave it open whether, in general, the beneficial effect of selective item repetition on nonrepeated items varies with repetition format, and whether Karpicke et al.'s (2014) variant of context retrieval theory can be applied to explain the beneficial effects of selective item repetition. It is another goal of this chapter to provide an answer on this issue (Experiment 2 and 3, see below).

1.3 Goals of Experiments 1-3

As emphasized above, it is the first goal of this chapter to examine the proposal included in Bäuml and Samenieh's (2012) two-factor account of selective retrieval that context reactivation processes mediate the beneficial effect of selective retrieval more directly. Although the proposal has proven consistency with several lines of findings on the two faces of selective retrieval (e. g., Aslan & Bäuml, 2014; Bäuml & Dobler, 2015; Bäuml & Samenieh, 2012; Bäuml & Schlichting, 2014), to date there is no direct evidence yet for this theoretical position. More direct evidence for the proposal would arise from an experiment, in which, after inducing impaired study context access - for instance, by increasing the retention interval between study and selective retrieval - participants' study context was mentally reinstated immediately before selective retrieval starts. (Partial) reinstatement of the study context should reduce the need for further retrieval-induced context reactivation processes and, following the two-factor account, should thus reduce or eliminate the beneficial effect of selective retrieval. Moreover, if reinstatement of the study context was complete, even detrimental effects of selective retrieval should arise. In fact, a complete reinstatement of the study context should also reinstate the items' interference level and thus trigger inhibition and blocking, leading to retrieval-induced forgetting. The issue was addressed in Experiment 1.

Subjects studied a list of unrelated items and after a prolonged retention interval, chosen to impair study context access (e. g., Estes, 1955), were asked to recall predefined target items of the list first or after prior selective retrieval of the list's other items. Immediately before recall started, subjects engaged in mental reinstatement of the study context or a neutral distractor task to leave context largely unaffected (e. g., Jonker et al., 2013; Sahakyan & Kelley, 2002). On the basis of the two-factor account and the comprised view that context reactivation processes underlie the beneficial effect of selective retrieval, the typical beneficial effect of selective retrieval was expected when the retention interval was prolonged and study context was not reinstated. In contrast, when study context was reinstated, no such beneficial effect was expected and selective retrieval might even impair recall of the other items.

To anticipate the results of Experiment 1, it will provide clear evidence that context reactivation processes mediate the beneficial effect of selective retrieval. On the basis of this result and the view that context reactivation processes mediate the beneficial effects of both selective retrieval and selective restudy (Bäuml & Dobler, 2015), it is the second goal of this chapter to examine the proposal that format of selective item repetition - difficult retrieval versus easy retrieval versus restudy - can influence repetition-induced context reactivation processes, and thus can influence the beneficial effects of selective item repetition on recall of the nonrepeated items. Following the two-factor account and Karpicke et al.'s (2014) variant of context retrieval theory, selective retrieval should induce stronger beneficial effects than selective restudy, and more difficult selective retrieval should induce stronger beneficial effects than easy selective retrieval. The issue was examined in three experiments (Experiments 2A, 2B, and 3) that employed different study materials and prolonged retention intervals of different length to vary the degree of study context access. The results of the experiments will provide important information on the role of repetition format for the beneficial effect of selective item repetition. In particular, together with the results of Experiment 1, they will offer a more conclusive picture of the role of context reactivation processes for the effects of selective item repetition.

In these experiments, subjects studied a list of unrelated items (Experiments 2A and 2B) or some more coherent prose material

(Experiment 3) and after a prolonged retention interval were asked to recall some predefined target items of the original study material. The target items were recalled first or after prior selective repetition of some of the material's other (nontarget) information. Repetition of the nontarget items occurred through restudy of the items, retrieval of the items in the presence of strong item-specific cues (easy retrieval), or retrieval of the items in the presence of weak item-specific cues (difficult retrieval).

1.4 Experiment 1: The Role of Study Context Reinstatement for the Beneficial Effect of Selective Retrieval

Experiment 1 aimed to come up with a rather direct test of the proposal that context reactivation processes mediate the beneficial effect of selective retrieval on recall of the nonretrieved items. In this experiment, subjects studied a list of unrelated items and, after a retention interval of 10 min, which included an imagination task to enhance contextual drift and the impairment in study context access (e. g., Bower, 1972; Estes, 1955), were asked to recall predefined target items of the list, either first or after prior selective retrieval of the list's remaining (nontarget) items. Immediately before recall started, two different testing conditions were induced. In the one testing condition, subjects' study context was (partially) reinstated by employing a mental context reinstatement technique. Subjects were asked to mentally reinstate their original list learning environment, and to recall and write down in brief phrases what they were doing prior to the study phase (e. g., Jonker et al., 2013; Sahakyan & Kelley, 2002). In the other testing condition, no such context reinstatement took place. On the basis of the two-factor account and the view that context reactivation mediates the beneficial effect of selective retrieval, selective retrieval was expected to improve recall of the other items when there was no preceding mental reinstatement of the study context. In contrast, when study context was mentally reinstated before recall started, there should be little or no need for further retrieval-induced context reactivation, and therefore, the beneficial effect of selective retrieval was expected to be reduced, if not reversed (see also Tab. 1).

Table 1. *Overview of Experiments 1-3: Methods, Predictions, and Results.*

	Methods	**Predictions**	**Results**
Exp. 1	(Easy) SR Study context reinstatement before test or no such context reinstatement 10 min delay Unrelated words	Reduced, if not reversed, beneficial effect of SR after study context reinstatement	Prediction confirmed
Exp. 2A	Easy or difficult SR or SS 10 min or 30 min delay Unrelated words	Equivalent beneficial effects after 10 min Larger beneficial effect after SR than SS, and after difficult than easy SR after 30 min	Most predictions confirmed But: equivalent beneficial effects of difficult and easy SR after 30 min
Exp. 2B	Easy or difficult SR or SS Control of number of successfully repeated items 24 hrs delay Unrelated words	Larger beneficial effect after SR than SS, and after difficult than easy SR	Predictions confirmed
Exp. 3	Easy or difficult SR Control of number of successfully repeated items 30 min delay Text passage	Larger beneficial effect after difficult than easy SR	Predictions confirmed

Note. SR = selective retrieval; SS = selective restudy.

Method

Participants. 48 students of Regensburg University participated in the experiment (M= 22.60 years, range = 19-29 years, 70.8% female). They were equally distributed across the two between-subjects conditions, resulting in n=24 participants in each condition. Sample size was based on prior work examining beneficial effects of selective memory retrieval (e. g., Bäuml &

Samenieh, 2012; Bäuml & Schlichting, 2014). All subjects spoke German as native language and received monetary reward or course credit for their participation.

Materials. Two study lists (A, B) were constructed, each containing 15 unrelated concrete German nouns. The items were drawn from the larger set of items used in Bäuml and Dobler (2015; see Appendix A). Half of the participants studied List A, the other half List B. Each of the two lists consisted of 5 predefined target and 10 predefined nontarget items. Among all items within a list, each target and each nontarget item had a unique initial letter.

Design. The experiment had a 2 × 2 design with the within-subjects factor of SELECTIVE RETRIEVAL (present, absent) and the between-subjects factor of CONTEXT REINSTATEMENT (present, absent). In the condition with context reinstatement, participants were instructed immediately before the test started to recall and write down details of what they were doing immediately prior to the study phase, whereas participants in the condition without context reinstatement were engaged in neutral filler tasks. Selective retrieval conditions differed in whether participants were asked to retrieve the target items first or after prior retrieval of the list's nontarget items.

Procedure. Each participant completed two experimental blocks. To provide a distinctive initial experimental context at the beginning of each experimental block, in all experimental conditions participants initially rated a list of pictures (nice places, food). Each participant evaluated in the one block how attractive the provided places were to them for traveling in future, and indicated in the other how much they liked the presented food's taste. Each of the two rating tasks lasted for 3 min. Subsequently, the study phase started and in each block, the items of one study list were exposed individually in the center of a computer screen and in random order for 5 s each. A retention interval of 10 min followed, in which participants were engaged in several distractor tasks, which included counting backward from a three-digit number, resolving decision tasks, and doing one of two imagination tasks. The imagination tasks lasted 3 min each. Subjects were either asked to imagine their parents' house and mentally walk through it, or they were asked to imagine being back on their last vacation and to remember and re-feel the most beautiful moments as intensively as possible (e. g., Delaney, Sahakyan, Kelley, & Zimmerman, 2010; Sahakyan & Kelley, 2002). They were also asked to write down their imaginations. After the retention interval, subjects who participated in the mental context reinstatement conditions were told to take 1 min to recall and write down as detailed as possible their thoughts, feelings, or emotions while rating the pictures from the initial phase. In addition, they should try to remember the strategies they used in the study phase. Subjects who did not participate in the context reinstatement condition solved arithmetic problems for the same time period.

At test, recall order of target items was controlled through the presentation of the items' unique initial letters, which were presented successively and in random order for 6 s each. Responses were given orally. Target items were either tested first or after selective retrieval of the nontarget items. Nontargets were retrieved successively for 6 s each and in two successive cycles, each with its own random order. The nontargets' word stems were provided as retrieval cues. Subjects who recalled the target items first in the first experimental block recalled the target items after prior recall of the nontarget items in the second block, and vice versa. Overall, assignment of rating tasks, imagination tasks, lists and conditions was counterbalanced (see Fig. 1).

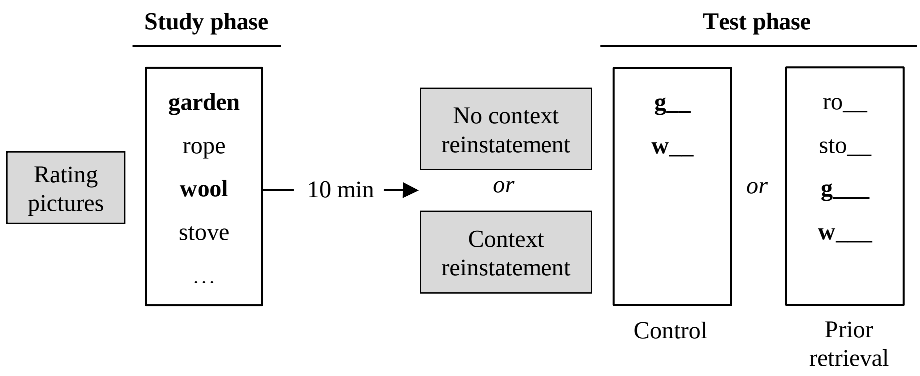

Figure 1. *Procedure and conditions employed in Experiment 1. Participants rated a list of pictures (food/nice places) and then studied a list of words. After a delay of 10 min, half of the participants engaged in mental reinstatement of the study context (context reinstatement), while the other half solved arithmetic problems as a control (no context reinstatement). At test, participants were asked to recall predefined target items from the list (e. g., garden, wool). The targets were tested first (control) or after prior recall of the list's remaining items (e. g., robe, stove; prior retrieval). Predefined target items are depicted in bold letters.*

Results

Fig. 2 shows mean recall rates for the target items. A 2 × 2 analysis of variance with the within-subjects factor of SELECTIVE RETRIEVAL (present, absent) and the between-subjects factor of CONTEXT REINSTATEMENT (present, absent) showed a significant interaction between the two factors, $F(1, 46) = 47.69$, $MSE = 236.23$, $p < .001$, $\eta^2 = 0.51$. There was no main effect of CONTEXT REINSTATEMENT, $F(1, 46) < 1$, but a main effect of SELECTIVE RETRIEVAL, $F(1, 46) = 4.52$, $MSE = 236.23$, $p = .039$, $\eta^2 = 0.09$, indicating that target recall was influenced by whether the nontarget items were previously retrieved. Planned comparisons showed that, whereas preceding selective retrieval improved target recall in the absence of context reinstatement, $t(23) = 6.82$, $p < .001$, $d = 1.35$, it impaired target recall in the presence of context reinstatement, $t(23) = 3.19$, $p = .004$, $d = 0.66$. [2] Access to study context at test thus modulated the effect of selective retrieval.

Half of the participants in this experiment started testing with target items being recalled first and the other half with target items being recalled last. Testing order did not affect results, however. There was no main effect of testing order, $F(1, 44) = 1.85$, $MSE = 729.55$, $p = .181$, $\eta^2 = 0.04$, and no interaction of testing order with any of the other factors, all $Fs(1, 44) < 1.76$, $MSEs > 237.12$, $ps > .192$, $\eta^2 s < 0.04$. Further analyses showed that, if no prior selective retrieval took place, target recall was higher when context reinstatement was present than when it was absent, $t(46) = 3.97$, $p < .001$, $d = 1.15$, thus showing the typical context reinstatement effect. [3] Nontarget recall was high (reinstatement: 73.75%; no reinstatement: 79.58%) and did not vary between reinstatement conditions, $t(46) = 1.35$, $p = .183$, $d = 0.15$.

---

[2] Note that all reported effect sizes of t-tests are Cohen's (1977) *d statistics*, and thus appropriate standardized levels of the effects, irrespective of whether the design was within or between measures (for a discussion of computing effect size appropriately from matched groups or repeated measures designs effect, see Dunlap, Cortina, Vaslow, and Burke, 1996).

[3] Among other things, the results of Experiment 2 reported below show that, after a 10-min delay and without preceding context reinstatement, selective retrieval eliminates time-dependent forgetting. If, under the same conditions, selective retrieval also eliminated time-dependent forgetting in Experiment 1, then the recall level in the prior retrieval-no context reinstatement condition of this experiment can serve as a baseline to determine how effective the context reinstatement was. The finding that, after context reinstatement, the recall level in the control condition was nearly indistinguishable from the recall level in the prior retrieval-no context reinstatement condition (see Fig. 2) then indicates that the context reinstatement was more or less complete in the present experiment.

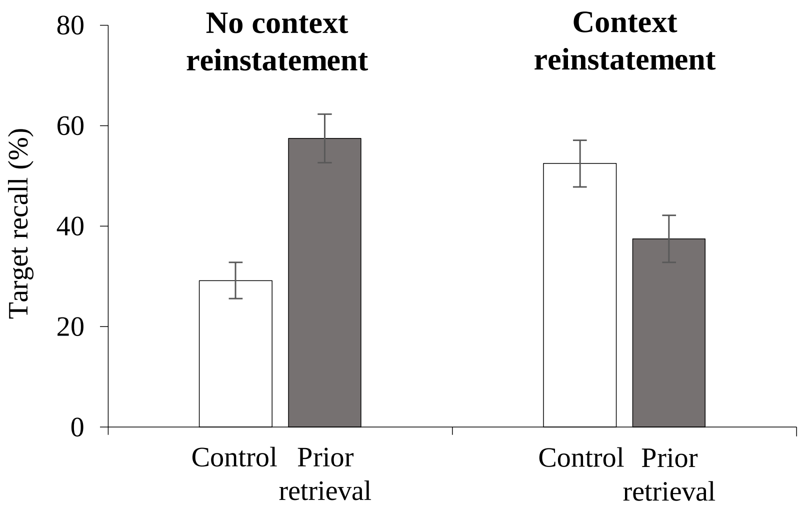

Figure 2. *Results of Experiment 1. Percentage of recalled target items is shown as a function of delay (10 min, 30 min) and repetition format (control, prior restudy, prior easy retrieval, prior difficult retrieval). Percentage of recalled target items is shown for the control and prior retrieval conditions, in the presence and absence of mental context reinstatement. Error bars represent standard errors.*

Discussion

The results replicate prior work (Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014) by demonstrating a beneficial effect of selective retrieval after a prolonged retention interval when there was no mental reinstatement of the study context before recall started. In contrast, for the same retention interval, the results showed a detrimental effect of selective retrieval when study context was mentally reinstated before recall started. These results underline the critical role of study context access for the beneficial effect of selective retrieval and indicate that, with constant study and selective retrieval conditions, selective retrieval can both improve and impair recall of other items, depending on whether study context access at test is impaired or not. The finding strongly supports the proposal that context reactivation processes mediate the beneficial effect of selective retrieval and thus supports the two-factor account of selective retrieval.

1.5 Experiment 2A: The Role of Repetition Format

Experiment 2A was aimed as a first step to investigate whether the context reactivation processes that supposedly underlie the beneficial effect of selective retrieval vary with repetition format. Following the two-factor account and Karpicke et al.'s (2014) variant of context retrieval theory, repetition format may influence the effects of selective item repetition on the recall of the nonrepeated items: selective retrieval may induce a higher degree of context reactivation and thus a stronger beneficial effect for nonrepeated items than selective restudy, and more difficult selective retrieval may induce a higher degree of context reactivation and thus a stronger beneficial effect for nonretrieved items than easy selective retrieval. In addition, Experiment 2A investigated whether such relationship would depend on the extent to which access to study context at test is impaired, so that the single repetition formats may create different beneficial effects when study context access is strongly impaired at test, but largely equivalent beneficial effects when the impairment is only moderate. Such finding could explain the discrepancy in results in the Bäuml and Dobler (2015) study, which reported equivalent beneficial effects of retrieval and restudy after moderate episodic forgetting (after a forget cue) but stronger beneficial effects of retrieval than restudy after strong episodic forgetting (after a 48-hrs retention interval).

Participants studied a list of unrelated words and after study were engaged in several distractor tasks. There were a shorter (10 min) and a longer (30 min) retention interval condition, with one (10-min condition) or three (30-min condition) imagination tasks included to enhance contextual drift. At test, participants in both delay conditions were asked to recall predefined target items of the list first or after prior selective repetition of the list's remaining (nontarget) items. There were three repetition conditions: in the restudy condition, subjects restudied the nontarget items; in the easy retrieval condition, they retrieved the nontarget items with the item's unique word stems as retrieval cues; in the difficult retrieval condition, they retrieved the nontarget items with the words' unique initial letters as retrieval cues). On the basis of the two-factor account of selective retrieval and Karpicke et al.'s variant of context retrieval theory, all three repetition formats were expected to induce beneficial effects on the nonrepeated items, after both the shorter and the longer retention interval. In particular, the three repetition formats were expected to differ in amount of the beneficial effect, with selective retrieval inducing a larger beneficial effect than selective restudy, and more difficult retrieval inducing a larger beneficial effect than easier retrieval. On the basis of the findings of Bäuml and Dobler (2015), however, one may expect this pattern to arise mainly for the longer (30-min) retention interval (see also Tab. 1).

Method

Participants. Another 192 students of Regensburg University took part in the experiment (M= 23.00 years, range: 18-33 years, 87.0% female). They were equally distributed across the eight between-subjects conditions, resulting in n=24 participants in each single condition. Sample size followed Experiment 1. All participants spoke German as native language and took part on a voluntary basis. They received monetary reward or course credits for their participation.

Materials. Materials were identical to Experiment 1.

Design. The experiment had a 2 × 4 factorial design with the between-subjects factors of REPETITION FORMAT (control, prior restudy, prior easy retrieval, prior difficult retrieval) and DELAY (10 min, 30 min). Between study of the list and the test phase participants took part in several distractor tasks that lasted either 10 min or 30 min. At test, subjects recalled the target items first (control) or after restudy of the nontarget items (prior restudy) or after retrieval of the nontarget items, with either the items' unique initial letters serving as retrieval cues (prior difficult retrieval) or their word stems serving as retrieval cues (prior easy retrieval). Assignment of conditions and lists was counterbalanced.

Procedure. List items were exposed on a screen individually and in random order for 5 s each. In the 10-min retention interval condition, participants were engaged in one block of distractor tasks; in the 30-min retention interval condition, they were engaged in three successive 10-min blocks of distractor tasks. In each block, subjects were first asked to participate in a number of tasks, like counting backward from a three-digit number, solving arithmetic problems, resolving some decision tasks, playing tetris, doing the Ravens Progressive Matrices, or rating pictures of food or nice places, for a total of 7 min. Afterwards, subjects were engaged in an imagination task of 3 min duration. In this task, participants were either asked to imagine their parents' house and to mentally walk through it, or to imagine the things they would like to do if they were invisible and did not have to take responsibility for their actions, or to imagine being back on their last vacation and to remember and re-feel the most beautiful moments as intensively as possible (e. g., Delaney, Sahakyan et al., 2010; Sahakyan & Kelley, 2002). They were also asked to write down their imaginations. For the 30-min delay condition, all three imagination tasks were employed in random order; for the 10-min delay condition, one of the tasks was randomly selected.

At test, in all four repetition conditions, recall order of the target items was controlled through the presentation of the items' unique initial letters, which were presented successively and in random order, for 6 s each. Responses were given orally. In the prior restudy condition, participants were asked to study the list's nontarget items a second time, for 6 s each and in random order, before being tested on the list's target items. In the prior easy retrieval condition, nontargets were tested before target items, providing the nontargets' word stems as retrieval cues; the stems were presented successively and in random order, for 6 s each. In the prior difficult retrieval condition, nontargets were also tested before target items, but the nontargets' initial letters only were provided as retrieval cues; the initial letters were presented successively and in random order, for 6 s each. In the control condition, targets were tested immediately at the beginning of the test phase. All nontargets were repeated in two successive cycles, with two different random orders, prior to target recall (see Fig. 3).

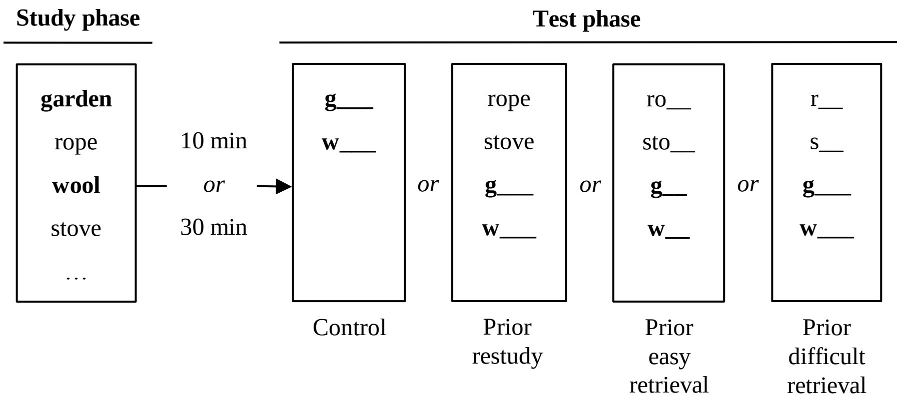

Figure 3. *Procedure and conditions employed in Experiment 2A. Participants studied a list of words and, after a 10-min or a 30-min delay, were asked to recall predefined target items from the list. The targets were tested first (control) or after prior restudy the list's remaining items (prior restudy) or after prior retrieval of the list's remaining items, with either the items' word stems serving as retrieval cues (prior easy retrieval) or their unique initial letters serving as retrieval cues (prior difficult retrieval). Predefined target items are depicted in bold letters.*

Additional baseline condition. Another 24 students ($M = 21.69$ years, range: 19-28 years, 66.7% female) took part in an additional, ninth experimental condition, in which they studied a list of items and recalled the list's target items after a 30-s distractor task, in which they counted backward from a three-digit number. Material and procedure were identical to the four repetition conditions above with the only exceptions that target items were tested only and there was a 30-s retention interval between study and test. This condition was included to serve as a baseline, in both Experiment 2A and Experiment 2B, to (i) measure amount of time-dependent forgetting after the single prolonged retention intervals, and (ii) measure the extent to which possible beneficial effects of selective item repetition eliminate time-dependent forgetting.

Results

Fig. 4 shows mean recall rates for the target items in the eight experimental conditions, and the baseline condition. A 2 × 4 analysis of variance with the between-subjects factors of REPETITION FORMAT (control, prior restudy, prior easy retrieval, prior difficult retrieval) and DELAY (10 min, 30 min) showed a main effect of DELAY, $F(1, 184) = 35.28$, $MSE = 416.67$, $p < .001$, $\eta^2 = 0.16$, and a main effect of REPETITION FORMAT, $F(3, 184) = 21.49$, $MSE = 416.67$, $p < .001$, $\eta^2 = 0.26$. There was also a significant interaction between the two factors, $F(3, 184) = 3.23$, $MSE = 416.67$, $p = .024$, $\eta^2 = 0.05$, suggesting that repetition format affected target recall in the two delay conditions differently.

Planned comparisons for the 10-min delay condition showed that all three repetition formats facilitated recall of the target items relative to the (no-repetition) control condition, all $ts(46) > 4.70$, $ps < .001$, $ds > 1.31$. The three repetition formats did not differ in recall rates, $F(2, 69) < 1$, indicating that they induced about the same beneficial effects on target recall. In the 30-min delay condition, again all three repetition formats facilitated target recall relative to the control condition, all $ts(46) > 2.35$, $ps < .023$, $ds > 0.67$. In contrast to the 10-min retention interval condition, however, the three repetition formats differed in recall rates in this condition, $F(2, 69) = 5.25$, $MSE = 410.87$, $p = .008$, $\eta^2 = 0.13$. In fact, both easy retrieval and difficult retrieval induced larger beneficial effects than restudy, $t(46) = 2.31$, $p = .025$, $d = 0.67$, and $t(46) = 3.11$, $p = .003$, $d = 0.90$, whereas the two retrieval formats did not create different recall rates, $t(46) < 1$.

Additional analyses showed that target recall in the 30-sec baseline condition was higher than in the 10-min control condition (70.83% vs. 40.83%), $t(46) = 5.38$, $p < .001$, $d = 1.55$, and was higher in the 10-min control condition than in the 30-min control condition (40.83% vs. 27.50%), $t(46) = 2.25$, $p = .029$, $d = 0.65$, thus showing the typical pattern of time-dependent forgetting. Interestingly, in the 10-min delay condition, recall rates in the three repetition conditions were statistically indistinguishable from recall in the (30-sec) baseline condition, all $ts(46) < 1.06$, $ps > .296$, $ds < 0.31$, indicating that selective item repetition did not only induce beneficial effects but eliminated all time-dependent forgetting. In contrast, in the 30-min delay condition, recall rates in all three repetition conditions were below baseline, all $ts(46) > 2.14$, $ps < .037$, $ds > 0.62$, indicating that selective item repetition induced only partial elimination of the time-dependent forgetting. Nontarget recall rates in the easy and difficult retrieval conditions differed in the 10-min delay condition (94.17% vs. 55.00%), $t(46) = 11.17$, $p < .001$, $d = 3.23$, as well as in the 30-min delay condition (87.50% vs. 55.00%), $t(46) = 7.20$, $p < .001$, $d = 2.08$. They did not depend on delay (74.58% vs. 71.25%), $F(1, 92) = 1.36$, $MSE = 196.01$, $p = .246$, $\eta^2 = 0.02$.

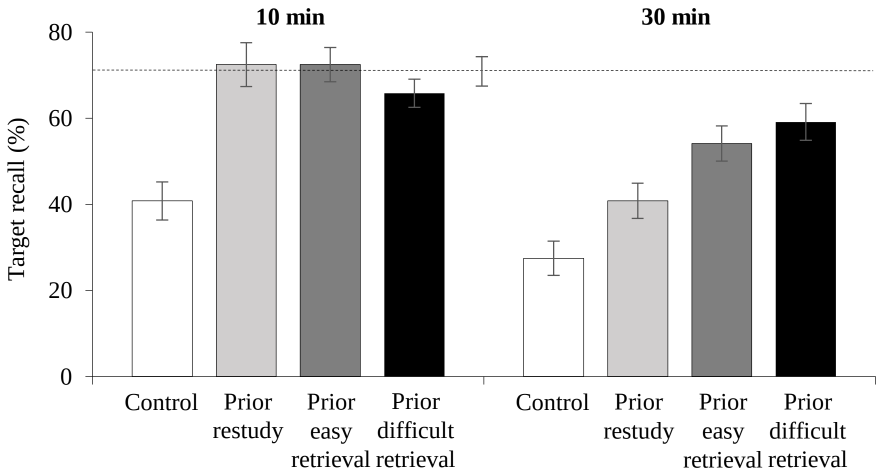

Figure 4. *Results of Experiment 2A. Percentage of recalled target items is shown as a function of delay (10 min, 30 min) and repetition format (control, prior restudy, prior easy retrieval, prior difficult retrieval). Error bars represent standard errors. The dashed line represents a 30-sec delay baseline condition, in which target items were recalled first.*

Discussion

For both delay conditions and all three repetition formats, the results show beneficial effects of selective item repetition on recall of the nonrepeated items. However, whereas after the shorter 10-min retention interval, the three repetition formats induced about the same beneficial effects and completely eliminated time-dependent forgetting, after the longer 30-min retention interval, there was only partial elimination of time-dependent forgetting and selective retrieval improved recall of the other items more than selective restudy did. The findings for selective retrieval and selective restudy thus simulate the results reported in Bäuml and Dobler (2015), which found selective retrieval and selective restudy to induce largely equivalent beneficial effects after presentation of a forget cue, i. e., after moderate episodic forgetting, but larger beneficial effects of selective retrieval than selective restudy after a retention interval of 48 hrs, i. e., after stronger episodic forgetting. Together, the results thus indicate that the degree to which access to study context at test is impaired can influence whether selective retrieval induces stronger beneficial effects than restudy (see also 1.8 Conclusions).

The finding that, after a longer retention interval, selective retrieval can improve recall of other items more than selective restudy is consistent with the two-factor account of selective retrieval and Karpicke et al.'s (2014) variant of context retrieval theory, according to which retrieval may trigger more context reactivation than restudy. However, the finding that, after the same retention interval, difficult retrieval did not improve target recall more than easy retrieval does not agree with the theory, indicating that context reactivation processes may not vary with retrieval difficulty and the beneficial effect of selective retrieval may not depend on how demanding retrieval is.

However, the procedure employed in Experiment 2A may have underestimated the beneficial effect of difficult retrieval. In fact, while in the selective restudy condition of the experiment participants should have repeated more or less all of the nontarget items and in the selective easy retrieval condition nearly all of the items (91% success rate for nontarget recall), in the difficult retrieval condition only about half of the nontarget items were repeated (55% success rate for nontarget recall). Because the beneficial effect of selective retrieval has been shown to increase with number of successfully retrieved nontarget items (Bäuml & Samenieh, 2010), this finding indicates that the beneficial effect of difficult selective retrieval may have been underestimated in this experiment. If so, an effect of retrieval difficulty on the beneficial effect of selective retrieval may emerge if number of successfully retrieved nontargets was controlled across retrieval conditions. Experiment 2B addresses the issue.

1.6 Experiment 2B: The Role of Repetition Format and Retrieval Success

Experiment 2B repeated Experiment 2A with two changes. The first change was that Experiment 2B was aimed at roughly equating number of successfully repeated nontarget items across the three repetition conditions to no longer underestimate the beneficial effect of difficult retrieval. This was achieved by reducing the number of to-be-repeated nontarget items in the restudy and easy retrieval conditions. In fact, while in the difficult retrieval condition, subjects should repeat all 10 nontarget items, only 6 of the 10 nontargets should be repeated in the restudy and easy retrieval conditions. On the basis of the results of Experiment 2A, we expected that this adjustment created similar numbers of successful repetitions in the three repetition conditions. As the second change, we employed a single retention interval condition only but increased the retention interval between study and test to 24 hrs. If differences in repetition format are more easily detected after longer than after shorter retention interval (see Experiment 2A), then the increase in retention interval to 24 hrs may enhance chances to find pairwise differences between the three repetition formats. On the basis of the results of the 30-min condition of Experiment 2A, we again expected all three repetition formats to show beneficial effects of selective item repetition and to find selective retrieval to create larger beneficial effects than selective restudy. By holding number of successfully repeated items roughly constant across repetition conditions, we additionally expected that difficult selective retrieval created a larger beneficial effect for the nonretrieved items than easy selective retrieval. Such pattern of results would indicate that the beneficial effect of selective item repetition indeed depends on difficulty of selective item repetition (see also Tab. 1).

Method

Participants. Another 128 students of Regensburg University took part in the experiment (M=22.89 years, range=19-35 years, 74.2% female). They were equally distributed across the four between-subjects conditions, resulting in n=32 participants in each single condition. Bäuml and Dobler (2015, Exp. 2) reported an effect size of d=.62 for the difference in recall levels between easy retrieval and restudy for a prolonged retention interval of several hours. To ensure that such difference would be detected in the present experiment, in which we also employed a prolonged retention interval, an analysis of test power with the G*Power program (version 3, Faul, Erdfelder, Lang, & Buchner, 2007) was conducted. Setting alpha=.05 and beta=.20, this analysis suggested a sample size of n=33 subjects per condition, which is close to the sample size of n=32 employed in Bäuml and Dobler (2015). We followed the prior work. All subjects spoke German as native language and received monetary reward or course credit for their participation.

Materials. Materials were identical to Experiments 1 and 2A.

Design and Procedure. The experiment had a unifactorial design with the between-subjects factor of REPETITION FORMAT (control, prior restudy, prior easy retrieval, prior difficult retrieval). The procedure was largely identical to Experiment 2A, with two exceptions only: First, we increased the retention interval between study and test to 24 hrs. Included in this interval was a 10-min distractor block conducted immediately after study and another 10-min distractor block conducted immediately before test. The two distractor blocks were identical in design to the distractor blocks employed in Experiment 2A. Second, whereas at test, participants were asked to recall all 10 nontargets in the difficult retrieval condition, in the easy retrieval and restudy conditions, they recalled or restudied only 6 of the 10 nontarget items; for each subject, the 6 items were randomly selected from the set of 10 nontarget items. These items remained constant across the two repetition cycles, although retrieval or presentation order were randomized within each single repetition cycle. In all other aspects Experiment 2B was identical to Experiment 2A (see Fig. 5).

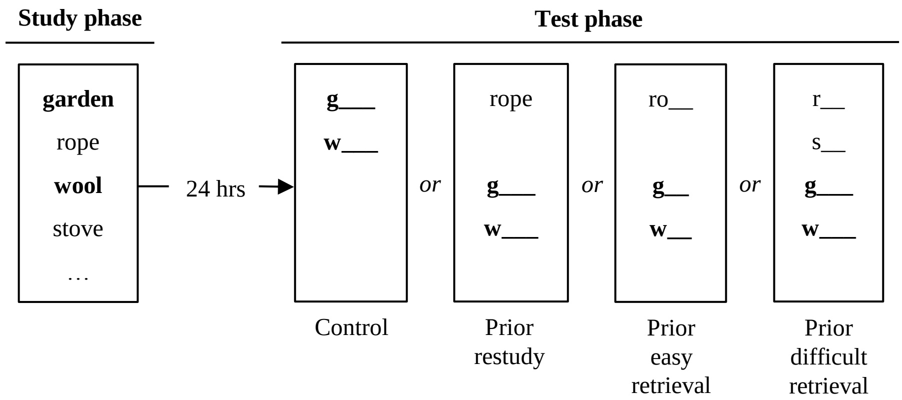

Figure 5. *Procedure and conditions employed in Experiment 2B. Participants studied a list of words and, after a 24-hrs delay, were asked to recall predefined target items from the list. The targets were tested first (control) or after prior restudy of 6 randomly selected items of the list's remaining items (prior restudy), after prior retrieval of 6 of the remaining items with the items' word stems serving as retrieval cues (prior easy retrieval), or after prior retrieval of all of the remaining items with their unique initial letters serving as retrieval cues (prior difficult retrieval). Predefined target items are depicted in bold letters.*

Results

In the easy retrieval condition, 5.41 of the (6) nontarget items were successfully recalled, and in the difficult retrieval condition 5.66 of the (10) nontarget items were successfully recalled. The difference between the two conditions was not significant, $t(62) < 1$, and the number of successfully retrieved items was also close to the number of restudied items (6) in the selective restudy condition. All this indicates that control of number of successfully repeated items was quite effective in this experiment.

Fig. 6 shows mean recall rates for the target items. A unifactorial analysis of variance with the between-subjects factor of REPETITION FORMAT (control, prior restudy, prior easy retrieval, prior difficult retrieval) showed a main effect of repetition format, $F(3, 124) = 23.95$, $MSE = 292.04$, $p < .001$, $\eta^2 = 0.37$. Planned comparisons revealed beneficial effects for all three repetition formats relative to the (no-repetition) control condition, all $ts(62) > 2.75$, $ps < .008$, $ds > 0.69$. In particular, repetition formats differed pairwise in the size of the beneficial effect, with difficult retrieval inducing higher target recall than easy retrieval, $t(62) = 2.15$, $p = .035$, $d = 0.54$, and easy retrieval inducing higher target recall than restudy, $t(62) = 2.98$, $p = .004$, $d = 0.75$.

Including the 30-sec baseline condition of Experiment 2A into the analysis also showed that, like in Experiment 2A, time-dependent forgetting was present, $t(54) = 11.44$, $p < .001$, $d = 3.06$. In particular, like in the 30-min retention interval condition of Experiment 2A, recall rates in all three repetition conditions were below the 30-sec baseline condition, indicating that in all three repetition conditions only partial elimination of the time-dependent forgetting was present, all $ts(54) > 2.92$, $ps < .005$, $ds > 0.80$.

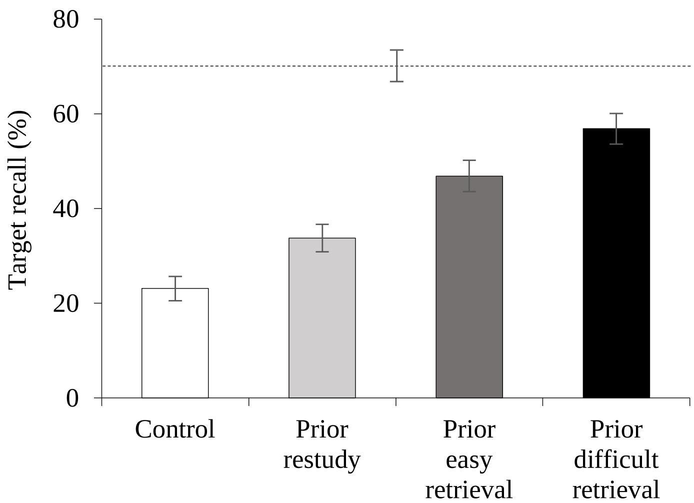

Figure 6. *Results of Experiment 2B. Percentage of recalled target items is shown as a function of repetition format (control, prior restudy, prior easy retrieval, prior difficult retrieval). Error bars represent standard errors. The dashed line represents the same baseline condition as shown in Fig. 4.*

Discussion

Employing a longer retention interval than in Experiment 2A and controlling the number of successfully repeated items across repetition conditions, we replicated the results of the 30-min condition of Experiment 2A by showing that (i) all three repetition formats can improve recall of target information, (ii) none of the three repetition formats eliminates time-dependent forgetting completely, and (iii) both forms of selective retrieval induce stronger beneficial effects on target recall than selective restudy. Unlike in Experiment 2A, however, the two retrieval conditions differed in amount of the beneficial effect, with difficult retrieval improving target recall more than easy retrieval. These findings are consistent with the view that context reactivation processes mediate the beneficial effects and Karpicke et al.'s (2014) variant of context retrieval theory, according to which difficulty of item repetition should influence the amount of context reactivation and thus should influence the size of the beneficial effect that arises for nonrepeated items in response to selective item repetition.

1.7 Experiment 3: The Role of Retrieval Difficulty with Integrated Prose Material

By showing that selective retrieval can induce larger beneficial effects for recall of other items than selective restudy, the results of Experiment 2B confirmed the results of Experiment 2A as well as the results from prior work on selective item repetition (Bäuml & Dobler, 2015). Besides, Experiment 2B provided the first indication that difficult selective retrieval may improve recall of nonrepeated items more than easy selective retrieval. It was therefore the goal of Experiment 3 to replicate this latter finding. In contrast to Experiment 2B, more coherent prose material was used as study material in this experiment. While prior work already demonstrated that, after prolonged retention interval, selective retrieval can induce beneficial effects on recall of other information also with coherent prose material (Bäuml & Schlichting, 2014), this prior work used difficult retrieval for selective item repetition only and did not address the role of repetition format in selective memory retrieval.

Experiment 3 compared the beneficial effects of easy and difficult selective retrieval employing the text passage *The Big Bang*, which was already used in previous studies (Bäuml & Schlichting, 2014; Chan, McDermott, & Roediger, 2006). Subjects studied the text passage and after a retention interval of 30 min were tested on some target questions, either first or after prior selective answering of some nontarget questions. Analogous to Experiments 2A and 2B, we employed two selective retrieval conditions: in the easy retrieval condition, subjects answered the nontarget questions with the word stems of the missing items serving as retrieval cues; in the difficult retrieval condition, nontarget questions were answered in the absence of any item-specific cues. Like Experiment 2B, Experiment 3 aimed at equating number of successfully retrieved items in the two retrieval conditions, which was achieved by reducing the number of presented nontarget questions in the easy retrieval condition. On the basis of the results of Bäuml and Schlichting (2014) on the effects of selective retrieval with coherent prose material, we expected that selective retrieval can induce beneficial effects on recall of the target information. Following the results of Experiment 2B of this study, we expected that this beneficial effect may be larger after difficult than after easy selective retrieval (see also Tab. 1).

Method

Participants. Another 138 students of Regensburg University participated in the experiment ($M=$ 22.09 years, range=17-35 years, 88.4% female). They were equally distributed across the three between-subjects conditions, resulting in n=46 participants in each condition. Number of subjects per condition was increased relative to Experiment 2B because, on the basis of the results of Bäuml and Schlichting (2014), we expected effect sizes to be reduced by the order of d=.10 with prose material relative to lists of unrelated items. With alpha=.05 and beta=.20, analysis of test power suggested a sample size of n=47 subjects per condition to detect an effect of d=.52 (relative to .62 in Experiment 2B). We followed this suggestion closely. All subjects spoke German as native language. In exchange for participation, course credit or monetary reward was provided.

Materials. The text passage *The Big Bang*, which was already used in prior work (Bäuml & Schlichting, 2014; Chan et al., 2006), served as study material. We used the German translation of the text used in Bäuml and Schlichting. The text was approximately 1800 words long. We selected the same 6 target and the same 12 nontarget questions as were employed in the prior work (gapped sentences like "The Hubble telescope found the heavy element ____ in extremely ancient stars." [Answer: boron] or "Arthur Eddington said: 'We must allow ____ an infinite amount of time to get started.' " [Answer: evolution], see Appendix B).

Design. The experiment had a unifactorial design with the between-subjects factor of REPETITION FORMAT (control, prior easy retrieval, prior difficult retrieval). Participants answered the target questions first (control) or after prior selective answering of the nontarget questions. In the two retrieval conditions, subjects answered the nontarget questions first with the missing information's word stem provided as a retrieval cue (prior easy retrieval), or in the absence of any item-specific cues (prior difficult retrieval). Assignment of conditions was counterbalanced.

Procedure. The procedure largely followed the one used in Bäuml and Schlichting (2014; Experiment 2), differing only in the length of the retention interval and format of selective retrieval. Participants had 16 min to read the study text with the instruction that all information can be relevant for the later test and without knowing what type of test would be conducted. After study, a 30-min retention interval followed that included three blocks of 10 min, with each block consisting of several distractor tasks including one imagination task (see Method of Experiment 2A above). At test, subjects had 25 s to answer a single target or nontarget question. Responses were given orally. Target questions were either tested first or after selective answering of the nontarget questions. In the prior easy retrieval condition, nontarget questions were answered providing the word stems of the missing items as retrieval cues; in the prior difficult retrieval condition, no such retrieval cues were provided. No item-specific cues were provided for the answers of the target questions. Both target and nontarget questions were presented successively and in random order. Like Experiment 2B above, the experiment was aimed at roughly equating number of successfully retrieved nontarget items between retrieval conditions, which was achieved by reducing the number of to-be-repeated nontarget questions in the easy retrieval condition. Accordingly, participants were asked to answer all 12 nontarget questions in the difficult retrieval condition, but were asked to answer only 8 of the 12 questions in the easy retrieval condition. For each subject, the 8 questions were randomly selected from the set of 12 nontarget questions. The selection of the nontarget questions remained constant across two repetition cycles, but order of questions was randomized within each of the two cycles (see Fig. 7).

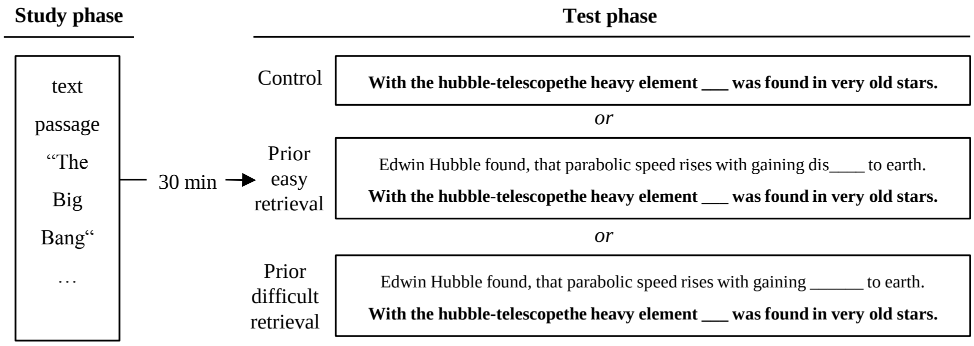

Figure 7. *Procedure and conditions employed in Experiment 3. Participants read a study text for 16 min. After a 30-min delay, they were asked to recall predefined target questions. The target questions were tested first (control) or after prior answering 8 randomly selected questions with the missing words' stems serving as retrieval cues (prior easy retrieval) or after prior answering 12 questions with no retrieval cues for the missing words (prior difficult retrieval). Predefined target questions are depicted in bold letters.*

Results

In the easy retrieval condition 5.79 of the (8) nontarget questions and in the difficult retrieval condition 5.65 of the (12) nontarget questions were correctly answered. Nontarget recall thus did not differ between retrieval conditions, $t(90) = 1.24$, $p = .217$, $d = 0.26$, indicating that control of number of successfully retrieved nontargets was effective.

Fig. 8 shows mean recall rates for the target questions. A unifactorial analysis of variance with the between-subjects factor of REPETITION FORMAT (control, prior easy retrieval, prior difficult retrieval) showed a main effect of REPETITION FORMAT, $F(2, 135) = 11.08$, $MSE = 383.89$, $p < .001$, $\eta^2 = 0.14$. Planned comparisons revealed beneficial effects in both retrieval formats relative to the (no-repetition) control condition, both $ts(90) > 2.30$, $ps < .024$, $ds > 0.48$. In particular, the two retrieval formats affected target recall differently, with higher recall in the difficult than the easy retrieval condition, $t(90) = 2.21$, $p = .030$, $d = 0.46$.

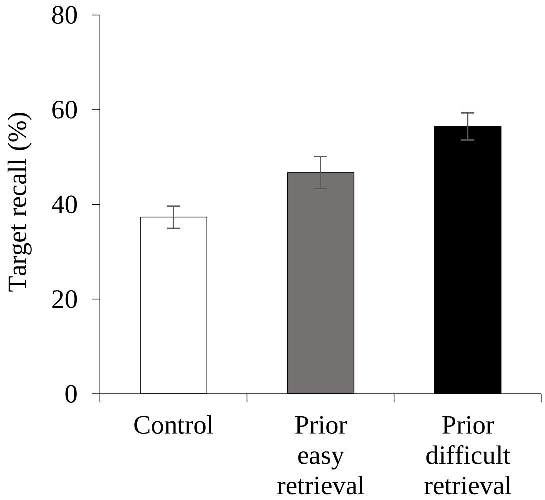

Figure 8. *Results of Experiment 3. Percentage of recalled target items is shown as a function of repetition format (control, prior easy retrieval, prior difficult retrieval). Error bars represent standard errors.*

Discussion

The results of Experiment 3 mimic recent findings of Bäuml and Schlichting (2014) by showing that, after a prolonged retention interval, selective retrieval can induce beneficial effects on other items also with coherent prose material. They go beyond the prior work by demonstrating that the beneficial effect can be modulated by retrieval difficulty, with difficult selective retrieval inducing a stronger beneficial effect than easy selective retrieval. Doing so, the results replicate the findings of Experiment 2B that employed unrelated word lists as study material with coherent prose material. Together with Experiment 2B, the results thus indicate that retrieval format can influence the size of the beneficial effect of selective retrieval and that the effect increases with difficulty of retrieval task.[4]

1.8 Discussion of Experiments 1-3

The first goal of this chapter was to examine the proposal included in Bäuml and Samenieh's (2010) two-factor account of selective retrieval that the beneficial effect of selective retrieval is mediated by context reactivation processes more directly. Impairing study context access through a prolonged retention interval, the results of Experiment 1 firstly showed that, in the absence of preceding mental reinstatement of the study context, selective retrieval of some studied information can in fact improve recall of the other information, thus replicating results from previous studies (e. g., Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014). However, when immediately before selective retrieval started, study context was mentally reinstated, no such beneficial effect arose and selective retrieval rather impaired recall of the other items, which mimics the typical detrimental effect of selective retrieval in the absence of any impairment in study context access (see Bäuml et al., 2017). These findings provide direct evidence of the critical role of impaired study context access for the beneficial effect of selective retrieval and thus support the proposal that context reactivation processes mediate the beneficial effect.

[4]Experiment 3 had a focus on the role of retrieval format and did not include a restudy condition. If a restudy condition had been included, the beneficial effect would have been expected to be larger after easy retrieval than restudy (see Experiments 2A and 2B). Inspection of Fig. 8 suggests that, in such case, recall in the restudy condition would not have been much different from recall in the (no-repetition) control condition, indicating that, with prose material, selective restudy may not induce any beneficial effects. However, whether a repetition format creates beneficial effects for some study material may depend on number of repeated nontarget items (see Bäuml & Samenieh, 2010). The results of Experiment 3 may thus suggest that, with prose material, a larger number of nontarget items must be repeated than with lists of unrelated items to improve recall of other information. Future work may address the issue directly.

The second goal of this chapter was to examine whether the size of the beneficial effect of selective retrieval on recall of the nonrepeated items is the same for easy and difficult selective retrieval, and can even generalize to selective restudy trials. On the basis of the view that context reactivation processes mediate the beneficial effect and Karpicke et al.'s (2014) variant of context retrieval theory, repetition format may influence the effect, with more difficult selective item repetition inducing a stronger beneficial effect on nonrepeated items than more easy item repetition does. The results of Experiments 2-3 show such pattern. Employing retention intervals of 30 min and 24 hrs, between study and selective item repetition, Experiments 2A and 2B showed that selective retrieval improves recall of other items more than selective restudy does. In addition, for the same retention intervals, Experiments 2B and 3 showed that, when number of successfully selectively retrieved items is controlled, difficult selective retrieval improves recall of other items more than easy selective retrieval does. These results converge on the view that difficulty of selective item repetition can influence the beneficial effect on nonrepeated items, which, together with the results of Experiment 1, support the proposal that context reactivation processes mediate the effect and are modulated by repetition format. For the shorter retention interval of 10 min, such effect of repetition format was absent and the single repetition formats created equivalent beneficial effects.

Implications for the Two-Factor Account of Selective Retrieval

The present results of the first chapter strengthen and extend the two-factor account of selective memory retrieval. This account claims that, in general, selective retrieval triggers inhibition and blocking as well as context reactivation processes. Critically, the contribution of the two types of processes is assumed to depend on access to study context at test, with a larger relative contribution of inhibition and blocking when study context access is (largely) maintained, and a larger relative contribution of context reactivation processes when access to study context at test is impaired. As a result, detrimental effects of selective retrieval on other items may be observed when access to study context is maintained at test, but beneficial effects may result when access to study context is impaired.

This account is supported by the present demonstration that, after a prolonged retention interval, mental reinstatement of the study context immediately before recall starts can eliminate and even reverse the beneficial effect. Indeed, in the absence of a mental context reinstatement, the delay-induced impairment in study context access should trigger retrieval-induced context reactivation and thus induce beneficial effects on the recall of related items. In contrast, in the presence of such reinstatement, the impairment in context access should be reduced, reducing the need for (further) retrieval-induced reactivation and thus attenuating possible beneficial effects of selective retrieval. Rather, the induced context reinstatement may revive item interference, leading to inhibition and blocking and detrimental effects of selective retrieval. By showing the two faces of selective retrieval in the presence versus absence of mental context reinstatement, the present findings are consistent with this proposal.

The present results also extend the two-factor account. Empirically, they extend the account by showing that beneficial effects on nonrepeated items do not only arise in response to (easy) selective retrieval trials, as has been shown in the prior work (see Bäuml et al., 2017), but do also arise in response to difficult selective retrieval and selective restudy trials. This holds while repetition format can modulate the size of the beneficial effect, with more difficult repetition formats creating larger beneficial effects than more easy repetition formats. Theoretically, the results extend the account by imposing a restriction on the proposed context reactivation processes, suggesting that amount of context reactivation varies with repetition format. This view on the underlying context reactivation processes fits with Karpicke et al.'s (2014) variant of context retrieval theory. In this variant, it was argued that the degree of context reactivation may be higher after retrieval than restudy, because context retrieval may not be obligatory during restudy cycles, and because, with retrieval, people deliberately search memory information about the prior occurrence of studied information. Similarly, difficult and easy retrieval conditions may also induce a difference in context reactivation, because mostly difficult retrieval and less easy retrieval requires reactivation of the study context.

Including this variant of context retrieval theory into the two-factor account leads to a more general two-factor account, which is able to explain the beneficial effects of selective item repetition on nonrepeated items, as they are reported in this study, but is also able to explain the detrimental effects of selective item repetition, as they were reported in prior work on retrieval-induced forgetting (e. g., Bäuml & Kliegl, 2017; Storm & Levy, 2012). These results showed mostly retrieval-specific detrimental effects of selective item repetition, a pattern well explained by inhibition and blocking processes (see above). Whether the size of the detrimental effect of selective retrieval also varies with difficulty of selective retrieval has not been examined yet. Because more difficult selective retrieval (e. g., providing weak item-specific cues) may create more interference from other items than easy selective retrieval (e. g., providing strong item-specific cues), on the basis of the inhibition view there would be reason to expect larger detrimental effects after difficult than easy retrieval.[5] Future work may address the issue and fill this empirical gap.

In a previous study, Bäuml and Dobler (2015) compared the effects of selective retrieval with the effects of selective restudy, both when a forget cue was provided after study and when a prolonged 48-hrs retention interval occurred between study and selective item repetition. Equivalent effects of selective retrieval and selective restudy were found after the forget cue (which created moderate episodic forgetting), whereas stronger beneficial effects of retrieval than restudy were found after the prolonged retention interval (which

[5] If difficult selective retrieval strengthened practiced items more than easy selective retrieval (e. g., Bjork & Kroll, 2015), then the same prediction would arise on the basis of the blocking account, arguing that stronger practiced items may block recall of unpracticed items more than weaker practiced items.

created strong episodic forgetting). Moreover, while both forms of selective item repetition eliminated the episodic forgetting induced by the forget cue completely, selective item repetition eliminated only about half of the episodic forgetting that was induced by the prolonged retention interval. Interestingly, the results of the present Experiments 2A and B mimic these findings using shorter (10 min) and longer (30 min, 24 hrs) retention intervals. They show equivalent effects of selective retrieval and selective restudy after the shorter (10-min) delay, with a complete elimination of the time-dependent forgetting. In contrast, they show larger beneficial effects of selective retrieval than selective restudy after the longer (30-min and 24-hrs) delays, with elimination of only about half of the time-dependent forgetting.

These parallels indicate that the effects of selective item repetition in listwise directed forgetting can be simulated using time-dependent forgetting with a moderate length of retention interval, thus supporting the view that, in both forms of forgetting, inaccess to study context plays a critical role for the induced forgetting (e. g., Estes, 1955; Geiselman, Bjork & Fishman, 1983; Sahakayn & Kelley, 2002). Moreover, on the basis of the parallels, the prediction arises that, if a forget cue created stronger episodic forgetting than the forget cue is doing in the standard listwise directed forgetting task (for an example, see Bäuml & Kliegl, 2013, Experiments 1A und 1B), then also in listwise directed forgetting selective retrieval may create larger beneficial effects than selective restudy. Why selective retrieval and selective restudy induced equivalent beneficial effects in Bäuml and Dobler's (2015) directed forgetting experiment and the moderate retention interval condition of the present Experiment 2A, but different beneficial effects in the other conditions of the present chapter is less clear. A possible reason, however, may be the presence of a ceiling effect. Indeed, both in Bäuml and Dobler's directed forgetting experiment and the moderate retention interval condition of the present Experiment 2A, selective restudy already eliminated all episodic forgetting, so that no further room may have been left for an enhanced beneficial effect of selective retrieval. If so, the present results would indicate that selective retrieval improves recall of other items more than selective restudy whenever room is left for retrieval to create more context reactivation than restudy.

Relation to Prior Work on Effects of Selective Retrieval in Social Recall

Abel and Bäuml (2015) recently suggested that the finding of beneficial effects of selective retrieval in individuals may generalize to social recall. Using the speaker-listener task - a task, in which two individuals study a list of items and one of the two persons (the "speaker") selectively retrieves a subset of the information before the other person (the "listener") recalls the remaining information (Cuc, Koppel, & Hirst, 2007) - these researchers found that, when access to study context at test was impaired for the two persons - by providing a forget cue after study, being engaged in an imagination task, or a prolonged retention interval between study and selective retrieval - the selective retrieval by the speaker improved the subsequent recall of the listener, similar to how selective retrieval had been shown to improve the recall of other items in individuals. In contrast, Hirst and colleagues reported that the selective retrieval of a speaker reduced the recall of a listener when recalling autobiographical or flashbulb memories, that is, memories that were encoded a long time before selective retrieval started and whose encoding context may therefore not have been easy to access (e. g., Coman, Manier, & Hirst, 2009; Stone, Barnier, Sutton, & Hirst, 2013).

The results of the present Experiment 1, which showed that mental reinstatement of the study context before test can eliminate and even reverse the beneficial effect, may bridge the gap between the two lines of studies. Indeed, while in the Abel and Bäuml (2015) study, nothing was done to reinstate participants' study context after access to study context was experimentally impaired, in the studies by Hirst and colleagues target memories were reactivated before selective retrieval started. In Coman et al. (2009) a questionnaire probed participants' flashbulb memories of the

September 11 attack before subjects engaged in selective retrieval; in Stone et al. (2013) participants first underwent an elicitation phase and generated the autobiographical memories, before, a day later, they studied each generated memory again right before selective retrieval began. Thus, in both studies, access to the encoding context may no longer have been impaired when selective retrieval started, providing little or no need for further context reactivation and any beneficial effects of selective retrieval. The findings by Hirst and colleagues thus are not in direct conflict with Abel and Bäuml's proposal that retrieval dynamics in individuals may generalize to social groups.

Relation to Prior Work on the Testing Effect

The present results of chapter 1 together with the two-factor account of selective retrieval suggest that repetition formats can differ in the degree to which they cause context reactivation and thus differ in the degree to which they induce beneficial effects on the recall of other items. This proposal parallels Karpicke et al.'s (2014) view on the testing effect. These authors argued that repetition formats can differ in the degree to which they cause context reactivation and thus differ in the degree to which they cause recall improvements for the repeated information itself. Indeed, the difference in context reactivation may induce a difference in the creation of unique context cues for the repeated items, and the larger number of unique context cues after retrieval may enhance retention of retrieved items more than of restudied items. Results from numerous studies in fact showed that (nonselective) retrieval practice can induce better recall of practiced items than restudy, and diffcult (nonselective) retrieval can induce better recall of practiced items than easy retrieval (e. g., Carpenter & DeLosh, 2006; Halamish & Bjork, 2011; Karpicke & Roediger, 2008; Roediger & Karpicke, 2006a).

While the results from the testing effect studies and the results from the present selective item repetition study thus suggest a similar role of repetition format for context reactivation processes in the two lines of studies, there is also an important difference in circumstances that surround the effects of repetition format in the two types of situations. Indeed, while the present results on beneficial effects of selective item repetition for nonrepeated items are tied to an impairment in study context access, with a reversal of effects when the impairment is absent (see present Experiment 1), the testing effect does not show such restriction. In fact, testing effects can easily be observed without inducing any major change in context between study and practice (see Roediger & Butler, 2011).

The apparent inconsistency between the two lines of findings is not in conflict with the view that item repetition triggers context reactivation processes already in the absence of major context change. Indeed, according to the two-factor account, possible beneficial effects of context reactivation in the absence of major context change can be masked by the simultaneous action of inhibition and blocking processes, thus inducing detrimental, rather than beneficial effects on nonrepeated items (e. g., Bäuml & Samenieh, 2012). The view that the role of context reactivation processes is increased in the presence of context change, as it is included in both the two-factor account and the episodic-context account, is supported by the finding of two faces of selective retrieval as well as results from testing effect studies. Indeed, at least the results of two such studies suggest that larger testing effects can arise when context between study and retrieval practice is changed than when it is left largely unaffected (Pyc,Balota, McDermott, Tully, & Roediger, 2014; Smith & Handy, 2014). Future work may examine the issue in more depth and investigate the effects of (selective) item repetition simultaneously on repeated and nonrepeated items. Such work should provide more detailed insights into the role of repetition-induced context reactivation.

Conclusions

So far, Experiment 1 in the first chapter served as the first indication that context reactivation processes mediate the beneficial effect of selective item repetition on recall of other items, thus empirically substantiating the two-factor account of selective retrieval and providing a possible explanation for supposed inconsistencies in studies on selective retrieval in social groups. The present Experiments 2A, 2B, and 3 furthermore showed that repetition format can influence such context reactivation processes, with the beneficial effect of item repetition being larger after selective retrieval than selective restudy, and being larger when selective retrieval is demanding. These findings extend the two-factor account of selective retrieval and fit with Karpicke et al.'s (2014) variant of context retrieval theory.

However, while chapter 1 focused on effects of selective retrieval of some items of a list on recall of the list's other items, chapter 2 aimed at investigating beneficial effects of recall across successive recall attempts, employing the hypermnesia paradigm.

Chapter 2

Beneficial Effects of Memory Retrieval in Hypermnesia

2.1 Effects of Retrieval on Subsequent Recall Tests

Hypermnesia

Assuredly, everybody knows the situation, when we are unable to remember a presumably forgotten fact that we encoded earlier, but that it comes to mind some time or repeated recall attempts later. For example, we cannot remember the name of the one English teacher in school, but her name comes to mind, when we repeatedly list all of the former teachers we can remember instantly. This phenomenon has also been examined experimentally for over 100 years (for a more detailed historical review on reminiscence and hypermnesia, see Payne, 1987). It was first studied experimentally by Ballard (1913), who termed the phenomenon *reminescence* and defined it as ”remembering again of the forgotten without re-learning”. Under a variety of study and testing conditions, he showed enhanced recall with repeated recall attempts. Because of the innovation at that time and the intuitive contrast to Ebbinhaus' (1964) finding of time-dependent forgetting, the reminescence finding attracted a great deal of attention and was soon replicated by lots of studies (e. g., Ammons & Irion, 1954; Bunch, 1938; McGeoch, 1935). Between the 1950s and 1970s, there was practically no research on reminescence because of methodological discrepancies, and consequently questionable reliability in the previous studies reviewed by Buxton (1943). However, motivated by the work of Erdelyi and colleagues (e. g., Erdelyi & Becker, 1974; Erdelyi, Buschke & Finkelstein, 1977; Erdelyi & Kleinbard, 1978; Shapiro & Erdelyi, 1974), there was a reformation of the research on improved memory performance across repeated tests. While there was a lack of clear determinations in the prior concept of reminescence, Erdelyi and colleagues used the more generic term *hypermnesia* to describe the phenomenon and underlined the improvement in *net* recall with repeated tests but without any relearning.

In a typical experiment on hypermnesia, participants study a set of items, like words or pictures, and are then presented with a series of successive recall tests, in each of which they are asked to recall the previously studied items. Across tests, some items are recalled on later tests that were not recalled in prior tests (item gains), whereas other items recalled on prior tests are not recalled on a later test (item losses). Hypermnesia arises if item gains exceed item losses, and consequently, a net increase in the number of items recalled across tests results. In contrast, net forgetting is generated if item losses exceed item gains (for reviews, see Erdelyi, 1996; Payne, 1987).

Particularly, since the studies of Erdelyi and colleagues hypermnesia is considered as a robust effect that was demonstrated in quite different experimental settings. It was shown in a variety of list-learning experiments, for instance, employing unrelated words, associated word pairs, pictures, foreign language vocabulary, or nonsense syllables (e. g., Belmore, 1981; Kelley & Nairne, 2003; Mulligan, 2001; Roediger & Payne, 1982). It arose with prose passages (Otani & Griffith, 1998; Wheeler & Roediger, 1992) and films (Montangero, Ivanyi, & de Saint-Hilaire, 2003), and was demonstrated in studies on eyewitness memory (Dunning & Stern, 1992) and autobiographical memory (Bluck, Levine, & Laulhere, 1999). However, despite the large number of studies that has been conducted on hypermnesia, to date it is still unclear exactly which mechanisms mediate the effect.

Theoretical Accounts of Hypermnesia

Over the years, a variety of different explanations for the effect have emerged. Three accounts with relevance for the present experiments are selected and subsequently, one further account is derived from the current literature on the testing effect.

One of the most prominent explanations is the *cumulative recall hypothesis* (Roediger & Challis, 1989; Roediger, Payne, Gillespie, & Lean, 1982). This hypothesis assumes that hypermnesia is a function of the cumulative level of recall of items and that study conditions producing high levels of recall are more likely to exhibit hypermnesia than study conditions producing lower levels of recall. In this approach, the end of the first recall test, which typically lasts between five and seven minutes, is considered an interruption of recall. Thus, if an experimental condition has not yet reached its asymptotic recall level at the end of this test – i. e., the level that could be produced given unlimited recall time –, then the additional retrieval time afforded by the subsequent test can produce item gains. Another account of hypermnesia is the *changes in cue set hypothesis* (Raaijmakers & Shiffrin, 1980; Roediger & Thorpe, 1978). This hypothesis suggests that the cue set that people use to sample and recover memories over longer test intervals can change depending on the items "sampled" as retrieval cues. Because new cue sets arise with newly recalled information, on a later test, alternative retrieval routes may be used, which may lead to retrieval of previously unrecalled information and thus improve recall performance. Yet another account of hypermnesia is the *retrieval strategy hypothesis*, which explains hypermnesia by improved retrieval strategies and enhanced organization arising from retrieval practice in repeated testing (Erdelyi & Becker, 1974; Mulligan, 2001). According to this view, accessibility of information on a later test may be greater than that on an earlier test, because the earlier test permits more efficient organization of recalled material, so that, on the later test, the already recalled material can be retrieved again more quickly, with time remaining for the recall of new material. More organized retrieval strategies may also limit the number of item losses between tests, thus further increasing net recall levels (McDaniel, Moore, & Whiteman, 1998).

Although each of the accounts can explain important findings in the hypermnesia literature, none of them can account for the full range of experimental results. For instance, while the cumulative recall hypothesis can explain the positive relation between variables affecting recall levels (e. g., imagery, semantic elaboration) and the magnitude of hypermnesia (Roediger

& Challis, 1989; Roediger et al., 1982), the functional equivalence between single and repeated recall tests of equal total duration, which is predicted by the hypothesis, has repeatedly been challenged (Mulligan, 2005, 2006). Similarly, while the retrieval strategy hypothesis can account for the fact that retrieval strategies become increasingly organized over multiple recall tests and appear to contribute to hypermnesia (MacDaniel et al., 1998; Mulligan, 2001), the hypothesis, for instance, has trouble explaining the picture-word difference, the very robust finding of higher hypermnesia for pictures than words (Payne, 1987). Finally, the changes in cue set hypothesis can describe several basic findings in the hypermnesia literature (Raaijmakers & Shiffrin, 1980), however, more direct tests of the hypothesis are rare. Moreover, like the cumulative recall hypothesis, this hypothesis focuses on item gains and is largely silent on item losses that may occur across subsequent recall tests

2.2 The Possible Role of Retention Interval between Study and Test

A factor that can speak to these accounts of hypermnesia is the role of delay between study and test. In a typical experiment on hypermnesia, the initial test occurs shortly after study without any major delay between study and test. Indeed, most hypermnesia studies employed a short delay between study and test of one or two minutes only, mainly to distribute the recall protocols or give detailed test instructions (e. g., Bergstein & Erdelyi, 2008; Kelley & Nairne, 2003; Mulligan, 2002; Payne & Roediger, 1987). Other studies additionally included filler tasks of two or three minutes to reduce possible recency effects (e. g., Mulligan, 2005; Otani, Widner, Whiteman, & Louis, 1999), employed a delay of five minutes with the subjects' instruction to think silently about the list items (Shapiro & Erdelyi, 1974), or employed a delay of about twelve minutes, asking subjects to participate in a distractor task and complete a questionnaire (Wheeler & Roediger, 1992). For this range of relatively short retention intervals, there is no indication yet that delay influences hypermnesia.

Expectations arising from the Single Accounts of Hypermnesia

However, longer retention intervals may well influence hypermnesia. On the basis of the cumulative recall hypothesis, for instance, one may expect that hypermnesia decreases with an increase in delay between study and test. Indeed, because longer delays generally reduce (cumulative) recall levels, and, according to the hypothesis, recall levels are positively related to the magnitude of hypermnesia, hypermnesia should be smaller after longer than shorter retention intervals and items gains should decrease with delay. In contrast, on the basis of the changes in cue set hypothesis, one may expect that hypermnesia increases with delay. Delay causes context shift (e. g., Bower, 1972; Estes, 1955) and, after context shift, retrieval of some first items can reactivate the study context and facilitate recall of the other items (e. g., Bäuml & Schlichting, 2014; Howard & Kahana, 1999; present Exp. 1). If such context reactivation was not yet complete at the end of the first recall test but extended to subsequent tests, then a longer delay between study and test may lead to more extensive changes in cue set across tests than a shorter delay, and thus enhance item gains and increase hypermnesia. Likewise, on the basis of the retrieval strategy hypothesis, one may also expect enhanced hypermnesia after longer delay. If delayed recall led to more organized retrieval strategies compared to recall after shorter delay, for instance, because recall after delay can be more challenging, then according to the hypothesis, repeated testing after longer delay may both enhance item gains and reduce item losses.

Changes in cue set and improved retrieval strategies after delay may not be the only reasons to expect increased hypermnesia after prolonged retention intervals. Differences in retrieval practice effects after short versus long delay may also influence hypermnesia, a view referred to as the *retrieval practice* *hypothesis* in the following. In fact, from the testing effect literature it is well known that (i) prior retrieval makes practiced items more accessible on subsequent tests and reduces the forgetting of the items (e. g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a), and (ii) such beneficial effects of retrieval practice are particularly strong if retrieval practice is demanding, like, for instance, in the presence of weak retrieval cues or in the presence of interference (e. g., Bäuml, Holterman, & Abel, 2014; Carpenter, 2011; Halamish & Bjork, 2011; Pyc & Rawson, 2009). Because, in general, longer delay should also make retrieval more demanding, the findings from the testing effect literature suggest that, after longer delay, retrieval on an initial test may increase hypermnesia by reducing the forgetting of the initially recalled items.

While the changes in cue set, the improved retrieval strategies, and the retrieval practice hypotheses lead to the expectation of increased hypermnesia after delay, the three hypotheses differ in their expectations on item gains and item losses. Because the changes in cue set hypothesis is primarily framed around item gains, it suggests increased item gains with delay, without making detailed suggestions regarding item losses; the retrieval strategy hypothesis leads to the expectation of both enhanced item gains and reduced item losses with delay; and the retrieval practice hypothesis suggests mainly a reduction in item losses with delay. Table 2 provides an overview of these expectations.

Table 2. *Overview of expectations from single accounts of hypermnesia regarding the effects of increased delay between study and test on net recall, item gains, and item losses.*

	Net recall	Item gains	Item losses
CR	↓	↓	–
CCS	↑	↑	–
RS	↑	↑	↓
RP	↑	0	↓

Note. CR = cumulative recall hypothesis; CCS = changes in cue set hypothesis; RS = retrieval strategy hypothesis; RP = retrieval practice hypothesis. Effects may be expected to increase (↑), decrease (↓), or remain constant (0). Alternatively, there may be no expectation (-) by the particular account.

Prior Work on the Role of Delay for Hypermnesia

To the best of my knowledge, there are only three studies in the literature yet that employed retention intervals of more than twelve minutes between study and test to examine the role of delay for hypermnesia. In one study, Dunning and Stern (1992; Experiment 2) investigated whether hypermnesia in eyewitness memory depends on delay between study and test. Subjects viewed videotapes reenacting several types of crimes and, after varying delay, were asked to provide accounts of the incident on three successive free recall tests.

The initial interview occurred immediately after watching the video tapes, after a 3-day delay, or a 1-week delay. Results revealed typical time-dependent forgetting for the number of correctly recalled facts with increasing delay. Above all, they showed hypermnesia without any influence of delay on the size of the effect. In the second study, Roediger and Payne (1982) presented subjects a list of pictures and then gave them three successive free recall tests. The first test was presented immediately after study, or was delayed by reading a prose passage by 18 min. Similar to Dunning and Stern (1992), the results showed hypermnesia, but again there was no effect of delay on the size of the effect. In the third study, Wheeler and Roediger (1992; Experiment 1) examined a number of factors of possible relevance for effects of repeated testing, but a subset of the experimental conditions is directly related to the present study. In this subset, subjects studied a list of pictures, either together with their names or embedded in a story, and, after study, received three immediate tests (the 3-3 condition) or three tests after a 1-week delay (the 0-3 condition). Results revealed typical time-dependent forgetting. More important, they showed hypermnesia after the short delay but no hypermnesia after the prolonged delay.

To date, it is not possible to draw firm conclusions on the role of delay between study and test for hypermnesia. Taken together, the hitherto existing studies on this issue produce inconsistent results. While both Dunning and Stern (1992) and Roediger and Payne (1982) reveal that repeated testing prompts increased recall after a longer delay, even if not significantly influenced by delay, Wheeler and Roediger (1992) showed no hypermnesia after a long delay of one week. Another reason for being cautious with conclusions may be that Dunning and Stern (1992) employed a very small sample of subjects, with 8-11 subjects only in each single delay condition, a sample that may have been too low in size to detect significant influences of delay on hypermnesia. Additionally, in Roediger and Payne (1982), the delay manipulation did not induce any time-dependent forgetting, which indicates that the manipulation may have been largely ineffective and thus have limited the room for influences of delay on hypermnesia. In contrast to Dunning and Stern (1992) and Roediger and Payne (1982), who used free recall at test, Wheeler and Roediger (1992) employed a forced recall format. In this format, subjects are given recall sheets with a separate line for each single to-be-recalled item at test; subjects are instructed to recall as many items as possible, but if unable to remember all studied items, to fill in the remaining spaces with their best guesses. Erdelyi and Becker (1974) established forced recall tests in hypermnesia literature in order to control for possible criterion changes over successive test trials, but Roediger and Payne (1985) showed no differences in recall levels obtained with free recall and forced recall tests, suggesting that changes in response criterion play little or no role in hypermnesia after a short delay. Remarkably, after a long delay, response criterion may change and the testing method may well influence the effect. Importantly, none of the three extant studies analyzed item gains and item losses.

2.3 Goals of Experiments 4-7

To clarify the role of delay between study and test for hypermnesia, fresh experiments are necessary that (i) include a sufficiently large sample of subjects, (ii) employ delay conditions that induce robust time-dependent forgetting, (iii) examine the possible influence of recall format (free versus forced recall) on hypermnesia after longer delay, and (iv) include not only an analysis of net recall but also of item gains and item losses.

The second chapter addresses the issue in four experiments, in each of which subjects studied a list of items and, after a varying delay, were repeatedly asked to recall the previously studied material. In Experiment 4, subjects rated a list of unrelated words to be living or nonliving, whereas in Experiments 5, 6 and 7, they studied a list of pictures. Both study conditions have repeatedly been found to induce hypermnesia (e. g., Mulligan, 2006; Roediger & Thorpe, 1978). Critically, in all experiments, delay between study and the initial test was manipulated, using a short retention interval of 3 min (Exp. 4,5) or 11.5 min (Exp. 6, 7) and a longer retention interval of 24 hrs (Exp. 4,5) or one week (Exp. 6, 7). At test, all subjects participated in a series of successive recall tests. Whereas in Experiments 4, 5 and 7 subjects participated in three free recall tests, in Experiment 6 they attended three forced recall tests. In free recall tests subjects are asked to recall as many of the previously studied items as possible, independently of what they have remembered in preceding tests without knowing how many items they studied. In contrast, in forced recall tests subjects are instructed to recall as many items as possible, but if unable to remember all studied items, to fill in the remaining spaces with their best guesses.

The four experiments were expected to replicate typical hypermnesia in the short delay condition with an increase in net recall across tests, irrespective of test format. The critical question was whether delay would influence this beneficial effect, and if so, whether it reduced hypermnesia, as may be expected on the basis of the cumulative recall hypothesis, or enhanced hypermnesia as may be expected by the changes in cue set, the retrieval strategy and the retrieval practice hypothesis. Additionally, the possible role of test format for hypermnesia after a longer delay was explored. The results of the experiments in the second chapter will fill an empirical gap in the literature on hypermnesia and may provide new information on the mechanisms contributing to the effect. Beyond that, they may expand our view on retrieval processes by providing implications on recent findings on the testing effect when evaluating the retrieval practice hypothesis and by comparing the present results on hypermnesia with the results on the beneficial effects of selective retrieval in the first chapter of the present work.

2.4 Experiment 4: Delay Effects on Hypermnesia with Words as Study Material

Experiment 4 examined the effect of delay between study and initial test on hypermnesia employing lists of unrelated words. Subjects were presented the words and, for each single word, were asked to indicate if it was living or nonliving (e. g., Belmore, 1981). After a delay of 3 min or 24 hrs, subjects participated in three successive free recall tests, in each of which they were asked to remember and write down as many of the previously rated items as possible, independent of what they had remembered in possible preceding tests. On the basis of the cumulative recall hypothesis, one may expect larger hypermnesia after the short than the long retention interval, which would be consistent with Wheeler and Roediger's (1992) finding. In contrast, on the basis of the changes in cue set, the retrieval strategy, and the retrieval practice hypotheses, one may expect larger hypermnesia after the long than the short delay. Following the changes in cue set hypothesis, such increased hypermnesia may be mediated mainly by enhanced item gains, whereas following the retrieval practice hypothesis it may be mediated mainly by reduced item losses. Following the retrieval strategy hypothesis, both effects may arise.

Method

Participants. To ensure that a possible effect of delay on hypermnesia could be detected in the present experiment, an analysis of test power was conducted with the G*Power program (version 3, Faul et al., 2007) to estimate the number of participants required. This analysis revealed that, to detect a small-to-medium sized effect (f=0.20; Cohen, 1988) for the critical interaction with a probability of 1-beta=.80 and alpha=.05, 42 participants were required.

Following this analysis, 42 students of Regensburg University took part in the experiment (M= 22.19 years, range: 18-30 years, 64.3% female). All participants spoke German as native language and took part on a voluntary basis. They received monetary reward or course credit for their participation.

Materials. For counterbalancing purposes, two study lists (A, B) were constructed, each containing 48 labels of line-drawing pictures selected from the Snodgrass and Vanderwart (1980) norms (see Appendix C). All items were high-imagery nouns. 30% of the items were selected as "living" and the rest as "nonliving". Items were chosen that elicited very high name agreement (98-100% according to the Snodgrass & Vanderwart norms) and had single word names. Two of the 48 items of a list served as primacy and two other items as recency items in this experiment. The remaining 44 items served as target items (see also Mulligan, 2006). All items were translated into German.

Design. The experiment had a 2 × 3 repeated measures design with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). Participants were tested on a study list 3 min after study (short delay) and after a delay of 24 hrs (long delay). At test, in both delay conditions subjects recalled the studied items in three successive free recall tests, which were separated by short distractor tasks. Assignments of conditions and lists were counterbalanced.

Procedure. Each participant completed two experimental blocks in counterbalanced order, one in the short and one in the long delay condition. The blocks were separated by a 5 min break, in which subjects played tetris. Prior to the study phase of each block, participants were informed that they would see a list of words and that they should try to rate the words whether they were "living" or "nonliving" (see Belmore, 1981). All words were presented individually on a screen for 5 s each and in random order. The entire list was presented twice in immediate succession (e. g., Mulligan, 2006). In the short delay condition, subjects were then asked to count backwards from a three-digit number for 3 min, while in the long delay condition, subjects were disbanded at this point and were asked to come back at the same time the next day. The test phase was identical for the two delay conditions. Participants completed three successive free recall tests, each lasting for 5 min. At the beginning of each test, a blank sheet was distributed with the instruction to report as many of the previously studied items as possible, independent of what they may have remembered in possible preceding tests. Between the tests, participants solved arithmetic problems for 3 min (see Fig. 9).

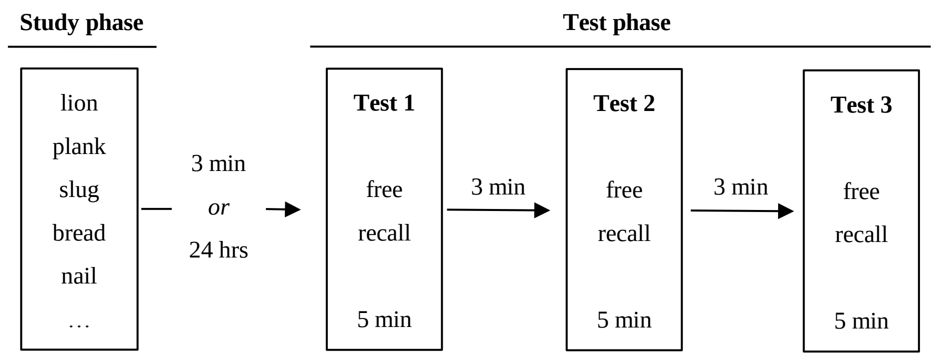

Figure 9. *Procedure and conditions employed in Experiment 4. Participants rated a list of words whether they were "living" or "nonliving" (see Belmore, 1981) and, after a 3-min or a 24-hrs delay, were asked to recall all items from the list in three successive free recall tests which were separated by distractor tasks of 3 min.*

Results

Separately for the short delay (3 min) and the long delay (24 hrs) conditions, Table 3 shows (i) net recall, i. e., number of correctly recalled words on each single test, (ii) item gains and item losses between test 1 and test 2, and between test 2 and test 3, and (iii) intrusion rates, i. e., number of recalled items not presented during study of the list.

Table 3. *Net recall, item gains, item losses, and intrusions in Experiment 4, separately for the short delay (3 min) and the long delay (24 hrs) condition.*

Words	*3-min delay*		*24-hrs delay*	
Free recall	*M*	*SD*	*M*	*SD*
Net recall				
Test 1	22.02	8.89	14.54	8.49
Test 2	21.95	9.13	15.58	9.29
Test 3	22.29	9.14	16.89	9.64
Gains and losses				
Gains 1 - 2	1.10	1.53	1.43	1.81
Gains 2 - 3	0.88	1.04	1.43	1.86
Losses 1 - 2	1.17	1.46	0.38	0.73
Losses 2 - 3	0.55	0.94	0.10	0.30
Intrusions				
Test 1	0.38	0.82	0.62	1.61
Test 2	0.40	0.86	1.14	2.18
Test 3	0.55	1.13	1.26	2.06

Net recall. A 2 × 3 analysis of variance (ANOVA) with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3) showed a main effect of DELAY, $F(1,41) = 22.45$, $MSE = 115.49$, $p < .001$, $\eta^2 = 0.35$, which demonstrates typical time-dependent forgetting. It also revealed a main effect of TEST, $F(2,82) = 12.69$, $MSE = 8.89$, $p < .001$, $\eta^2 = 0.24$, indicating hypermnesia. More important, there was a significant interaction between the two factors, $F(2, 82) = 11.36$, $MSE = 2.04$, $p < .001$, $\eta^2 = 0.22$, suggesting that the amount of increase in net recall across tests varied with delay. Consistently, two follow-up unifactorial ANOVAs with the within-subjects factor of TEST showed no significant main effect of TEST after 3 min, $F(2, 82) < 1$, but a significant main effect of TEST after 24 hrs, $F(2, 82) = 20.67$, $MSE = 2.82$, $p < .001$, $\eta^2 = 0.34$, suggesting that hypermnesia arose after the long but not the short delay. After 24 hrs, recall on the second test exceeded that on the first test, $t(41) = 3.50$, $p = .001$, $d = 0.12$, and recall on the third test exceeded that on the second test, $t(41) = 4.23$, $p < .001$, $d = 0.14$.

Item gains and item losses. Next, item gains and item losses across tests were analyzed. Gains on the second test were studied items reported on the second test but not on the first test, and gains on the third test were items reported on the third test but not on the second test. Likewise, losses on the second test were items reported on the first test but not the second, and losses on the third test were items reported on the second test but not the third. Regarding item gains, a 2 × 2 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) revealed no main effect of DELAY, $F(1, 41) = 3.53$, $MSE = 2.31$, $p = .067$, $\eta^2 = 0.08$, no main effect of TEST, $F(1, 41) < 1$, and no interaction between the two factors, $F(1, 41) < 1$. Regarding item losses, the same ANOVA showed a main effect of DELAY, $F(1, 41) = 21.01$, $MSE = 0.77$, $p < .001$, $\eta^2 = 0.34$, as well as a main effect of TEST, $F(1, 41) = 8.62$, $MSE = 0.10$, $p = .005$, $\eta^2 = 0.17$, with more losses in the short delay condition than in the long delay condition, and more losses between the first and the second test than between the second and the third test. The interaction was not significant, $F(1, 41) = 1.35$, $MSE = 0.86$, $p = .215$, $\eta^2 = 0.03$.

Intrusions. Analysis of intrusions may provide information on whether response criteria change across tests and delay conditions. Intrusions were analyzed with a 2 × 3 ANOVA with the within-subjects factors of DELAY (long, short) and TEST (test 1, test 2, test 3). It revealed significant main effects of DELAY, $F(1,41) = 4.35$, $MSE = 4.60$, $p = .043$, $\eta^2 = 0.10$, and TEST, $F(2,82) = 10.07$, $MSE = 0.36$, $p < .001$, $\eta^2 = 0.20$, showing that there were more intrusions after 24 hrs than after 3 min, and that intrusions increased across tests. There was also a significant interaction between the two factors, $F(2,82) = 4.95$, $MSE = 0.38$, $p = .009$, $\eta^2 = 0.11$, suggesting that delay enhances the increase in intrusions with repeated testing.[6]

Discussion

Results show an increase of net recall across tests reflecting typical hypermnesia. This increase, however, varied with the delay between study and test. Hypermnesia was larger after the long than the short delay and was even nonsignificant in the short delay condition. Moreover, the increase in hypermnesia with delay was primarily driven by reduced item losses across tests and was hardly affected by enhanced item gains. The findings on net recall are inconsistent with the cumulative recall hypothesis, which predicts reduced hypermnesia with prolongation of delay, but are consistent with the changes in cue set, the retrieval strategy, and the retrieval practice hypotheses. The finding that the effect is mainly due to a reduction in item losses but less, if at all, to enhanced item gains favors the retrieval practice hypothesis over the other two accounts (compare Table 2). Intrusions increased across tests and with delay, which at first points to changes in response criteria. It is unlikely that changes in response criteria mediated the effect of delay on hypermnesia in the present experiment, however. In fact, loosing the criterion with delay should increase item gains more than affecting item losses, which is not what the present results show. Before drawing more firm conclusions on the issue, it is the goal of Experiment 5 to replicate the present pattern of results

[6] Number of intrusions was fairly low in this experiment. Therefore, it was not differed between inter-list intrusions (words from block 1 intruding during recall in block 2) and intrusions caused by items not presented in any of the two experimental blocks. The same holds for Experiments 5 and 7.

2.5 Experiment 5: Delay Effects on Hypermnesia with Pictures as Study Material

A factor critically contributing to hypermnesia is stimulus material. Since Ballard's (1913) demonstration of the role of stimulus material for hypermnesia, many studies showed that hypermnesia effects arise fairly easy with some kind of stimulus material (e. g., lists of pictures; Erdelyi & Kleinbard, 1978; Madigan, 1976; Madigan & Lawrence, 1980), but may be harder to get with others (e. g., lists of unrelated words; Nelson & MacLeod, 1974; Tulving, 1967; Wilkinson & Koestler, 1983). In his review, Payne (1987) integrated 172 studies, and summarized that 96% of the experiments using simple pictures produced hypermnesia, whereas only 46% of the experiments using word lists did. Hence, the finding of nonsignificant hypermnesia with words in the short delay condition of Experiment 4 is not atypical in research on hypermnesia. Because hypermnesia is more readily found when pictures are used as study material and because words and pictures sometimes produce different results regarding hypermnesia (e. g., Erdelyi & Becker, 1974; Payne, 1986), it was the aim to repeat Experiment 4 with pictures as study material. The same set of items as in Experiment 4 was presented in Experiment 5, but showed the items' pictorial representations in the study phase. Doing so, reliable hypermnesia was expected in the short delay condition. The critical question then was if hypermnesia was again increased in the prolonged retention interval condition and whether such increase in net recall was again mainly driven by reduced item losses.

Method

Participants. Another 42 students of Regensburg University took part in the experiment (M= 22.14 years, range: 17-32 years, 64.3% female). All participants spoke German as native language and took part on a voluntary basis. Again, they received monetary reward or course credit for their participation.

Materials. The same two study lists (A, B) as in Experiment 4 were employed. However, in contrast to Experiment 4, not the labels of the pictures were presented in the study phase, but the line-drawings themselves (see Snodgrass & Vanderwart, 1980). Like in Experiment 4, the same four buffer items of each list were applied to control for primacy and recency effects.

Design and Procedure. Design and procedure were identical to Experiment 4, with the only exception that participants in the study phase were not instructed to rate the words to be "living" or "nonliving". Rather, participants were informed that they would see a list of pictures and that they should try to remember them for a later memory test (e. g., Mulligan, 2006; see Fig. 10).

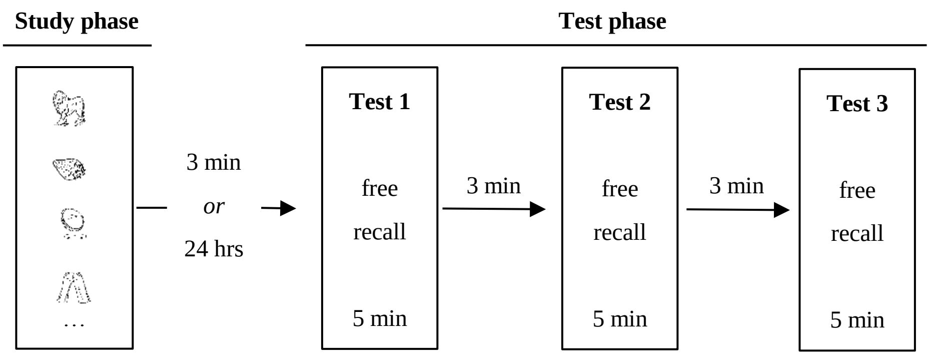

Figure 10. *Procedure and conditions employed in Experiment 5. Participants studied a list of pictures and, after a 3-min or a 24-hrs delay, were asked to recall the labels of all items from the list in three successive free recall tests which were separated by distractor tasks of 3 min.*

Results

Separately for the short delay (3 min) and the long delay (24 hrs) conditions, Table 4 shows (i) net recall, i. e., number of correctly recalled pictures on each single test, (ii) item gains and item losses between test 1 and test 2, and between test 2 and test 3, and (iii) intrusions on each single recall test.

Table 4. *Net recall, item gains, item losses, and intrusions in Experiment 5, separately for the short delay (3 min) and the long delay (24 hrs) condition.*

Pictures	*3-min delay*		*24-hrs delay*	
Free recall	*M*	*SD*	*M*	*SD*
Net recall				
Test 1	27.57	7.60	19.02	8.64
Test 2	27.74	8.27	20.12	8.76
Test 3	28.71	8.31	21.00	9.08
Gains and losses				
Gains 1 - 2	1.38	1.50	1.52	2.18
Gains 2 - 3	1.86	2.18	1.12	1.27
Losses 1 - 2	1.21	1.57	0.43	0.74
Losses 2 - 3	0.74	1.36	0.26	0.63
Intrusions				
Test 1	0.45	0.59	1.05	2.27
Test 2	0.48	0.67	1.19	2.88
Test 3	0.57	0.77	1.26	2.80

Net recall. The net recall data were scored using a conservative scoring method, in which the recalled name had to match the German translation of the picture name given by the Snodgrass and Vanderwart (1980) norms.[7] The net recall data were analyzed with a 2 × 3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). There was a main effect of DELAY, $F(1, 41) = 57.07$, $MSE = 69.97$, $p < .001$, $\eta^2 = 0.58$, showing typical time-dependent forgetting, and a main effect of TEST, $F(2, 82) = 23.12$, $MSE = 2.24$, $p < .001$, $\eta^2 = 0.36$, indicating increased recall across tests, i. e., hypermnesia. In addition, there was a significant interaction between the two factors, $F(2, 82) = 3.20$, $MSE = 1.71$, $p = .046$, $\eta^2 = 0.07$, suggesting that the test-induced increase in recall varied with delay condition. This held while there was significant hypermnesia in both delay conditions. In fact, two follow-up unifactorial ANOVAs with the within-subjects factor of TEST showed a significant main effect of test in both the short delay condition, $F(2, 82) = 8.63$, $MSE = 1.85$, $p < .001$, $\eta^2 = 0.17$, and the long delay condition, $F(2, 82) = 19.67$, $MSE = 2.09$, $p < .001$, $\eta^2 = 0.32$. In the short delay condition, recall on the first and the second tests did not differ significantly, $t(41) < 1$, but recall on the second and third tests did, $t(41) = 3.65$, $p = .001$, $d = 0.12$. In contrast, in the long delay condition both recall on the second test exceeded that on the first test, $t(41) = 3.10$, $p = .004$, $d = 0.13$, and recall on the third test exceeded that on the second, $t(41) = 4.03$, $p < .001$, $d = 0.10$.

Item gains and item losses. Regarding item gains, a 2 × 2 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) revealed no main effect of DELAY, no main effect of TEST, and no interaction between the factors, all $Fs(1, 41) < 2.25$, $MSEs < 3.62$, $ps > .141$, $\eta^2 s < 0.05$. The same ANOVA for item losses showed a significant main effect of

[7] There was an alternative, more liberal scoring method, in which the experimenter decided if the recalled name corresponded to a picture from the study list, even if the label was a variant of the normed name (e. g., Mulligan, 2006). The two scoring methods produced equivalent results, thus, only the data from the conservative scoring method are reported here. The not matched labels of recalled pictures, e. g., crab instead of lobster, were added to the number of intrusions.

DELAY, $F(1, 41) = 12.24$, $MSE = 1.37$, $p = .001$, $\eta^2 = 0.23$, suggesting that item losses in the short delay condition exceeded item losses in the long delay condition. There was a significant main effect of TEST, $F(1, 41) = 4.40$, $MSE = 0.99$, $p = .042$, $\eta^2 = 0.10$, indicating that item losses significantly decreased across tests, with more losses between the first and the second test than between the second and the third test. The interaction was not significant, $F(1, 41) = 1.04$, $MSE = 0.97$, $p = .314$, $\eta^2 = 0.03$.

Intrusions. Intrusions were analyzed with a 2 × 3 ANOVA with the within-subjects factors of DELAY (3 min, 24 hrs) and TEST (test 1, test 2, test 3). It revealed no main effect of DELAY, $F(1, 41) = 3.03$, $MSE = 9.23$, $p = .089$, $\eta^2 < 0.07$, no main effect of TEST, $F(2, 82) = 2.50$, $MSE = 0.23$, $p = .089$, $\eta^2 < 0.06$, and no interaction between the two factors, $F(2, 82) < 1$.

Discussion

Using pictures as stimulus material, the results of this experiment showed expected hypermnesia in the short delay condition. More important, like in Experiment 4, hypermnesia was influenced by the delay between study and test and was larger after the longer than the shorter delay. Also like in Experiment 4, this effect of delay was mainly driven by a reduction in item losses across tests in the long delay condition. There were no effects regarding intrusions, suggesting that, in this experiment, response criteria were roughly constant. The observed increase in net recall with delay is again consistent with the changes in cue set, the retrieval strategy, and the retrieval practice hypotheses, although the observed reduction in item losses favors the retrieval practice explanation of the present results.

Additional Analysis: Control of Order Effects

In contrast to the between-subjects design employed in the three extant studies on the issue (Dunning & Stern, 1992; Roediger & Payne, 1992; Wheeler & Roediger, 1992), in Experiments 4 and 5, each subject participated in both the short delay and the long delay conditions. Because this feature may have created order effects, the data of the two experiments were reanalyzed, this time including each subject's first block data only into the analysis. To maintain sufficient statistical power (see Methods of Experiment 4 above), the data of the two experiments were pooled to get again 42 participants in each delay condition. Table 5 shows (i) net recall, (ii) item gains and item losses, and (iii) intrusions for the pooled data.

Table 5. *Net recall, item gains, item losses, and intrusions pooled over the first experimental blocks of Experiments 4 and 5. Results are shown separately for the short delay (3 min) and the long delay (24 hrs) condition.*

Words + pictures	*3-min delay*		*24-hrs delay*	
Free recall	*M*	*SD*	*M*	*SD*
Net recall				
Test 1	24.38	8.34	16.27	7.98
Test 2	24.33	9.06	17.20	8.37
Test 3	25.05	9.06	17.98	8.55
Mean number of gains and losses				
Gains 1 - 2	1.17	1.40	1.33	1.78
Gains 2 - 3	1.29	1.52	1.00	1.25
Losses 1 - 2	1.21	1.44	0.40	0.77
Losses 2 - 3	0.45	0.94	0.24	0.58
Intrusions				
Test 1	0.48	0.80	1.14	2.61
Test 2	0.40	0.83	1.64	3.46
Test 3	0.52	1.04	1.76	3.21

Statistical analysis of the pooled data replicated the main results for the two single experiments. Regarding net recall, a 2 × 3 ANOVA with the within-subjects factor of TEST (test 1, test 2, test 3) and the between-subjects factor of DELAY (short, long) showed a main effect of DELAY, $F(1, 82) = 16.13$, $MSE = 216.04$, $p < .001$, $\eta^2 = 0.16$, a main effect of TEST, $F(2, 164) = 14.66$, $MSE = 2.08$, $p < .001$, $\eta^2 = 0.15$, and a significant interaction between the two factors, $F(2, 164) = 3.46$, $MSE = 2.08$, $p = .034$, $\eta^2 = 0.04$. Recall increased across tests in the long delay condition, $F(2, 82) = 15.09$, $MSE = 2.05$, $p < .001$, $\eta^2 = 0.27$, and in the short delay condition, $F(2, 82) = 3.19$, $MSE = 2.10$, $p = .046$, $\eta^2 = 0.07$. In the long delay condition, recall on the second test exceeded that on the first test, $t(41) = 3.18$, $p = .003$, $d = 0.11$, and recall on the third test exceeded that on the second, $t(41) = 3.50$, $p = .001$, $d = 0.09$. In the short delay condition, recall on the second test did not differ to that on the first test, $t(41) < 1$, but recall on the third test exceeded that on the second, $t(41) = 3.06$, $p = .004$, $d = 0.08$.

Regarding item gains, a 2 × 2 ANOVA with the between-subjects factor of DELAY (short, long) and the within-subjects factor of TEST (test 2, test 3) revealed no main effects, both $Fs(1, 82) < 1$, and no interaction between the two factors, $F(1, 82) = 1.26$, $MSE = 1.72$, $p = .267$, $\eta^2 = 0.02$. The same ANOVA for item losses showed a significant main effect of DELAY, $F(1, 82) = 8.98$, $MSE = 1.23$, $p = .004$, $\eta^2 = 0.10$, and a significant main effect of TEST, $F(1, 82) = 12.64$, $MSE = 0.72$, $p < .001$, $\eta^2 = 0.13$, indicating that item losses in the short delay condition exceeded item losses in the long delay condition and that there were more losses between test 1 and test 2 than between test 2 and test 3. There was also a significant interaction between the two factors, $F(1, 82) = 5.20$, $MSE = 0.72$, $p = .025$, $\eta^2 = 0.06$, suggesting that the reduction in item losses in the long delay condition was present mainly from the first to the second recall test. At least numerically, this same interaction was also present in the two single experiments reported above.

Regarding intrusions, a 2 × 3 ANOVA with the between-subjects factors of DELAY (long, short) and the within-subjects factor of TEST (test 1, test 2, test 3) showed significant main effects of DELAY, $F(1, 82) = 4.62$, $MSE = 14.98$, $p = .035$, $\eta^2 = 0.05$, and TEST, $F(2, 164) = 6.41$, $MSE = 0.37$, $p = .002$, $\eta^2 = 0.07$, suggesting that there were more intrusions after a long delay and that intrusions raised across tests. Like in Experiment 1, there was also a significant interaction between the two factors, $F(2, 164) = 6.12$, $MSE = 0.37$, $p = .003$, $\eta^2 = 0.07$.

2.6 Experiment 6: Delay Effects on Hypermnesia with Forced Recall Testing

The results of Experiments 4 and 5 disagree with those reported in the two previous studies by Dunning and Stern (1992) and Roediger and Payne (1982), who reported no effect of delay on hypermnesia. Still, they are not in direct conflict with these previous findings. In fact, the present experiments included larger samples of subjects than Dunning and Stern's study did, and they employed longer retention intervals than Roediger and Payne's study did, which may account for the difference in results. However, there is a possible conflict between the results of present Experiments 4 and 5 and those reported by Wheeler and Roediger (1992), who across three successive tests observed hypermnesia after a short delay but no hypermnesia after a prolonged delay.

There are several methodological differences between the present experiments and the one reported in Wheeler and Roediger (1992). For instance, Wheeler and Roediger employed a short delay of 11.5 min and a long delay of 1 week, whereas, in the present experiments, the short delay lasted 3 min and the long delay 24 hrs; Wheeler and Roediger tested subjects in groups, ranging in size from 3 to 9, whereas in the present experiments, subjects were tested individually; and Wheeler and Roediger presented 60 items for study, which were shown in the same serial order to all subjects, whereas here 44 items were presented in a random order. It is speculated that these differences are not at the core of the conflict in results.

A more critical methodological difference between studies may be recall format. Whereas in the present experiments, free recall tests were applied across the series of recall tests, Wheeler and Roediger employed forced recall tests. In these tests, subjects were given recall sheets with a separate line for each single to-be-recalled item and were asked to recall as many items as possible. In particular, if unable to remember all studied items, subjects should fill in the remaining spaces with their best guesses. Although there is evidence that recall format does not influence hypermnesia after a short delay (Roediger & Payne, 1985), an influence after long delay can not be excluded. For instance, allowing subjects to fill in the remaining spaces of a recall sheet with their best guesses may not much reduce subjects' effort to recall further previously studied items after a short delay, when recall is still relatively easy. But it may do so after a prolonged delay when recall becomes more demanding. If so, free and forced recall may lead to similar hypermnesia after short delay, but free recall may lead to higher hypermnesia than forced recall after prolonged delay. Experiments 6 and 7 examined the possible role of recall format for hypermnesia directly.

There were two goals with Experiments 6 and 7. The goal of Experiment 6 was to replicate Wheeler and Roediger's (1992) finding of decreased hypermnesia with delay using forced recall at test, the same number of study items, and the same delay intervals as were used in their previous study. The goal of Experiment 7 then was to examine whether forced recall was critical for the results of Experiment 6 and whether results would change if a free recall format was applied at test. If recall format was the critical difference between the present Experiments 4 and 5 and the experiment reported in Wheeler and Roediger (1992), then the results of Experiment 6 using forced recall should replicate those of Wheeler and Roediger (1992) and the results of Experiment 7 using free recall should replicate those of Experiments 4 and 5.

Experiment 6 examined the role of delay for hypermnesia, closely following the methods employed by Wheeler and Roediger (1992). Subjects were presented 60 pictures and, after a short delay of 11.5 min or a long delay of 1 week, were asked to recall the study items. In both delay conditions, three successive recall tests were conducted, each test using a forced recall format, thus deviating from the recall format used in Experiments 4 and 5 above. The expectation was to replicate the results by Wheeler and Roediger

(1992) and to find hypermnesia after the short delay but no hypermnesia after the long delay.

Method

Participants. On the basis of the analysis of test power in Experiment 4 and because of counterbalancing purposes, 48 students of Regensburg University participated in the experiment (M = 20.83 years, range: 19-30 years, 77.1% female). All participants spoke German as native language and took part on a voluntary basis. Again, they received monetary reward or course credit for their participation.

Materials. The two study lists (A, B) of Experiments 4 and 5 were extended by adding 12 further line-drawing pictures from the Snodgrass and Vanderwart (1980) norms to each single list (see Appendix D). Doing so, list length became equal to that applied in Wheeler and Roediger (1992). Like in this previous study, there was no control for primacy and recency effects in Experiment 6.

Design and Procedure. The design of the experiment was identical to Experiments 4 and 5. Each participant completed two experimental blocks, one in the short and one in the long delay condition in counterbalanced order. Again, the blocks were separated by a 5 min break, in which subjects played tetris. All 60 line-drawings were presented individually on a screen for 7 s each in random order. With presentation, the label of the drawing was enunciated by the experimenter. Each list was presented once. Following Wheeler and Roediger's (1992) procedure, after study, subjects in both delay conditions recalled as many U.S. presidents (one experimental block) or capital cities (other experimental block) as they could. They were then given a questionnaire on which they guessed how many pictures they had seen, how long each picture had appeared, and the total length of the entire presentation. In addition, they were asked to recall the instructions they had received before item presentation. Doing so, a delay of 11.5 min arose before subjects in the short delay condition were tested. In the long delay condition, subjects were disbanded at this point and asked to return at the same time 7 days later.

At test, subjects completed three successive forced recall tests, each lasting for 7 min, with a 1 min break between tests. The experimenter distributed test sheets, with lines numbered 1 to 60 with the instruction to the subjects to recall as many of the previously studied items as possible, independent of what they may have remembered in possible preceding tests. If they felt unable to remember all 60 objects, they should fill the remaining spaces with their best guesses. If the 60 spaces were not complete after 7 min, the subjects were instructed to fill in the remaining spaces as quickly as possible, thus again following Wheeler and Roediger's (1992) procedure (see Fig. 11).

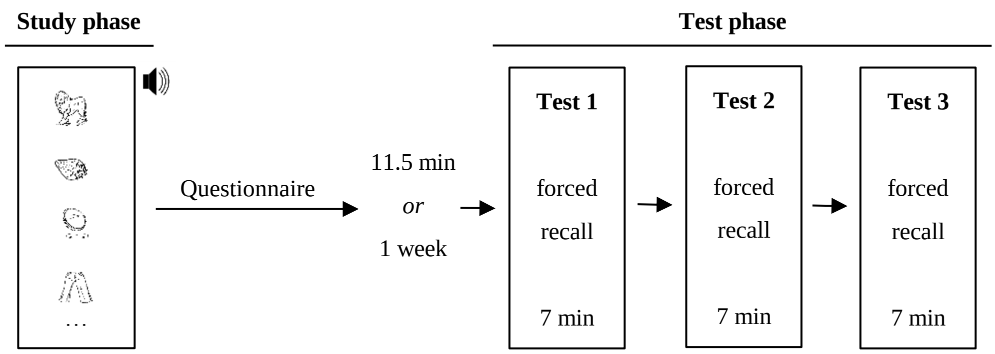

Figure 11. *Procedure and conditions employed in Experiment 6. Participants studied a list of pictures and completed a questionnaire on the study phase. After a 11.5-min or a 1-week delay, they were asked to recall the labels of all items from the list in three successive forced recall tests of 7 min each.*

Results

Table 6 shows, separately for the short delay (11.5 min) and the long delay (1 week) conditions, (i) net recall on each single test, (ii) item gains and item losses between test 1 and test 2, and between test 2 and test 3, and (iii) intrusions on each single test.

Table 6. *Net recall, item gains, item losses, and intrusions in Experiment 6, separately for the short delay (11.5 min) and the long delay (1 week).*

Pictures	*11.5-min delay*		*1-week delay*	
Forced recall	*M*	*SD*	*M*	*SD*
Net recall				
Test 1	30.00	7.85	15.29	6.29
Test 2	31.00	8.18	15.79	6.44
Test 3	32.23	8.41	15.96	6.40
Gains and losses				
Gains 1 - 2	2.88	2.18	2.54	1.88
Gains 2 - 3	2.50	1.91	2.04	1.43
Losses 1 - 2	1.88	1.66	2.04	1.88
Losses 2 - 3	1.38	1.38	1.85	1.60
Intrusions				
Test 1	21.54	12.64	33.29	16.07
Test 2	21.88	12.72	36.94	15.37
Test 3	21.10	11.86	36.94	15.36

Net recall. The net recall data were analyzed by means of a 2 × 3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). There was a main effect of DELAY, $F(1, 47) = 205.20$, $MSE = 83.17$, $p < .001$, $\eta^2 = 0.81$, showing typical time-dependent forgetting, and a main effect of TEST, $F(2, 94) = 19.18$, $MSE = 2.62$, $p < .001$, $\eta^2 = 0.29$, indicating increased recall across tests, i. e., hypermnesia. In addition, there was a significant interaction between the two factors, $F(2,94) = 4.14$, $MSE = 3.69$, $p = .019$, $\eta^2 = 0.08$, suggesting that the test-induced increase in recall varied with delay. In fact, two follow-up unifactorial ANOVAs with the within-subjects factor of TEST showed a significant main effect in the short delay condition, $F(2,94) = 18.01$, $MSE = 3.32$, $p < .001$, $\eta^2 = 0.28$, but no such effect in the long delay condition, $F(2,94) = 1.93$, $MSE = 2.99$, $p = .171$, $\eta^2 = 0.04$. In the short delay condition, recall on the second test exceeded that on the first test, $t(47) = 2.75, p = .008, d = 0.12$, and recall on the third test exceeded that on the second test, $t(47) = 3.99$, $p = .001$, $d = 0.15$.

Item gains and item losses. Regarding item gains, a 2 × 2 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) revealed a main effect of TEST, $F(1,47) = 4.39$, $MSE = 2.09$, $p = .042$, $\eta^2 = 0.09$, indicating more gains between test 1 and test 2 than between test 2 and test 3. There was no main effect of DELAY, $F(1,47) = 2.46$, $MSE = 3.06$, $p = .124$, $\eta^2 = 0.05$ and no interaction between the two factors, $F(1,47) < 1$. The same analysis for item losses showed no main effect of TEST, $F(1,47) = 3.78$, $MSE = 1.50$, $p = .058$, $\eta^2 = 0.07$, no main effect of DELAY, $F(1,47) = 2.19$, $MSE = 2.28$, $p = .145$, $\eta^2 = 0.05$, and no interaction between the factors, $F(1,47) < 1$.

Intrusions. As expected from the nature of the forced recall test, intrusion rates were high in this experiment. Intrusions were analyzed by means of a 2 × 3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). The analysis revealed significant main effects of DELAY, $F(1,47) = 80.03$, $MSE = 130.00$, $p < .001$, $\eta^2 = 0.63$, and TEST, $F(2,94) = 4.73$, $MSE = 23.20$, $p = .011$, $\eta^2 = 0.09$, showing that, unsurprisingly, there were more intrusions after 1 week than after 11.5 min, and that intrusions differed across tests. There was also a significant interaction between the two factors, $F(2,94) = 7.00$, $MSE = 16.80$, $p = .001$, $\eta^2 = 0.13$, reflecting the fact that intrusions in the long, but not the short delay condition, increased across tests.[8]

[8] It was not possible to push each subject to fill in all missing spaces of the test sheets.

Discussion

Using forced recall at test, the same number of study items, and the same delay conditions as employed in Wheeler and Roediger (1992), the results of this experiment replicate Wheeler and Roediger's prior finding. While net recall increased significantly across tests in the short delay condition, repeated testing left net recall largely unaffected in the long delay condition. Analysis of item gains and item losses did not reveal significant effects of delay, but there were numerical trends for higher item gains and lower item losses after the short delay, which together created the significant effect of delay on hypermnesia. Wheeler and Roediger did not report item gains and item losses, so there is no way to compare the present results on gains and losses with the prior work.

Experiments 4 and 5 on the one hand and Experiment 6 on the other differ in more than one methodological detail. But if recall format was the main methodological difference, then the difference in results between Experiments 4 and 5 and Experiment 6 suggests that recall format can influence the effect of delay on hypermnesia. Whereas both recall formats may create hypermnesia after short delay, after long delay, free recall may increase hypermnesia even further, while forced recall may decrease, or even eliminate, the effect. Experiment 7 examines this proposal directly.

2.7 Experiment 7: Delay Effects on Hypermnesia and the Role of Testing Format

Experiment 7 repeated Experiment 6 but replaced the forced recall format

As a result, in none of the single conditions, do mean net recall and mean intrusions sum up to 60 (see Table 6).

of Experiment 6 by the free recall format used in Experiments 4 and 5. Thus, again subjects were presented 60 pictures and their labels and, after a short delay of 11.5 min or a long delay of 1 week, were asked to recall the labels of the studied pictures. After the delay, three successive free recall tests were conducted. It was expected to replicate the finding of Experiment 6 of significant hypermnesia after the short delay. However, in contrast to Experiment 6, it was expected to find an increase in hypermnesia in the long delay condition, mainly driven by reduced item losses. Such pattern of results would mimic the findings of Experiments 4 and 5, indicating that, with free recall, delay can increase hypermnesia. In addition, the same pattern would suggest that recall format can be critical for hypermnesia and influence whether delay has a beneficial or a detrimental effect on hypermnesia.

Method

Participants. Another 48 students of Regensburg University participated in this experiment (M= 20.48 years, range: 18-26 years, 68.8% female). All participants spoke German as native language and took part on a voluntary basis. They received monetary reward or course credit for their participation.

Materials, Design, and Procedure. Materials and design were identical to Experiment 6. The procedure was also largely identical. However, unlike in Experiment 6, a free recall format was employed at test. At the beginning of each test, a blank sheet was distributed with the instruction to report as many of the previously studied items as possible, independent of what they may have remembered in possible preceding tests. Guessing was not encouraged (see Fig. 12).

Results

Table 7 shows, separately for the short delay (11.5 min) and the long delay (1 week) conditions, (i) net recall on each single test, (ii) item gains and item

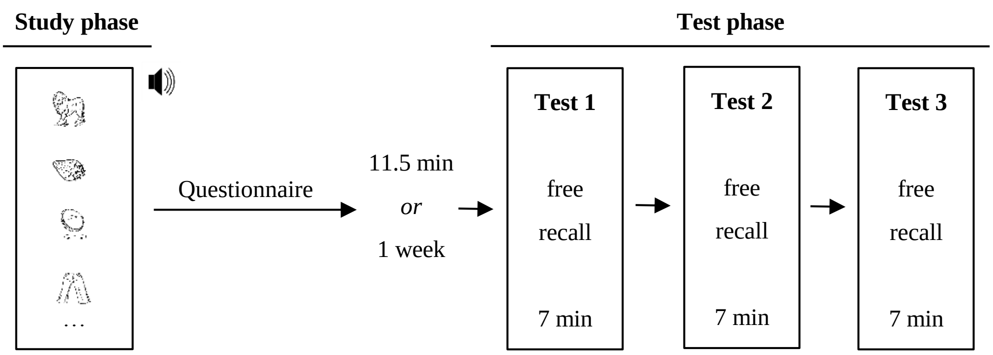

Figure 12. *Procedure and conditions employed in Experiment 7. Participants studied a list of pictures and completed a questionnaire on the study phase. After a 11.5-min or a 1-week delay, they were asked to recall the labels of all items from the list in three successive free recall tests of 7 min each.*

losses between test 1 and test 2, and between test 2 and test 3, and (iii) intrusions on each single test.

Table 7. *Net recall, item gains, item losses, and intrusions in Experiment 7, separately for the short delay (11.5 min) and the long delay (1 week).*

Pictures	*11.5-min delay*		*1-week delay*	
Free recall	*M*	*SD*	*M*	*SD*
Net recall				
Test 1	27.27	8.62	12.38	7.14
Test 2	28.13	9.12	13.31	7.43
Test 3	28.75	9.33	15.00	8.25
Gains and losses				
Gains 1 - 2	2.46	2.21	1.44	1.50
Gains 2 - 3	1.10	1.82	1.92	2.20
Losses 1 - 2	1.60	1.51	0.50	0.74
Losses 2 - 3	1.44	1.57	0.25	0.53
Intrusions				
Test 1	1.00	1.19	4.23	4.46
Test 2	1.10	1.32	5.29	4.66
Test 3	1.83	2.66	5.94	5.59

Net recall. The net recall data were analyzed by means of a 2 × 3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). There was a main effect of DELAY, $F(1, 47) = 108.39$, $MSE = 139.40$, $p < .001$, $\eta^2 = 0.70$, showing typical time-dependent forgetting, and a main effect of TEST, $F(2, 94) = 35.06$, $MSE = 2.90$, $p < .001$, $\eta^2 = 0.43$, indicating increased recall across tests. In addition, there was a significant interaction between the two factors, $F(2, 94) = 4.21$, $MSE = 2.33$, $p = .018$, $\eta^2 = 0.08$, indicating that the test-induced increase in recall was larger in the long than the short delay condition. This holds while there was significant hypermnesia in both delay conditions. Indeed, two follow-up unifactorial ANOVAs with the within-subjects factor of TEST showed a significant main effect of TEST in both the short delay condition, $F(2, 94) = 9.39$, $MSE = 2.82$, $p = .001$, $\eta^2 = 0.17$, and the long delay condition, $F(2, 94) = 35.31$, $MSE = 2.41$, $p < .001$, $\eta^2 = 0.43$. In the short delay condition, recall on the second test exceeded that on the first test, $t(47) = 2.70$, $p = .010$, $d = 0.10$, but recall on the third test did not exceed that on the second, $t(47) = 1.96$, $p = .056$, $d = 0.07$. In contrast, in the long delay condition both recall on the second test exceeded that on the first test, $t(47) = 4.05$, $p < .001$, $d = 0.13$, and recall on the third test exceeded that on the second, $t(47) = 5.43$, $p < .001$, $d = 0.22$.

Item gains and item losses. Regarding item gains, a 2 × 2 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) revealed a main effect of DELAY, $F(1, 47) = 6.36$, $MSE = 2.76$, $p = .015$, $\eta = 0.12$, indicating that there were more gains after the short than the long delay. The main effect of TEST and the interaction between the factors were not significant, both $Fs(1, 47) < 2.85$, $MSEs > 2.16$, $ps > .056$, $\eta^2 s < 0.08$. The same ANOVA for item losses showed a significant main effect of DELAY, $F(1, 47) = 51.09$, $MSE = 1.23$, $p < .001$, $\eta^2 = 0.52$, suggesting that item losses in the short delay condition exceeded item losses in the long delay condition. The main effect of TEST and the interaction between the two factors were nonsignificant, both $Fs(1, 47) < 1.63$, $MSEs > 1.28$, $ps > .207$, $\eta^2 s < 0.03$.

Intrusions. A 2 × 3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3) revealed significant main effects of DELAY, $F(1, 47) = 41.68$, $MSE = 25.47$, $p < .001$, $\eta^2 = 0.47$, and TEST, $F(2, 94) = 14.16$, $MSE = 2.74$, $p < .001$, $\eta^2 = 0.23$, showing that there were more intrusions after 1 week than after 11.5 min, and that intrusions increased across tests. The interaction was not significant, $F(2, 94) = 2.17$,

$MSE = 3.12$, $p = .120$, $\eta^2 = 0.04$.

Discussion

The results of the experiment demonstrate hypermnesia in both delay conditions, but the effect was larger in the long than the short delay condition. This effect of delay was driven by a reduction in item losses across tests in the long delay condition. The reduction in item losses was numerically larger than the simultaneously observed reduction in item gains, which is why an increase in net recall arose with delay. Intrusions also increased with delay. Again, this increase could reflect a more liberal recall threshold in the long than the short delay condition and thus, in principle, could underlie the observed increase in hypermnesia. However, as already mentioned in the discussion of Experiment 4, there is reason to reject such proposal, because loosing the criterion across tests should increase item gains more than affecting item losses, which is not what the present results show. All in all, the results of Experiment 4 thus mimic those of Experiments 4 and 5 above and indicate that, with free recall at test, delay can increase hypermnesia.

The results of Experiment 7 clearly differ from those of Experiment 6. While the results of Experiment 7 show that usage of a free recall format can lead to an increase in hypermnesia with delay, the results of Experiment 6 show that usage of a forced recall format can lead to a decrease with delay. This holds while the two recall formats lead to similar results after short retention interval. Prior work already demonstrated that recall format has no major influence on hypermnesia after short delay (Roediger & Payne, 1985). The present results support this equivalence proposal, but they also show that the proposal does no longer hold when retention interval is prolonged.

Additional Analysis: The Role of Testing Format

The results of Experiments 6 and 7 above suggest similar hypermnesia for free and forced recall after the short retention interval, but different hypermnesia for the two recall formats after the long retention interval. Statistical analyses support this suggestion.

Short delay conditions. Regarding net recall, a 2 × 3 ANOVA with the between-subjects factor of RECALL FORMAT (forced, free) and the within-subjects factor of TEST (test 1, test 2, test 3) revealed a main effect of TEST, $F(2,188) = 26.87$, $MSE = 3.07$, $p < .001$, $\eta^2 = 0.22$, indicating hypermnesia. The effect of RECALL FORMAT, $F(1,94) = 3.06$, $MSE = 215.80$, $p = .084$, $\eta^2 = 0.03$, and the interaction, $F(2,188) = 1.24$, $MSE = 3.07$, $p = .291$, $\eta^2 = 0.01$, were nonsignificant. Regarding both item gains and item losses, a 2 × 2 ANOVA with the factors of RECALL FORMAT (forced, free) and TEST (test 2, test 3) showed no main effects and no interactions, all $Fs(1,94) < 2.98$, $MSEs > 1.79$, $ps > .088$, $\eta^2 s < 0.03$.

Long delay conditions. Regarding net recall, a 2 × 3 ANOVA with the factors of RECALL FORMAT (forced, free) and TEST (test 1, test 2, test 3) revealed a main effect of TEST, $F(2,188) = 24.22$, $MSE = 2.70$, $p < .001$, $\eta^2 = 0.21$, but no main effect of RECALL FORMAT, $F(1,94) = 2.26$, $MSE = 142.72$, $p = .136$, $\eta^2 = 0.02$. The interaction was significant, $F(2,188) = 29.40$, $MSE = 2.70$, $p < .001$, $\eta^2 = 0.09$, pointing to higher hypermnesia with free recall testing. Regarding item gains, a 2 × 2 ANOVA with the factors of RECALL FORMAT and TEST revealed a main effect of RECALL FORMAT, $F(1,94) = 4.58$, $MSE = 3.96$, $p = .035$, $\eta^2 = 0.05$, with more item gains with forced than free recall testing. There was no main effect of TEST, $F(1,94) < 1$, but a significant interaction, $F(1,94) = 4.85$, $MSE = 2.37$, $p = .030$, $\eta^2 = 0.05$. Regarding item losses, the same ANOVA showed a main effect of RECALL FORMAT, $F(1,94) = 57.62$, $MSE = 2.06$, $p = .001$, $\eta^2 = 0.38$, with less item losses with free than forced recall testing. There was no main effect of TEST and no interaction, all $Fs(1,94) < 1.65$, $MSEs > 1.40$, $ps > .203$, $\eta^2 s < 0.02$.

2.8 Discussion of Experiments 4-7

The results of the four experiments in chapter 2 are summarized in Fig. 13 for comparison of the hypermnesia effects after a short and a long interval between study and test, between employing words and pictures as study material and between free and forced recall testing.

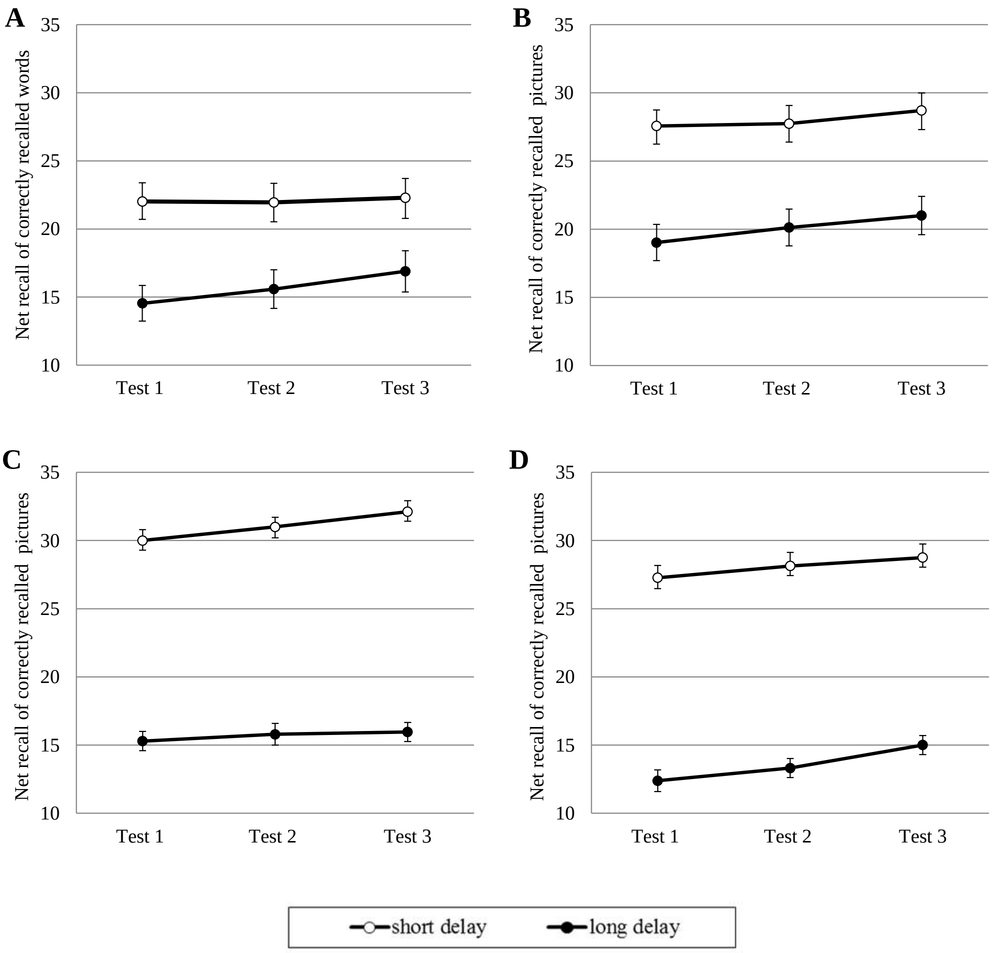

Figure 13. **(A)** *Results of Experiment 4,* **(B)** *Experiment 5,* **(C)** *Experiment 6, and* **(D)** *Experiment 7. Net recall of correctly recalled words (Experiment 4) and pictures (Experiment 5-7) is shown as a function of delay (short, long), and test (test 1, test 2, test 3). Error bars represent standard errors.*

These experiments replicate prior work by showing that net recall increases with multiple tests, and that this effect can be larger with pictures as stimuli than with words (e. g., Erdelyi & Becker, 1974; Madigan & Lawrence, 1976). Going beyond the prior work, the present results show that the delay between study and test can influence hypermnesia. Indeed, when free recall was used as testing format, hypermnesia was larger after a long delay of 24 hrs (Experiments 4 and 5) or 7 days (Experiment 7) than after a short delay of 3 min (Experiments 4 and 5) or 11.5 min (Experiment 7). Moreover, in all three experiments, the delay-induced influence on hypermnesia was driven mainly by differences in item losses, with less previously recalled items being forgotten between tests in the long delay than the short delay condition. There was no increase in item gains with delay. Together, these results indicate that a longer delay between study and test can increase hypermnesia and does so primarily by reducing the forgetting across recall tests.

The present experiments also show that recall format can influence the effect of delay on hypermnesia. Employing forced recall (Experiment 6) instead of free recall (Experiment 7) at test, the results firstly showed equivalent hypermnesia in the two recall formats after short delay, which replicates prior work (Roediger & Payne, 1992). Increasing the delay, however, led to nonequivalent hypermnesia effects, with an increase in hypermnesia with free recall testing and a decrease with forced recall testing (see also Wheeler & Roediger, 1992). The decrease was reflected in both reduced item gains and increased item losses, although both effects were present numerically only but not statistically. These findings suggest a role of recall format in hypermnesia, indicating that different mechanisms may mediate the effects of repeated testing in the two recall conditions.

Implications of the Free Recall Results for the Accounts of Hypermnesia

The present free recall results on net recall are consistent with the changes in cue set hypothesis (Raaijmakers & Shiffrin, 1980; Roediger & Thorpe, 1978), the retrieval strategy hypothesis (Erdelyi & Becker, 1974; Mulligan, 2001), and the retrieval practice hypothesis (Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a). On the basis of the changes in cue set hypothesis, hypermnesia is expected to increase with delay if, after delay-induced contextual drift, retrieval of some items reactivates the study context and such reactivation is not completed at the end of the first test but extends to later recall tests. Because context reactivation can change the cue set that people use to sample and recover items, it can induce alternative retrieval routes on later recall tests, which may enhance item gains and increase hypermnesia. According to the retrieval strategy hypothesis, enhanced organization across repeated tests leads to hypermnesia, improving recall by increasing item gains and reducing item losses. If such organization was further advanced after longer delay, for instance, because retrieval after delay becomes more challenging, then item gains should be further enhanced with delay and item losses be limited, again increasing hypermnesia. Also the retrieval practice hypothesis can account for the present free recall results. Because retrieval practice should be more demanding after longer than after shorter delay, and retrieval practice effects have been shown to be particularly strong if retrieval practice is demanding, retrieval after longer delay may lead to enhanced hypermnesia by reducing the forgetting of the initially recalled items. In contrast, the present finding of increased hypermnesia with delay is not easily reconciled with the cumulative recall hypothesis (Roediger & Challis, 1989; Roediger et al., 1982). This hypothesis claims that study conditions producing high levels of asymptotic recall should induce more hypermnesia than conditions producing lower levels of recall, which is the opposite of what the present free recall results show.

Although the changes in cue set hypothesis, the retrieval strategy hypothesis, and the retrieval practice hypothesis are consistent with the finding of increased hypermnesia with delay, they differ in the degree to which they can explain the observed presence of a delay effect in item losses and the observed absence of a delay effect in item gains. The changes in cue set hypothesis is largely focused on item gains and thus explains the effect of delay mainly by attributing it to enhanced item gains. The retrieval strategy hypothesis makes assumptions about both item gains and item losses and suggests a beneficial effect of delay on item gains and a detrimental one on item losses. Finally, the retrieval practice hypothesis focuses mainly on item losses and thus explains the effect of delay primarily by a reduction in item losses. The present finding that, with free recall as testing format, delay increases hypermnesia mainly by reducing item losses thus favors the retrieval practice hypothesis, indicating that retrieval practice effects can contribute to hypermnesia and do so particularly when the delay between study and test is increased.

The finding of Experiments 4, 5, and 7 that the increase in hypermnesia with delay is due to a reduction in item losses, arose by analyzing absolute differences in recall levels between tests, which is typical for prior work on hypermnesia (e. g., Dunning & Stern, 1992; Mulligan, 2005; Wheeler & Roediger, 1992; but see Goernert, Widner, & Otani, 2007). However, one may also take a different view on the issue. Indeed, because after prolonged delay, fewer items are recalled than after short delay (see Tables 3-7), one could argue that there are also fewer items to drop between tests after the longer delay, which raises the question of whether the results reported in Experiments 4, 5, and 7 above would replicate if a measure of proportion of items dropped was employed for analysis. Using such proportion measure, corresponding analyses showed that the pattern of results outlined earlier indeed replicates and item losses remain reduced with delay in each of the three experiments.[9] The main

[9] Regarding item losses, 2 × 2 ANOVAs with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) showed a significant main effect of DELAY in all three experiments, $F's > 6.88$, $MSEs < 40.17$, $ps < .012$, $\eta' s^2 > 0.14$, indicating that losses were indeed reduced after prolonged delay. Regarding item gains, the pattern of results reported above was also replicated.

results of the present experiments of chapter 2 thus do not depend on whether absolute or proportion measures are used for analysis.

Relation of the Free Recall Results to Prior Work

The present finding that, with free recall as testing format, longer delay increases hypermnesia disagrees with the results of two previous studies that also examined the role of delay between study and test for hypermnesia and found no effect of delay. In the one study, Dunning and Stern (1992; Experiment 2) investigated hypermnesia in eyewitness memory using films about crime scenes as stimulus material and employing a single experiment with, on average, less than 10 subjects per condition. The present study reports the results of three experiments with at least 42 subjects per condition, using both words and pictures as stimulus material. While it cannot be excluded that stimulus material can affect hypermnesia results (e. g., Ballard, 1913), it appears more likely that the difference in results between the previous study and present experiments on free recall has to do with the difference in statistical power, in particular, as the statistical power employed in Dunning and Stern's experiment should have been too low to detect a possible effect of delay on hypermnesia (see also Methods of Experiment 4 above). In the other study, Roediger and Payne (1982) reported another single experiment, in which delay was manipulated by conducting the initial recall test immediately after study, or after subjects read a prose passage for 18 min. Because, in contrast to the present study, delay did not induce any time-dependent forgetting in this previous study, the difference in results between the previous study and the present experiments on free recall may reflect the difference in degree to which the employed delay manipulations were effective. The present results thus are not in direct conflict with the results from these two previous studies and may rather indicate that, in order to observe an effect of delay on hypermnesia, sufficient statistical power and a delay interval that induces robust time-dependent forgetting are required.

Also the present results of chapter 2 are in line with the testing effect literature, which shows that retrieval practice can improve recall of practiced items and does so even more if retrieval practice is demanding (e. g., Bäuml et al., 2014; Carpenter, 2011; Halamish & Bjork, 2011; Pyc & Rawson, 2009). In particular, this literature has shown that retrieval practice can reduce the forgetting of practiced items and thus enhance longterm-retention (e. g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a). Using a different paradigm, the present Experiments 4-7 reveal a similar pattern by showing that, after longer delay between study and test, retrieval practice on an initial recall test can reduce the forgetting of practiced items on subsequent recall tests relative to a short delay condition. Enhanced hypermnesia after longer delay can thus serve as another demonstration of the role of difficulty of retrieval practice task for beneficial effects of retrieval practice. Moreover, on the basis of the testing effect literature, the present findings also suggest that hypermnesia may be enhanced whenever the initial test is demanding. Hypermnesia may thus be increased not only after longer delay, but also after a change in context between study and test, or in the presence of interference. Future work may investigate the issue in more depth.

Results from several recent studies (e. g., Bäuml & Schlichting, 2014) and especially the present Experiments 1-3 suggest that, after longer delay between study and test, selective retrieval can improve recall of other items. The finding was interpreted as evidence that, after a delay and induced context change (e. g., Bower, 1972; Estes, 1955), retrieval of some first items reactivates the items' study context, which then serves as a retrieval cue for the remaining items and improves recall performance (see also Bäuml & Samenieh, 2012; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988). Critically, if context reactivation was still incomplete at the end of the first recall test, reactivation might still operate on the subsequent test, leading to retrieval of further items and increased hypermnesia. The present findings do not show such an increase in item gains, however. Increases in recall due to reactivated context thus may be largely restricted to the first test and not easily extend to subsequent recall tests. One reason might be that in typical hypermnesia experiments, pictorial study material is used, be it actual pictures or even words that should be imagined pictorially. Those type of items may reactivate the original study context very fast, i. e., even at the first recall trial, so that there is no additional benefit by dint of the following recall trials.

Aside from the fact that the beneficial effect of selective retrieval and the beneficial effect of nonselective retrieval in hypermnesia both represent retrieval-induced benefits in recall, they differ in their focuses: While the beneficial effect of selective retrieval is restricted to retrieval-induced benefits on related material, hypermnesia indeed also bears on such benefits by involving item gains across repeated recall tests, but furthermore, it additionally includes reduced item losses across tests. Therefore, hypermnesia embodies also a variant of the testing effect by means of their common finding that repeated recall (or retrieval) makes practiced items more accessible on subsequent tests and reduces the forgetting of these items (e. g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a).

Basically, the two paradigms agree in their findings that selective retrieval of some of the previously studied items can lead to further recall of the other, not repeated items. But the processes of selective retrieval in the two paradigms are different. Whereas in experiments on selective memory retrieval, the subset that should be selectively retrieved is provided by the experimenter, in experiments on hypermnesia, subjects are asked to repeatedly retrieve all of the studied items. Accordingly, selective retrieval in hypermnesia is rather unintentionally selective. This difference in selective retrieval and the fact that reactivation of the study context may be already complete at the end of the first recall test may account for why the present findings of chapter 2 do not show an increase in item gains after impaired study context access by dint of a prolonged delay, whereas impaired study context access is a mandatory requirement for the beneficial effect of selective retrieval (see Experiment 1).

The results of Experiments 4, 5, and 7, which address the role of delay between study and test for hypermnesia, complement prior work by Mulligan

(2006) who, also employing free recall tests, investigated the role of inter-test delay for hypermnesia. Using free recall, many previous studies already showed hypermnesia after very long inter-test delays, like days, weeks, months, or even a year (e. g., Campbell, Nadel, Duke, & Ryan, 2011; Erdelyi, 1996), but Mulligan was the first to explicitly compare hypermnesia under different inter-test delay conditions. In this study, hypermnesia was found to increase with inter-test delay (7 min vs. immediate recall), with the increase being due to increases in item gains and hardly to reductions in item losses. Together with the present results, these findings suggest that, with free recall testing, both delay between study and test and inter-test delay can influence hypermnesia, though in different ways. Whereas increased delay between study and test seems to affect mainly item losses (present experiments), increased inter-test delay seems to affect mainly item gains (Mulligan, 2006). Future work may investigate the possible interaction between the two delay factors and examine whether the present free recall results generalize to conditions in which delay between tests is increased, and whether the results by Mulligan generalize to conditions in which delay between study and test is increased.

Free Recall versus Forced Recall Testing

Erdelyi and Becker (1974) introduced forced recall testing in the hypermnesia literature in order to control for possible criterion changes across successive recall tests. Comparing the effects of forced recall and free recall on hypermnesia, the results of several studies, however, reported equivalent hypermnesia effects, first of all indicating that changes in response criteria may play little, if any, role in hypermnesia (e. g., Roediger & Payne, 1985). While this prior work focused on short delay conditions between study and test, the present study includes both short and long delay conditions. Doing so, the results of the present Experiments 6 and 7 show nonequivalent hypermnesia effects for the two recall formats after prolonged delay, suggesting an effect of recall format on response criterion. However, it is unlikely that differences in response criteria mediated the difference in hypermnesia results. Indeed, if the nonequivalence between free and forced recall was caused by loosened criterion with free recall testing, then the increase in hypermnesia with delay observed with free recall testing should have been accompanied by an increase in item gains rather than a reduction in item losses, which is not what the present results show.

The results rather suggest that different mechanisms may mediate hypermnesia after delay with forced versus free recall testing. Whereas the increase in hypermnesia with free recall testing is in line with the retrieval practice hypothesis but is inconsistent with the cumulative recall hypothesis (see above), the opposite is true for the decrease in hypermnesia with forced recall testing. Indeed, because (cumulative) recall levels after longer delay are lower than after short delay (see Table 6), the finding of decreased hypermnesia after delay agrees with the cumulative recall hypothesis, which assumes that study conditions producing low levels of asymptotic recall should induce less hypermnesia than conditions producing higher levels of recall. The observed numerical (though not statistical) reduction in item gains with delay fits also with this view. Answering the questions of why different mechanisms may mediate hypermnesia with forced recall than with free recall testing, and why repeated testing after delay reduces item losses with free recall only, is beyond the scope of the present study and future work is required to address these issues. Such work may improve not only our understanding of hypermnesia but also of the relation between free and forced recall testing in general.

Conclusions

To conclude, the four experiments in chapter 2 focused on beneficial effects of repeated recall across successive recall attempts employing the hypermnesia paradigm. In particular, the results showed that hypermnesia varied with the delay between study and test. When free recall was used at test, hypermnesia increased with delay and the effect was driven mainly by reduced item losses between tests. When forced recall was used at test, hypermnesia decreased with delay and was even absent after longer delay. These findings indicate that recall format can influence hypermnesia and different mechanisms may mediate the effects of repeated testing with free and forced recall testing.

Chapter 3

General Discussion

3.1 Summary of Findings

In the first chapter of this thesis, the evidence on the beneficial effect of selective memory retrieval of some studied items on related items is collected by investigating retrieval dynamics after a prolonged interval between study and test. All experiments listed in chapter 1 support the findings of recent studies, which consistently show that the selective retrieval of some studied items from a list improves the recall of the list's remaining items after a prolonged delay between study and retrieval (e. g., Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014). Consistently, the present Experiments 2-3 replicate Bäuml & Dobler's (2015) study by showing that the beneficial effect of selective retrieval is not retrieval specific. Thus, it is not restricted to selective retrieval trials (as the detrimental effect), but rather generalizes to selective restudy trials. Moreover, they firstly reveal that the repetition format can influence the size of the beneficial effect on nonrepeated items with the beneficial effect being stronger after retrieval than after restudy, and again stronger after difficult than after easy retrieval. Most importantly, the results of Experiment 1 firstly show directly that the beneficial effect of selective retrieval critically depends on an impaired access to the study context during retrieval by demonstrating a beneficial effect of selective retrieval after a prolonged retention interval when there is no mental reinstatement of the study context before recall starts, but a detrimental effect when study context is mentally reinstated before recall starts.

In the second chapter of this thesis, evidence for the beneficial effect of repeated recall in hypermnesia is collected by investigating the role of delay between study and test when either free or forced recall tests are employed. By showing reliable item gains in all experiments of chapter 2, the results confirm the findings of chapter 1, namely that the selective retrieval of some of the previously studied items can lead to a further recall of the other nonrepeated items. In addition, they conform to the findings that the repeated retrieval of some previously studied items strengthens these retrieved items by reducing item losses in later recall tests. Thus, the results replicate prior findings by showing reliable hypermnesia after a short delay across repeated free recall tests, as well as across repeated forced recall tests (e. g., Erdelyi, 1996; Payne, 1987). Notably, the results firstly indicate different effects of a prolongation of the delay between study and test for varying test formats on the amount of hypermnesia. Employing free recall tests, hypermnesia increases with delay. This effect is mainly driven by reduced item losses between tests, whereas item gains are less influenced by delay. Employing forced recall tests, however, hypermnesia decreases with delay and is even absent after longer delay.

For both investigated beneficial effects of memory retrieval, the underlying mechanisms have not been clearly identified yet. Indeed, to date there has only been rather indirect empirical evidence for Bäuml and Samenieh's (2012) assumption that retrieval-induced context reactivation processes underlie the beneficial effect of selective memory retrieval. Hence, Experiment 1 makes a protruding contribution to this branch of memory research by firstly providing direct evidence for this proposal, consequently underlining the critical role of study context access for the beneficial effect of selective retrieval. Based on these results, another still open-ended question has been whether the format of selective item repetition can influence such repetition-induced context reactivation processes, and thus impact the beneficial effects of selective item repetition. All in all, across three Experiments and varying study material, recall performance and, given the findings of Experiment 1, consequently also the amount of context reactivation was modulated by the retrieval format. Especially the finding that easy and difficult forms of selective retrieval differ in the degree of the inducement of the context reactivation is new to the literature and allows to create a link between these results and the assumptions of the episodic-context account on the testing effect by Karpicke et al. (2014), who assume that the difficulty of nonselective retrieval can influence the context reactivation.

Considering the literature on hypermnesia, a variety of different accounts have emerged, but none of them can account for the whole range of experimental results. One essential and so far not clearly acknowledged question has been the role of the interval between study and test, which is an important factor contributing to the different accounts. By showing different effects of the prolongation of the delay between study and test on hypermnesia depending on the test format, the present Experiments 4-7 suggest that different mechanisms mediate the effects of repeated testing with free and forced recall. While the finding that hypermnesia across free recall tests increases after a prolonged delay favors the retrieval practice explanation of hypermnesia, the finding that hypermnesia across forced recall tests, in contrast, decreases after a prolonged delay is rather consistent with the changes in cue set hypothesis.

To sum up, the results presented in the present eight experiments emphasize the potency of retrieval for our episodic memory by representing two forms of retrieval-induced beneficial effects on recall. Intriguingly, they pose a more detailed perspective on the retrieval's potential for effective remembering by shedding further light on the mechanisms that underlie the beneficial effects of selective and nonselective retrieval for our episodic memories.

3.2 Theoretical Implications

Initially, there are important implications for Bäuml and Samenieh's (2012) two-factor account of selective memory retrieval. They assume that whether selective retrieval is detrimental or beneficial for related material depends on the accessibility to the original study context. According to them, selective retrieval reduces the recall of related material via inhibition or blocking when access to the original study context is (largely) maintained and it enhances the recall of related material via context reactivation when access to the original study context is impaired. The present chapter 1 supports this account by showing reliable beneficial effects when access to the original study context is impaired after a long retention interval between study and retrieval, but a detrimental effect when there is a context reinstatement before the retrieval starts. Prior results have been consistent with a basic form of the two-factor account, which assumes that one type of process is active when the access to the study context is maintained and the other type of process is active when the access to the study context is impaired. However, the present results of Experiment 1 suggest a more realistic form of the two-factor account. They indicate that both types of processes are active under both conditions with one type of process predominating one condition and the other type of process predominating the other condition. Indeed, Experiment 1 shows both types of processes in identical procedures with the one exception that under one condition, the access to the study context is reinstated before the retrieval starts and under the other it is not. In doing so, it additionally serves as the first direct evidence for context reactivation processes underlying the beneficial effect of selective memory retrieval. Above all, the findings of Experiments 2-3 broaden the two-factor account by showing that, unlike the detrimental effect, the beneficial effect is not retrieval specific but is rather modulated by the retrieval format.

The idea that retrieval-induced context reactivation processes enhance the final recall performance has also been successfully applied to explain the recency effect and the contiguity effect. The principle of contiguity (Kahana, 1996) refers to the finding that during the free recall of previously studied item lists, the items that have been studied in a serial neighbored position tend to have a neighboring output position. The recency effect describes the decline in memory performance with the passage of time or the presence of intervening events (Murdock, 1962). On the basis of the assumption that contexts of serial nearby presented items overlap and that this overlap increases with a decreasing lag between those items, findings regarding these two effects support the assumption that the retrieval of prior contextual states leads to an advantage to forward recall (Kahana & Howard, 2005).

Yet, the context retrieval theory suggests that not only repetition by virtue of retrieval but also repetition by virtue of restudy can induce context reactivation processes (Greene, 1989; Thios & D'Agostino, 1976). Evidence is provided by the spacing effect (Bjork, 1970), which is based on the early findings of Ebbinghaus (1964) regarding a memory advantage for spaced compared to massed presented items, i. e., it is easier to remember items when they are studied repeatedly and thus spaced over a long time, rather than studied many a time in a short time span. One of the first accounts for the spacing effect was the contextual variability theory (Melton, 1967). Fundamentally, it is assumed that spaced items occur in multiple contexts. Those multiple contexts may lead to different retrieval routes by which they can be accessed easier than massed items do in the following test . Thus, findings on the spacing effect are in line with the present findings that the beneficial effect of selective retrieval and hence also context reactivation processes are not retrieval specific but arise in response to both, retrieval and restudy trials.

The spacing effect is also included in the desirable difficulties perspective (e. g., Bjork, 1994; Bjork& Bjork, 2011), which suggests that making learning feel more difficult and challenging enhances long-term retention and transfer. Contiguous to the proposal that learning should be spaced, Bjork and his colleagues have specified three further ways in which learning should be made difficult. Based on a review of the relevant literature (e. g., Carrier & Pashler, 1992; Smith, Glenberg, & Bjork, 1978; Shea & Morgan, 1979), they additionally have recommended that conditions of learning should be varied, feedback should be reduced and study of different topics should be interleaved. While their concept of desirable difficulties is restricted to the difficulties at learning, the present thesis yields the idea that more difficult retrieval also benefits long-term retention. Indeed, the results of chapter 1 directly show that more difficult selective retrieval (i. e., when only the studied item's initial letters are given as retrieval cues) enhances later recall of related material more than easy retrieval (i. e., when the studied item's word stems are given as retrieval cues). On the basis of the retrieval practice hypothesis and the assumption that, in general, longer delays make retrieval more demanding, the free recall results of chapter 2 also indicate that difficult retrieval on an initial test (i. e., after a long delay) increases hypermnesia compared to easier retrieval on an initial test (i. e., after a short delay). Thus, difficulty does not only benefit the positive effects of learning, but also seems to make the retrieval of our episodic memories more effective. Nevertheless, future research is required to figure out more precisely which role the difficulty of retrieval plays for the beneficial effects of recall in hypermnesia.

Recently, the testing effect has also been attributed to context reactivation processes (Karpicke et al., 2014). Because the present results of chapter 1 on the beneficial effect of selective retrieval suggest that repetition formats can differ in the degree to which they cause context reactivation and thus differ in the degree to which they induce beneficial effects on the recall of other items, it parallels Karpicke et al.'s (2014) episodic context account. Likewise, Karpicke et al. adopt repetition formats to differ in the degree to which they cause context reactivation and thus differ in the degree to which they cause recall improvements, not for related information, but rather for the repeated information itself (for more details, see chapter 1.8).

In a slightly different way, the present free recall results of chapter 2 on hypermnesia are also in line with the testing effect literature. They yielded that, after a longer delay between study and test, the retrieval practice of some previously studied items can reduce the forgetting of practiced items on subsequent recall tests more than after a shorter delay between study and test. Assuming that the retrieval after a longer delay is more demanding than after a short delay, these results are in line with findings on the testing effect which showed that retrieval practice can improve the recall of practiced items and enhances it even more if retrieval practice is demanding (e. g., Carpenter, 2011; Halamish & Bjork, 2011; Pyc & Rawson, 2009). Thus, employing different paradigms, both prior findings on the testing effect and the present findings on hypermnesia demonstrate how the difficulty of retrieval practice influences the positive effects of retrieval. In contrast to the parallels between the beneficial effect of selective retrieval and the testing effect, that are based on similar context reactivation processes, no evidence for such processes can be provided by the present experiments on hypermnesia. Because the findings on hypermnesia did not show an increase in item gains after a long delay, context reactivation processes may be largely restricted to the first test and not easily extend to subsequent recall tests (for more details, see chapter 2.9).

There could also be a relation between the present experiments and studies on test-potentiated learning. Test-potentiated learning refers to the finding that retrieval practice on previously studied material can enhance the ability of a learner to benefit from a subsequent restudy opportunity (Arnold & McDermott, 2013a, 2013b; Izawa, 1966). If this retrieval-practice effect was also (partly) mediated by retrieval-induced context reactivation processes, then the effect may also increase if the context between study and retrieval practice is changed, and it may be larger after a more difficult than an easy retrieval practice. Potentially, even if context reactivation processes played a minor role for the test-potentiated learning, like in the present hypermnesia experiments, retrieval practice could reduce the forgetting of practiced items more after difficult retrieval trials than after easy retrieval trials and thus enhance the effect of test-potentiated learning. Future work may address the issue, providing information on whether context reactivation processes and/ or strengthen processes because of the difficulty of the retrieval trials can contribute to test-potentiated learning as well.

Finally, however, selective retrieval and selective restudy are just two options to increase the contextual overlap between study and test to improve the recall performance at test. For instance, different forms of external and internal cuing can lead to a similar reactivation of the study context. As a very famous and often-cited example of external context reactivation effects, Godden and Baddeley (1975) showed that recall performance is enhanced when there is a match between the places where encoding and test are conducted (e. g., study on dry land, test on dry land), compared to when the test is carried out in a place mismatching the encoding condition (e. g., study on dry land, test under water; for similar results, see Smith et al., 1978). In his review, Eich (1980) accented the role of internal context reactivation effects in a similar way by indicating that the recall performance in a test declines when the subjects pharmacological state changes between study and test (e. g., boozed at study, sober at test), compared to conditions in which their state remains the same (e. g., boozed at study, boozed at test; for similar results, see Eich & Metcalfe, 1989).

3.3 Further Directions

The findings on the retrieval-induced beneficial effects, particularly with regard to the present findings on the beneficial effects of selective retrieval and repeated recall in hypermnesia, may be of great interest for practical applications. For instance, such findings may play an eminent role in education. Optimizing learning strategies for students at schools or universities has always been a central issue for our educational system. For a long time, tests have served only as assessment devices to test what a student knows. Especially since many studies on the testing effect from the last decade have emphasized the power of testing for long-term retention and the implications on educational practice (e. g., McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger & Karpicke, 2006b), the fact that retrieval can change memories is more and more taken into consideration and implemented in school lessons, as well as tutorials and practical courses at university. But the still persisting question is, how retrieval as it occurs in classrooms, tutorials or during self-study can be most effective, considering that usually not the complete subject factor, but rather only a part of it is retrieved.

Moreover, in the psycho-legal research of the past 30 years, the focus has been on retrieval processes as an influencing factor for eyewitness accuracy. Considering that eyewitnesses typically report what they have seen several times as they are interrogated by the police, consultants and jurists, it is not surprising that hypermnesia turned out to be relevant to this branch of research. Indeed, the basic finding of hypermnesia for eyewitness memory has repeatedly been shown after relatively short intervals (e. g., La Rooy, Pipe & Murray, 2005; Scrivner & Safer, 1988). But if you imagine interrogations regarding a crime or an accident in real life, the first interrogations of eyewitnesses occur after varying delays. Therefore, the examination of the role of delay between study and test on hypermnesia may be of special interest.

Although selective retrieval is investigated in chapter 1 and the role of delay in hypermnesia is investigated in chapter 2, the direct application of the present results to educational and psycho-legal concerns is challenged by the transferability of the employed study material. Like most studies in cognitive psychology, the present experiments are conducted by employing lists of unrelated items, like pictures and words. Employing more coherent material, however, would offer us an opportunity to investigate effects relevant to our everyday lives, because such material reproduces the complex information that we are confronted with every day much better than lists of unrelated words or pictures do (e. g., Bäuml & Schlichting, 2014).

While the general finding of hypermnesia has already been investigated employing more coherent material like prose passages (e. g., Otani & Griffith, 1998), films (Montangero, et al., 2003), and was also directly demonstrated in studies on eyewitnesses (e. g., Dunning & Stern, 1992; Scrivner & Safer, 1988) and autobiographical memory (Bluck et al., 1999), the role of delay between study and test has hardly been investigated yet. Although the role of employed material seems to be of lesser importance after short intervals between study and test (see Roediger & Wheeler, 1992), there might be some differences caused by the material after longer delays. Because of the coherence of the single units that have to be remembered, more integrated prose material may be less likely to be forgotten. Hence, the first recall test may not be more demanding after a long delay than after a short delay and the prolongation of the interval between study and test may not influence hypermnesia across successive tests. Alternatively, it may parallel findings on selective memory retrieval and similar findings on the testing effect, which showed that the retrieval-induced facilitation is increased after a long delay, compared to a short delay for both unrelated word lists and more complex material (e. g., Bäuml & Schlichting, 2014; Roediger & Karpicke, 2006a). Because up to now, there is no clear answer on this issue, it might be interesting to examine the role of delay for hypermnesia with more complex material in future studies.

Whereas there have already been studies on hypermnesia employing more complex study material, the beneficial effect of selective retrieval has been examined for only a relatively restricted set of study materials. Indeed, most studies employed lists of unrelated words (for exception see Bäuml & Schlichting, 2014). Although the present Experiment 3 shows the beneficial effect of selective retrieval for more coherent prose material and thus points out that the beneficial effect does not largely depend on the study material, additional work is needed to demonstrate the beneficial effect of selective retrieval within the scope of a wider range of study materials. Particularly interesting is, whether the effect arises for autobiographical and eyewitness memories. Since the present Experiment 3 extended prior studies by showing that the beneficial effect can also be modulated by retrieval difficulty, when coherent prose material is used, there may be a greater beneficial effect, if the study material is more complex such as autobiographical or eyewitness memories.

Educational and psycho-legal research areas serve as examples which reflect that various areas of application may benefit from the re-investigation of the present results, employing more complex material. Therefore, future work may enable to learn how the retrieval of more complex study material can be most effective for later recall and notably, whether it is conducive for later recall to extent the delay between encoding and retrieval or to make retrieval more difficult, as it is indicated by the present results.

3.4 Final Conclusions

Depending on the contextual overlap between study and retrieval, two different processes operate and cause the two opposing effects of selective retrieval on related material. While the detrimental effect is supposed to be mediated by inhibition and blocking processes (e. g., Storm & Levy, 2012), the present thesis provides the first direct evidence that the beneficial effect is mediated by context reactivation processes. Moreover, the detrimental effect of selective retrieval has repeatedly been shown to be largely retrieval specific with selective retrieval but not restudy impairing recall of related items (e. g., Bäuml & Dobler, 2015). The present thesis shows that the beneficial effect, however, is not retrieval specific but rather generalizes to restudy trials. Furthermore, the beneficial effect is actually modulated by retrieval format with more difficult repetition formats leading to stronger beneficial effects than easier repetition formats do. Thus, by showing that the repetition format can influence context reactivation processes, a more conclusive explanation of the role of context for the effects of selective item repetition is offered.

By demonstrating that with free recall testing, recall performance across repeated tests increases with delay and that the effect is mainly driven by reduced item losses between tests, evidence for the retrieval practice hypothesis of hypermnesia is provided. In contrast, showing that with forced recall testing, hypermnesia decreases with delay argues for the cumulative recall hypothesis. Thus, the present thesis firstly indicates that the recall format can influence hypermnesia after a long delay and that different mechanisms may mediate the effect with free versus forced recall testing.

Moreover, the present results on both effects yield a link to other eminent memory phenomena. In particular, there are interesting parallels to findings on the testing effect. Like previous findings on the testing effect, the present findings on both paradigms are (albeit in a fairly different manner) also a demonstration of the powerful role of retrieval difficulty for later recall. In fact, the present work delivers insight into the more detailed study of these different, robust and meaningful memory phenomena and beyond that, opens the window for practical applications to educational and psycho-legal settings.

Literature

Abbott, E. E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs, 11*, 159-177.

Abel, M., & Bäuml, K.-H. T. (2015). Selective memory retrieval in social groups: When silence is golden and when it is not. *Cognition, 140*, 40-48.

Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior, 8,* 463-470.

Ammons, H., & Irion, A. L. (1954). A note on the Ballard reminiscence phenomenon. *Journal of Experimental Psychology, 48*, 184-186.

Anderson, M. C. (2003). Rethinking interference theory: executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49,* 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*, 1063-1087.

Arnold, K. M., & McDermott, K. B. (2013a). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal* *of Experimental Psychology: Learning, Memory and Cognition, 39*, 940-945.

Arnold, K. M., & McDermott, K. B. (2013b). Free recall enhances subsequent learning. *Psychonomic Bulletin and Review, 20*, 507-513.

Aslan, A., & Bäuml, K.-H. T. (2014). Later maturation of the beneficial than the detrimental effect of selective memory retrieval. *Psychological Science, 25*, 1025-1030.

Aslan, A., Schlichting, A., John, T., & Bäuml, K.-H. T. (2015). The two faces of selective memory retrieval: Earlier decline of the beneficial than the detrimental effect with older age. *Psychology and Aging, 30*, 824-834.

Baddeley, A. (2001). The concept of episodic memory. *Philosophical Transactions of the Royal Society of London. Series B, 356*, 1345-1350.

Bäuml, K.-H. T. (2002). Semantic generation can cause episodic forgetting. *Psychological Science, 13*, 356-360.

Bäuml, K.-H. T., Aslan, A., & Abel, M. (2017). The two faces of selective memory retrieval - cognitive, developmental, and social processes. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 66, pp. 167-209). Academic Press: Elsevier Inc.

Bäuml, K.-H. T., & Dobler, I. M. (2015). The two faces of selective memory retrieval: recall specificity of the detrimental but not the beneficial effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41*, 246-253.

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*, 1568-1581.

Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language, 68*, 39-53.

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

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

Bäuml, K.-H. T., & Samenieh, A. (2012). Infuences of part-list cuing on different forms of episodic forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 366-375.

Bäuml, K.-H. T., & Schlichting, A. (2014). Memory retrieval as a self-propagating process. *Cognition, 132*, 16-21.

Ballard, P. B. (1913). Obliviscence and Reminiscence. *British Journal of Psychology Monograph Supplements, 1,* 1-82.

Belmore, S. M. (1981). Imagery and semantic elaboration in hypermnesia for words. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 191-203.

Bergstein, J., & Erdelyi, M. (2008). Recognition hypermnesia: How to get it. *Memory, 16*, 689-702.

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

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217-235). Washington, DC: Winston & Sons.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *The American Journal of Psychology, 128*, 241-252.

Bluck, S., Levine, L. J., & Laulhere, T. M. (1999). Autobiographical remembering and hypermnesia: a comparison of older and younger adults. *Psychology and Aging, 14*, 671-682.

Bornstein, B. H., Liebel, L. M., & Scarberry, N. C. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology, 12*, 119-131.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. *Coding Processes in Human Memory,* 85-123.

Bunch, M. E. (1938). The measurement of reminiscence. *Psychological Review, 4*5, 525-531.

Buxton, C. E. (1943). The status of research in reminiscence. *Psychological Bulletin, 40,* 313-340.

Campbell, J., Nadel, L., Duke, D., & Ryan, L. (2011). Remembering all that and then some: Recollection of autobiographical memories after a 1-year delay. *Memory, 19*, 406-415.

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*, 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*, 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 and Cognition, 34*, 268-276.

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

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

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

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1403-1414.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Erlbaum.

Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable through conversation Socially shared retrieval-induced forgetting of September 11 memories. *Psychological Science, 20*, 627-633.

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,* 535-551.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Cuc, A., Koppel, J., & Hirst, W. (2007). Silence is not golden a case for socially shared retrieval-induced forgetting. *Psychological Science, 18*, 727-733.

Delaney, P. F., Sahakyan, L., Kelley, C. M., & Zimmerman, C. A. (2010). Remembering to forget: The amnesic effect of daydreaming. *Psychological Science, 21*, 1036-1042.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation, 53,* 63-147.

Dobler, I. M., & Bäuml, K.-H. T. (2012). Dissociating the two faces of selective memory retrieval. *Memory, 20,* 478-486.

Dunning, D., & Stern, L. B. (1992). Examining the generality of eyewitness hypermnesia: A close look at time delay and question type. *Applied Cognitive Psychology, 6,* 643-657.

Ebbinghaus, H. (1964). *ber das Gedlichtnis.* Leipzig: Duncker and Humblot, 1885.

Eich, J. E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory and Cognition, 8*, 157-173.

Eich, J. E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. Journal of Experimental *Psychology: Learning, Memory, and Cognition, 15,* 443-455.

Erdelyi, M. H. (1996). *The recovery of unconscious memories: Hypermnesia and reminiscence.* Chicago: The University of Chicago Press.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology, 6*, 159-171.

Erdelyi, M., Buschke, H., & Finkelstein, S. (1977). Hypermnesia for Socratic stimuli: The growth of recall for an internally generated memory list abstracted from a series of riddles. *Memory and Cognition, 5,* 283-286.

Erdelyi, M. H., & Kleinbard, J. (1978). Has Ebbinghaus decayed with time? The growth of recall (hypermnesia) over days. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 275-289.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62*, 145-154.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A exible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 60*, Vol. 40.

Geiselman, R. E., Bjork, R. A., & Fishman, D. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General, 112*, 58-72.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology, 70,* 401-412.

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

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

Goernert, P. N., Widner, R. L., & Otani, H. (2007). Classification accuracy across multiple tests following item method directed forgetting. *Quarterly Journal of Experimental Psychology, 60,* 1178-1186.

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 371-377.

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*, 801-812.

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,* 562-567.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 923-941.

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

Hulbert, J. C., Shivde, G., & Anderson, M. C. (2012). Evidence against associative blocking as a cause of cue-independent retrieval-induced forgetting. *Experimental Psychology, 59*, 11-21.

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, 18*, 879-919.

Jonker, T. R., Seli, P., & MacLeod, C. (2013). Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review, 4*, 852-872.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition, 24*, 103-109.

Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin and Review, 12,* 159-164.

Kang, S. H. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory and Cognition, 38*, 1009-1017.

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

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

Kelley, M. R., & Nairne, J. S. (2003). Remembering the forgotten? Reminiscence, hypermnesia and memory for order. *The Quarterly Journal of Experimental Psychology, 56*, 577-599.

La Rooy, D., Pipe, M. E., & Murray, J. E. (2005). Reminiscence and hypermnesia in childrens eyewitness memory. Journal of Experimental Child Psychology, 90, 235-254.

Madigan, S. (1976). Reminiscence and item recovery in free recall. *Memory and Cognition, 4*, 233-236.

Madigan, S., & Lawrence, V. (1980). Factors affecting item recovery and hypermnesia in free recall. *The American Journal of Psychology, 93*, 489-504.

Masson, M. E., & McDaniel, M. A. (1981). The role of organizational processes in longterm retention. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 100-110.

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*, 494-513.

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 371-385.

McDaniel, M. A., Moore, B. A., & Whiteman, H. L. (1998). Dynamic changes in hypermnesia across early and late tests: A relational/item-specific account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 173-185.

McGeoch, G. O. (1935). The conditions of reminiscence. *The American Journal of Psychology, 47,* 65-89.

Melton, A. W. (1967). Repetition and retrieval from memory. *Science, 158,* 532-532.

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

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

Montangero, J., Ivanyi, C. T., & de Saint-Hilaire, Z. (2003). Completeness and accuracy of morning reports after a recall cue: Comparison of dream and film reports. *Consciousness and Cognition, 12,* 49-62.

Mulligan, N. W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 436-450.

Mulligan, N. W. (2002). The emergent generation effect and hypermnesia: Influences of semantic and nonsemantic generation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 541-554.

Mulligan, N. W. (2005). Total retrieval time and hypermnesia: Investigating the benefits of multiple recall tests. *Psychological Research, 69*, 272-284.

Mulligan, N. W. (2006). Hypermnesia and total retrieval time. *Memory, 14*, 502-518.

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin, 140*, 1383-1400.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*, 482-488.

Nelson, T. O., & MacLeod, C. M. (1974). Fluctuations in recall across successive test trials. *Memory and Cognition, 2,* 687-690.

Otani, H., & Griffith, J. D. (1998). Hypermnesia for prose. *The Journal of General Psychology, 125*, 147-155.

Otani, H., Widner, R. L., Whiteman, H. L., & Louis, J. P. S. (1999). Hypermnesia: The role of multiple retrieval cues. *Memory and Cognition, 27*, 928-934.

Pastötter B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology, 5,* 1-10.

Payne, D. G. (1986). Hypermnesia for pictures and words: Testing the recall level hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 16-29.

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101*, 5-27.

Payne, D. G., & Roediger , H. L., III. (1987). Hypermnesia occurs in recall but not in recognition. *The American Journal of Psychology,* 100, 145-165.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*, 129-156.

Pyc, M. A., Balota, D. A., McDermott, K. B., Tully, T., & Roediger, H. L., III, (2014). Between-list lag effects in recall depend on retention interval. *Memory and Cognition, 42*, 965-977.

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*, 437-447.

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

Raaijmakers, J. G. W., & Jakab, E. (2012). Retrieval-induced forgetting without competition: Testing the retrieval specificity assumption of inhibition theory. *Memory and Cognition, 40*, 19-27.

Raaijmakers, J. G., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In B. Ross (Ed.), *The Psychology of Learning and Motivation (Vol.14, pp. 207-262).* Academic Press: Elsevier Inc.

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

Roediger, H. L., III. (1974). Inhibiting effects of recall. *Memory and Cognition, 2*, 261-262.

Roediger, H. L., III. (1978). Recall as a self-limiting process. *Memory & Cognition, 6*, 54-63.

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

Roediger, H. L., III., & Challis, B. H. (1989). Hypermnesia: Improvements in recall with repeated testing. In C. Izawa (Ed.), *Current Issues in Cognitive Processes: The Tulane Floweree Symposium on Cognition.* Hillside, NJ: Erlbaum.

Roediger, H. L., III. & Karpicke, J. D. (2006a). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255.

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

Roediger, H. L., III., & Neely, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology, 36*, 213-242.

Roediger, H. L., III., & Payne, D. G. (1982). Hypermnesia: the role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 66-72.

Roediger, H. L., III., & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory and Cognition, 13*, 1-7.

Roediger, H. L., III., Payne, D. G., Gillespie, G. L., & Lean, D. S. (1982). Hypermnesia as determined by level of recall. *Journal of Verbal Learning and Verbal Behavior, 21,* 635-655.

Roediger, H. L., III., & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory and Cognition, 6*, 296-305.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140,* 1432-1463.

Rowland, C. A., & DeLosh, E. L. (2014). Benefits of testing for nontested information: retrieval-induced facilitation of episodically bound material. *Psychonomic Bulletin and Review, 21,* 1516-1523.

Rupprecht, J., & Bäuml, K.-H. T. (2016). Retrieval-induced forgetting in item recognition: Retrieval specificity revisited. *Journal of Memory and Language, 86*, 97-118.

Rupprecht, J., & Bäuml, K.-H. T. (2017). Retrieval-induced versus context-induced forgetting: Can restudy preceded by context change simulate retrieval-induced forgetting? *Journal of Memory and Language, 93*, 259-275.

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

Schlichting, A., Aslan, A., Holterman, C., & Bäuml, K.-H. T. (2015). Working memory capacity predicts the beneficial effect of selective memory retrieval. *Memory, 23*, 786-794.

Scrivner, E., & Safer, M. A. (1988). Eyewitnesses show hypermnesia for details about a violent event. Journal of Applied Psychology, 73, 371-377.

Shapiro, S. R., & Erdelyi, M. H. (1974). Hypermnesia for pictures but not words. *Journal of Experimental Psychology, 103*, 1218-1219.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. Journal of Experimental Psychology: Human Learning and Memory, 5, 179-187.

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*, 384-197.

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

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Enviromental context and human memory. *Memory and Cognition, 6*, 342-353.

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*, 1582-1593.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174-214.

Stone, C. B., Barnier, A. J., Sutton, J., & Hirst, W. (2013). Forgetting our personal past: Socially shared retrieval-induced forgetting of autobiographical memories. *Journal of Experimental Psychology: General, 142*, 1084-1094.

Storm, B. C., & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory and Cognition, 40*, 827-843.

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

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86*, 255-262.

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior, 15,* 529-536.

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

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*, 175-184.

Tulving, E. (1972). Episodic and semantic memory. *Organization of memory, 1,* 381-403.

Tulving, E. (1985). Memory and consciousness. Canadian Psychology, 26, 1-12.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*, 1-25.

Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and verbal behavior, 1*, 321-334.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 353-373.

Verde, M. F. (2013). Retrieval-induced forgetting in recall: Competitor interference revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1433-1448.

Walker, M. P., Stickgold, R., Alsop, D., Gaab, N., Schlaug, G. (2005). Sleep-dependent motor memory plasticity in the human brain. *Neuroscience, 133,* 911-917.

Wheeler, M. A., & Roediger, H. L., III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3*, 240-245.

Wilkinson, A. C., & Koestler, R. (1983). Repeated recall: A new model and tests of its generality from childhood to old age. *Journal of Experimental Psychology: General, 112*, 423-451.

Zaromb, F. M., & Roediger, H. L., III. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory and Cognition, 38,* 995-1008.

Appendix

APPENDIX A

English translations of words presented to subjects in Exp. 1, 2A, 2B. Target items and target cues are depicted in bold letters.

List A	Cue List A *prior easy retrieval*	Cue List A *prior difficult retrieval*	List B	Cue List B *prior easy retrieval*	Cue List B *prior difficult retrieval*
garden	**g_____**	**g_____**	**beaker**	**b_____**	**b_____**
saloon	**s_____**	**s_____**	**rose**	**r_____**	**r_____**
pipe	**p_____**	**p_____**	**varnish**	**v_____**	**v_____**
nail	**n_____**	**n_____**	**seat**	**s_____**	**s_____**
wool	**w_____**	**w_____**	**factory**	**f_____**	**f_____**
oven	ov____	o_____	pea	pe____	p_____
knife	kn____	k_____	island	isl____	i_____
antenna	ant____	a_____	ladder	lad____	l_____
jacket	ja_____	j_____	gatherer	gä____	g_____
beekeeper	be_____	b_____	urne	ur____	u_____
rope	ro_____	r_____	traffic	tra____	t______
vinegar	vin____	v_____	writer	wr____	w_____
docket	doc____	d_____	moon	mo____	m_____
herring	her____	h_____	hotel	ho_____	h_____
loupe	lo_____	l______	curtain	cu_____	c_____

APPENDIX B

English translations of target and nontarget sentences presented to subjects in Exp. 3. Solutions for the missing words are depicted in bold letters.

Targets	• Following the oscillating closed model, the universe will be permanently arranged between a big bang and a big _______ **(*crunch*).** • After some period of time following the big bang, gravity condensed clumps of matter together which eventually formed _______ *(**galaxies**)*. • Due to the Doppler shifting, the wavelength emitted by something moving away from us is shifted to a _____ frequency (***lower***). • NASA built the satellite _____ to detect background radiation (acronym suffices, ***COBE; Cosmic Background Explorer***). • Astronomers using Hubble have found the element _____ in extremely ancient stars (**boron**). • Arthur Eddington said: "We must allow _____ an infinite amount of time to get started" (***evolution***).
Nontargets prior easy retrieval	• In 1964, two astronomers, Arno Penzias and Robert Wilson, inadvertently discovered a noise that they thought belonged to an ext_____ origin (***extraterrestrial***). • It has taken every galaxy the same amount of ti____ to move from a common starting position to its current position (***time***). • At the beginning of the universe, there was an asymmetry between two kinds of particles. As these two materials collided and destroyed one another, they created pure ene_____ (***energy***). • Einstein resisted the idea of a beginning of the universe by introducing a constant into his equations. It is named fu____ factor (***fudge***). • Later, it became obvious that what Arno Penzias and Robert Wilson heard was co_____ background radiation (***cosmic***). • The Hubble Constant refers to how fast the velocities of the galaxies increase with their dis_____ from the Earth (***distance***). • Immediately after the big bang, the universe was tremendously hot as a result of particles of both matter and anti _____ rushing apart in all directions (***antimatter***). • The theory that argues that the Universe began with a period of exaggerated outward expansion is called Inf_____ theory (***Inflation***). • Visible wavelengths emitted by objects moving away from us are shifted towards the re_____ part of the visible spectrum (red). • The Big Bang theory states that in order to have mass condense and form galaxies, there must be inh_____ left over from the big bang that will be detectable (***inhomogeneities***). • Using Hubble, a certain element has been found in extremely ancient stars. According to the present theory, such a heavy and com_____ atom could not have existed (***complex***). • When protons and neutrons began to react with each other, they form deuterium, an isotope that is also called heavy hyd_____ (***hydrogen***).

Note. In the prior easy retrieval condition, randomly selected 8 of the12 nontarget sentences were presented to subjects. In the prior difficult retrieval condition, all 12 nontarget sentences were presented to subjects prior to target retrieval. Like for the target sentences, no cues for the missing words were given in the prior difficult retrieval condition.

APPENDIX C

English translations of the words presented to subjects in Experiment 4 and also original labels of line-drawing pictures presented in to subjects in Experiment 5.

Set 1

parrot	bat	pear	lamp	radio	mirror
whale	bird	plank	mixer	girl	carpet
slug	bug	letter	strawberry	rain	bowl
baby	bride	violin	rainbow	slide	seesaw
camel	dentist	cheese	nail	shovel	tent
zebra	boy	bone	pizza	ship	sheet
bread	elephant	helmet	cup	bag	queen
pirate	broom	tie	puzzle	pants	moose

Set 2

cowboy	panda	brush	microscope	plate	bomb
crab	octopus	egg	nest	stairs	bridge
fish	penguin	feather	patch	clock	chain
ghost	policeman	glass	rocket	colcannon	cross
king	chest	towel	rose	balloon	cactus
lizard	desert	comb	box	bathtub	butter
llama	lion	tape	rope	belt	doctor
monkey	drill	ladder	scarf	bench	fireman

APPENDIX D

Original labels of line-drawing pictures additionally presented to participants in Experiments 6, 7.

Set 1

plane	book	apple	bus	flower	dog
ant	axe	cake	mushroom	ear	moon

Set 2

carrot	pipe	ruler	ring	banana	scissors
key	refrigerator	toaster	flag	chair	cat