

**When Testing Improves Memory Performance:
The Forward Testing Effect With Prose Materials**

Inaugural-Dissertation zur Erlangung der Doktorwürde der Fakultät
für Humanwissenschaften der Universität Regensburg

vorgelegt von

Verena Margaretha Kriechbaum

aus Antwort

Regensburg 2026

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

Zweitgutachter: Prof. Dr. Peter Fischer

Acknowledgements

My special thanks go to my supervisor, Prof. Dr. Karl-Heinz Bäuml, who is likely the best out there in the world. I am absolutely grateful that I was able to do my PhD with him. Heinz, you were a great support and always there when I needed help.

I would also like to thank my colleagues at the lab: Petra Witzmann, Lena Abel, Oliver Kliegl, Lisa Wallner, Michael Wirth, Anna Nickl, Johannes Bartl, Sarah Meixensperger, Marilena Hirsch, and Zsuzsanna Nemecz. You always had an open ear during the years, and you all contributed to the joy I felt at work.

Lukas, I cannot thank you enough for all your support. For some years now, we are dancing through life together. May there be many more years to come.

Big thanks also need to go to my parents, Franz and Barbara, who were always there for me. This dissertation is dedicated to them with love.

Last but not least, I am so grateful for this journey. There were so many ups and downs, and I yet still made it through. This work really has my heart.

Parts of the present thesis have been published as:

Kriechbaum, V. M., & Bäuml, K.-H. T. (2024). Retrieval practice can promote new learning with both related and unrelated prose materials. *Journal of Applied Research in Memory and Cognition*, 13(3), 319–328.

Kriechbaum, V. M., & Bäuml, K.-H. T. (2025). Retrieval practice versus elaborative studying with concept mapping – Both promote new learning with related prose materials. *Journal of Applied Research in Memory and Cognition*. Advance online publication.

Contents

Abstract	8
Preface.....	9
1. The Forward Testing Effect: Empirical and Theoretical Overview.....	17
1.1 The Forward Testing Effect With More Complex Study Material.....	20
1.2 Theoretical Accounts of the Forward Testing Effect and Empirical Support.....	23
1.3 The Forward Testing Effect With Complex Material: Empirical Findings and Theoretical Implications	27
1.4 The Potential Role of Relatedness of Study Materials	31
2. Open Questions and Goals of the Present Dissertation.....	33
2.1 Open Questions.....	33
2.2 Goals of the Present Dissertation	35
3. Examining the Potential Role of Context Change in the Forward Testing Effect With Prose Materials.....	39
3.1 Overview of Experiments 1–3	40
3.2 Experiment 1	41
3.2.1 Method	42
3.2.2 Results	46
3.2.3 Discussion	49
3.3 Experiment 2.....	50
3.3.1 Method	51

3.3.2 Results	52
3.3.3 Discussion	55
3.4 Experiment 3	56
3.4.1 Method	57
3.4.2 Results	58
3.4.3 Discussion	60
3.5 Discussion of Experiments 1–3	60
 4. Examining the Potential Role of Facilitated Comprehension in the Forward Testing Effect With Prose Materials	 63
4.1 The Potentiating Effects of Knowledge on New Learning	64
4.2 Retrieval Practice and Elaborative Studying With Concept Mapping as Knowledge-Acquisition Tools.....	66
4.3 The Potential Role of Relatedness of Study Materials	68
4.4 Overview of Experiments 4–6	69
4.5 Experiment 4	71
4.5.1 Method	72
4.5.2 Results	77
4.5.3 Discussion	80
4.6 Experiment 5	81
4.6.1 Method	81
4.6.2 Results	84
4.6.3 Discussion	87
4.7 Experiment 6	88
4.7.1 Method	88
4.7.2 Results	89
4.7.3 Discussion	91
4.8 Discussion of Experiments 4–6	92

5. General Discussion.....	96
5.1 Implications for Theoretical Accounts of the Forward Testing Effect	97
5.1.1 A Two-Factor Account of the Forward Testing Effect	
With Prose Materials	97
5.1.2 Further Implications for Accounts of the Forward Testing Effect	
With Prose Materials	99
5.1.3 The Potential Role of Release From Proactive Interference	
for the Forward Testing Effect With Prose Materials	101
5.2 Relation to the Memory Literature.....	103
5.2.1 Relation to Prior Studies on the Forward Testing Effect	
With More Complex Study Material.....	103
5.2.2 Relation to Prior Studies on the Forward Testing Effect	
With Simple Word Lists.....	105
5.2.3 Relation to the Literature on Knowledge-Enhanced Learning.....	107
5.2.4 An Additional Role of Attentional Factors	
in the Forward Testing Effect With Prose Materials?	110
5.2.5 Relation to Prior Studies on the Role of Release From Proactive	
Interference for the Forward Testing Effect.....	112
5.3 Implications for Educational Practice and Research.....	115
5.4 Possible Limitations of the Present Experiments.....	117
5.5 Directions for Future Research	119
5.5.1 The Potential Roles of Retention Interval and Lag	120
5.5.2 The Effects of Interim Retrieval Practice on Initially Studied Material	121
5.5.3 The Potential Roles of Type of Retrieval Practice and Text Coherence....	123
5.5.4 Potential Implications for Educational Practice	126
5.6 Conclusions.....	127
Literature.....	129

Abstract

The forward testing (FT) effect refers to the finding that retrieval practice on studied material can promote learning and memory of subsequently studied, new material. Most research on the FT effect has focused on the learning of relatively simple materials. This dissertation had the goal to examine whether the FT effect can be equally present for related and unrelated prose materials, and whether the FT effect is influenced by relatedness of study material. In six experiments, subjects studied either related or unrelated prose passages and were tested on the final critical passage immediately after study. In all experiments, recall of the final passage was higher when, between study of the single passages, subjects engaged in retrieval practice compared to restudy of the preceding passage, demonstrating an FT effect with both related and unrelated prose passages. Strikingly, only with unrelated passages did an FT-like effect arise when subjects between the passages generated semantic information unrelated to the studied passages (Experiments 1 and 2), and only with related passages did the FT effect lead to recall levels superior to those when only the final passage was studied (Experiments 1–6). Moreover, only with related passages was an FT-like effect observed when, instead of interim retrieval practice, elaborative restudy practice with concept mapping was applied between study of the single passages (Experiments 4–6). Together, these findings demonstrate a critical role of relatedness of study material for the FT effect with prose materials. Theoretically, the findings align with a new two-factor explanation of the FT effect, which suggests that the FT effect is mainly driven by context change with unrelated prose materials and is mainly driven by facilitated comprehension with related prose materials. The findings provide clues for students on how to optimize new learning during exam preparation both when faced with highly related and when faced with rather unrelated course materials.

Preface

Imagine you are reading a textbook chapter for an upcoming exam. After you have read the chapter one time, would you rather:

- A. Go back and restudy either the entire chapter or certain parts of the chapter.
- B. Try to recall material from the chapter (without the possibility of restudying the material).
- C. Use some other study technique.

In a survey conducted by Karpicke, Butler, and Roediger (2009), 57% of students chose to restudy after reading, 21% indicated that they would use some other study technique, and only 18% of the students indicated that they would engage in self-testing. You may well agree that rereading the material would be more effective than attempting to retrieve the material from memory. Research on this question is spectacularly clear, however: Retrieval practice is by far the superior learning activity.

Education in the twenty-first century is undergoing a substantial transformation. With the advent of online learning, flipped classrooms, and massive open online courses, students spend more and more time on learning outside the walls of classroom environments and thus are increasingly challenged to manage their study behaviors (see Carpenter, 2023). Unfortunately, when faced with a decision about how to approach a study session, students often do not choose the strategies that are optimal for learning (for a review, see Miyatsu et al., 2018). For instance, students tend to avoid the use of retrieval practice as a tool to potentiate learning (Hartwig & Dunlosky, 2012; Karpicke et al., 2009; Kirk-Johnson et al., 2019; Kornell & Bjork, 2007; Tricio et al., 2023; Tullis & Maddox, 2020), but the reasons for such underemployment are still not entirely clear. A popular proposal is that students lack metacognitive insight into the virtues of retrieval practice and rather

consider tests as tools to merely assess the current state of learning (Karpicke & Blunt, 2011; Morehead et al., 2016; Roediger & Karpicke, 2006b). More recent work provided evidence, however, that many students can distinguish fairly well between more versus less effective learning strategies, suggesting that the infrequent use of retrieval practice may be due to some other reason (Rea et al., 2022). Researchers have often claimed that retrieval practice is underutilized as a learning activity at all levels of education (for detailed discussions, see Roediger & Karpicke, 2006a; Yang et al., 2021). The present dissertation aimed to provide some guidance for students on how to approach their self-regulated learning, comparing the effectiveness of different learning activities that may be applied in-between a study session.

About a century ago, researchers began to consider that learning could happen not only when people encode knowledge and experiences, but also when information is retrieved from memory (Abbott, 1909; Thorndike, 1914; for a historical review, see Roediger & Karpicke, 2006a). Two classic studies on retrieval-based learning were carried out by Gates (1917) and Spitzer (1939), both of which demonstrated strong beneficial effects of retrieval practice on memory of previously studied information. Gates (1917) had children study prose passages about biographies and found that learning of the materials was best when the children spent about 60% of the total learning period involved in self-testing, as opposed to spending the majority of time rereading the materials. Similarly, Spitzer (1939) had students study prose passages on peanuts or bamboo plants and showed that performance on later memory tests could be greatly improved when the students were given an initial memory test relatively soon after the study session. Retrieval-based learning has received relatively little attention over a long period of time, but the past two decades have witnessed a substantial increase in research on retrieval practice (see Yang et al., 2021, for an illustration). A myriad of studies has now cemented the view that retrieval practice affords tremendous benefits for learning and represents a more powerful learning activity than a restudy opportunity (for reviews, see Agarwal et al., 2021; Karpicke, 2017; Roediger & Butler, 2011; Yang et al., 2021).

Retrieval practice does not only promote the learning of the previously studied material, a finding that has been termed the (backward) testing effect, but can also promote the learning of newly studied material encountered after retrieval practice (for

reviews, see Chan, Meissner, & Davis, 2018; Pastötter & Bäuml, 2014; Yang et al., 2018, 2021). The present dissertation is focused on the latter finding, which has sometimes been referred to as the forward testing (FT) effect, and offers three main contributions to the literature on retrieval-based learning. First, retrieval practice between study of prose passages can improve learning of future material compared to interim restudy, demonstrating that the FT effect arises with prose passages. The study materials used in prior work on the effect have mostly not reflected the complex information students are typically confronted with in educational practice. Employing educationally relevant study contents, the present work provides evidence that interim retrieval practice can indeed promote future learning when students face the study of more complex materials. Second, as the present results demonstrate, beneficial effects on future learning can also be obtained when, instead of interim retrieval practice, interim elaborative studying with concept mapping is implemented. Many researchers strongly recommended the use of retrieval practice in educational settings, and findings of similar levels of learning in response to retrieval practice and concept mapping have been rare, if not absent in the memory literature (see e.g., Karpicke & Blunt, 2011; Lechuga et al., 2015). The present work provides first evidence that interim concept mapping can be equally effective as interim retrieval practice for student learning, raising the possibility that concept mapping may be profitably used in lieu of retrieval practice in educational settings. Third, the present dissertation also addressed the impact of a retrieval activity other than retrieval practice, an activity that may reflect a quite typical way of how students spend breaks in-between a study session: thinking about what present to choose for a friend's birthday, what items to pack for vacation, or simply setting up the grocery list for the evening. Of course, such kinds of retrieval may not be considered a learning activity per se. Nonetheless, in some situations, they can be just as effective as retrieval practice in aiding future learning of complex information, as the present results show.

The FT effect was first discovered in research on proactive interference during the end of the verbal learning era (Darley & Murdock, 1971; Tulving et al., 1967; Tulving & Watkins, 1974; see also Malis, 1970). Before introducing the landmark studies on the FT effect, however, I briefly set the stage by describing what proactive interference actually means. Proactive interference refers to the finding that memory for recently studied,

critical information can be impaired by the prior study of further information (e.g., Underwood, 1957; for reviews, see Anderson & Neely, 1996; Crowder, 1976; Kliegl & Bäuml, 2021a). Buildup of proactive interference can be reflected in a number of measures. For instance, the preceding study of further material can reduce recall of the critical material, slow down retrieval of the critical material, and lead to intrusions from the preceding material when the critical material is attempted to be retrieved. Over the years, a number of explanations have been put forth to account for proactive interference, but the most widely accepted account is temporal discrimination theory. According to this account, proactive interference accrues due to a failure to distinguish the most recent material from material that was studied previously. Specifically, the account assumes that at the time of test, subjects are unable to restrict their memory search to the critical material and instead include the earlier material into their mental search set, leading to recall impairment (Baddeley, 1990; Crowder, 1976; Wixted & Rohrer, 1993). Proactive interference is considered to be one of the major forms of forgetting in what has been termed the episodic memory, which refers to the capacity to recollect specific events from the past and allows “mental time travel” (Tulving, 1972, 2002). Critically, research from the past decades has shown that there are tools available to reduce such proactive interference, and among these tools is interim retrieval practice (for a review on further tools, see Kliegl & Bäuml, 2021a).

A classic example comes from Tulving and Watkins (1974) who had subjects study a first list of word pairs (A–B) that was followed by the study of a second, related list of word pairs (A–D). Surprisingly, on a final recall test (A–?), subjects recalled more words from the first than from the second study list (i.e., $B > D$), which contrasted with the typical finding of higher recall of the most recent list (i.e., $D > B$) in the verbal learning literature (e.g., Keppel, 1968). The researchers discovered that these results emerged due to the absence of interim retrieval practice, and that inclusion of the customary retrieval practice between study of the lists led to release from proactive interference for the second list (see also Malis, 1970). Another classic example comes from Darley and Murdock (1971) who had subjects study 10 lists of words in anticipation of a final cumulative test. Half of the lists were followed by retrieval practice of the just studied list, and the other half were followed by a distractor activity. Strikingly, 81% of prior-list intrusions that subjects produced during the retrieval-practice tasks were words from previously

nonretrieved lists, indicating that retrieval practice on prior study lists inoculated subjects against proactive interference from those lists.

In 2008, Szpunar and colleagues published their seminal paper that has revived interest in the FT effect. In their Experiment 2, Szpunar et al. demonstrated that with an increasing number of prior study lists, recall of a final critical list decreased and the number of prior-list intrusions increased on an immediate recall test on the critical list. More important, the researchers showed that these detrimental effects were completely prevented when retrieval practice was inserted between study of the lists. These findings indicate that proactive interference builds up across successive study lists, but interim retrieval practice can lead to recovery from proactive interference and thus induce a recall benefit. Researchers since have used a variety of terms to describe the beneficial effects of interim retrieval practice on new learning. One of these terms has been the forward testing effect (see Pastötter & Bäuml, 2014), which is also used in the present dissertation.

To date, proactive-interference buildup and release has mostly been demonstrated using word lists as study material (for an early exception, see Blumenthal & Robbins, 1977). The present dissertation contributes to the literature by showing that proactive interference can accrue when prose passages are studied in succession, and by showing that, even with such complex materials, proactive interference can be reduced when retrieval practice is inserted in the study session. For the sake of completeness, it should be mentioned that whether the FT effect reflects release from proactive interference is a subject of debate in the memory literature (see e.g., Ahn & Chan, 2022; Chan, Manley, et al., 2018; Wissman et al., 2011; Yang et al., 2022). For instance, some researchers recently pointed out that the beneficial effects of interim retrieval practice emerged even in the absence of proactive interference, implying that proactive-interference release could be a byproduct rather than a causal factor of the FT effect (Ahn & Chan, 2022). Although the present experiments were not designed to directly address this question, the results may bear some implications for this debate as well (see chapters 5.1.3 and 5.2.5).

While from an interference point of view, preceding study should be detrimental to future learning, other lines of research indicated that prior learning experiences can be beneficial for future learning. In fact, beginning with the pioneering work by de Groot (1966), research since then has fostered the view that learning of new information can be vastly

improved by prior knowledge in the domain of the information (for reviews, see Ericsson et al., 2006; Gobet, 2015; Vicente & Wang, 1998). For instance, in one classic demonstration, Chase and Simon (1973) had participants recreate chess boards that were shown to them for a mere five seconds. When the chess boards were based on real game positions, chess masters were able to reconstruct, on average, the positions of 16 out of 28 pieces, while advanced-rated players could only remember seven, and chess novices could only remember four. When the pieces were positioned randomly in a way that did not “make sense”, however, all three groups could only remember three or four positions, suggesting that prior knowledge can facilitate learning of new information when the preexisting knowledge can be applied to the information. Another classic demonstration comes from Bransford and Johnson (1972) who had participants study a prose passage. The passage was written in an ambiguous manner that made it hard for participants to figure out what the passage was about (i.e., doing laundry). Consistently, when the topic of the passage was not provided, participants rated the passage as incomprehensible and successfully recalled, on average, only about three ideas from the passage. When provided with the topic “washing clothes”, however, participants rated the passage as easy to understand and were able to recall about six ideas on average. Crucially, both comprehension and recall were found to benefit only when the topic was provided before as opposed to after study, indicating that the learning of new information can be aided when relevant background knowledge can be used during study.

A more recent study by Wissman et al. (2011) provided preliminary support for the idea that background knowledge may also aid the learning of new material when such knowledge has been promoted by interim retrieval practice. In their Experiment 4, the researchers had subjects study prose passages comprising highly related material, and it was manipulated whether subjects between study of the single passages engaged in retrieval practice or not. Strikingly, interim retrieval practice not only improved recall of the final passage relative to the no-retrieval-practice condition, but also improved recall relative to a baseline condition in which only the final passage was studied. These results provide evidence that interim retrieval practice not only reduces forgetting as caused by the buildup of proactive interference (see above), but can even increase memory relative to baseline performance and thus induce proactive facilitation (for related findings in paired-associate learning, see Wahlheim, 2015). The present dissertation extends on these

results by showing that with highly related but not with unrelated prose passages, interim retrieval practice induces such proactive facilitation. These findings basically fit with the idea that effects of background knowledge on learning are restricted to domain-related material.

The present dissertation relates the FT effect with prose materials to context-dependent memory. In a number of studies, context change between the study of materials has been shown to reduce proactive interference, be it an external context change (e.g., a change in the physical environment; Dallett & Wilcox, 1968; Smith et al., 1978) or an internal context change (e.g., deliberately thinking about another time or place; Pastötter & Bäuml, 2007; Sahakyan & Kelley, 2002). The proposal is that, in the absence of such context change, the mental context in which the materials are encountered drifts slowly over time, leading to similar context cues stored along with the studied information (Bower, 1972; Estes, 1955; Mensink & Raaijmakers, 1988). Context change between study of the materials may accelerate the contextual drift, making it easier to discriminate the subsequent from the prior material on a temporal basis and allowing for a more focused memory search at test (Bäuml & Kliegl, 2013; Sahakyan & Kelley, 2002; Smith & Vela, 2001). Context change can be triggered by various mental tasks, and one of them is interim retrieval practice (Jang & Huber, 2008; Shiffrin, 1970). Prior work has already suggested a link between context-dependent memory and the FT effect with relatively simple study materials (Kliegl & Bäuml, 2021b; Pastötter et al., 2011; Szpunar et al., 2008; see also Divis & Benjamin, 2014). The present dissertation extends on the prior work by pointing to a link between context-dependent memory and the FT effect with unrelated prose materials.

The present dissertation also relates the FT effect with prose materials to knowledge-enhanced learning. A large body of research has established that accumulating knowledge in a certain domain can potentiate the learning of material related to this domain, sometimes even dramatically (see above). Crucially, when it comes to the learning of text information, the potentiating effects of prior knowledge on learning have been attributed to knowledge integration. The proposal is that, when essential prior knowledge lacks, learners experience comprehension gaps during study, making encoding of the text relatively hard. Prior knowledge in a text's domain may

enable learners to make inferences on how the single parts of the text fit together, resulting in better encoding of the material due to its integration with the knowledge base of the learner (Kintsch, 1988, 1994, 1998). The present thesis now addressed the question of whether text comprehension can also be aided when requisite background knowledge is accumulated by interim retrieval practice. Prior research on the backward testing effect has already demonstrated that knowledge on text information can indeed be promoted by retrieval practice on the information (Butler, 2010; Hinze et al., 2013; Karpicke & Blunt, 2011; McDaniel et al., 2009). Furthermore, prior work on the FT effect has shown that interim retrieval practice can enhance knowledge integration from across a study session (Jing et al., 2016) and that interim retrieval practice can lead to proactive facilitation (Wissman et al., 2011; see above). The present dissertation extends on these prior findings and points to parallels between the FT effect with highly related prose materials and knowledge-enhanced learning.

In sum, relatedness of materials seems to influence how retrieval practice promotes new learning. Prior work already demonstrated a critical influence of study material for the FT effect with simple word lists (Kliegl & Bäuml, 2021b, 2023). Extending on the prior work, the present dissertation now demonstrates a critical role of study material for the FT effect with prose passages. A new two-factor explanation of the FT effect will be proposed, according to which the FT effect primarily relies on context-dependent memory with unrelated prose materials and primarily relies on knowledge-enhanced learning with highly related prose materials. The suggested two-factor explanation raises a couple of novel predictions for the FT effect with prose passages, some of which will be introduced in the general discussion of the present thesis (see chapters 5.5.1 to 5.5.3).

1

The Forward Testing Effect: Empirical and Theoretical Overview

Courses are a central part of a student's educational experience. Students of psychology, for instance, often participate in several courses during a semester, like courses on clinical psychology, organizational psychology, or educational psychology. At the end of the semester, when preparing for the final exams, the students may therefore during some weeks have to study quite unrelated materials relevant to different courses, and during other weeks highly related materials relevant to a single course. In both scenarios, it may be hard for students to learn material encountered later in the study sessions, considering the vast amount of material that will be tested, and considering that learning tends to decline the more material is studied in succession (e.g., Johnstone & Percival, 1976; Lindquist & McLean, 2011; Risko et al., 2012). The question thus arises of how the students can optimize new learning in the two putative scenarios during exam preparation. The present dissertation aimed to find some answers to this question.

Exploring which activities can serve to improve learning has long been a goal in memory research (Ebbinghaus, 1913). While, over the past decades, various activities have been found to aid student learning (for reviews, see Dunlosky et al., 2013; Putnam et al., 2016), one of the most widely studied findings in the memory literature is that active retrieval practice enhances learning and memory. Hundreds of studies have repeatedly demonstrated that retrieval practice on studied material can enhance its long-term retention much more than restudy of the same material, a finding that has been termed the testing effect (e.g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006b). Typically, participants study a set of materials, and then are asked to practice the materials by virtue of restudy or (repeated) retrieval. Finally, the participants often take a final recall test which may be conducted from a few minutes after practice up to several days later.

In most studies, recall has been found to be higher after retrieval practice than restudy practice. The testing effect is a robust effect that has been observed across a wide range of materials, subject groups, and experimental settings (for reviews, see Karpicke, 2017; Roediger & Butler, 2011). Testing can enhance long-term retention even in the classroom with real course materials (for reviews, see Agarwal et al., 2021; Yang et al., 2021). Given these and related findings, testing has often been advocated as a powerful tool to improve learning, contrasting with the widely-held assumption among students and educators that tests serve primarily to assess learning (Morehead et al., 2016; Roediger & Karpicke, 2006b).

The testing effect is sometimes called a direct benefit of testing (Roediger & Karpicke, 2006a). Indeed, the act of practicing retrieval has been shown to directly improve retention of the practiced material, even without restudying or receiving feedback after retrieval. However, testing can also have indirect benefits for learning. For instance, retrieval practice on studied material has been shown to enhance relearning of the same material, a finding that has been termed test-potentiated learning (e.g., Arnold & McDermott, 2013; Izawa, 1971). Critically, and of central interest for the present dissertation, retrieval practice can also improve learning of new material encountered after practice of the previously studied material. The finding is often referred to as the forward testing (FT) effect, in contrast to the classic testing effect which is sometimes referred to as the backward testing effect (Pastötter & Bäuml, 2014; Yang et al., 2018).

The FT effect has been demonstrated in multiple-list learning paradigms (for different variants of these paradigms, see Chan, Meissner, & Davis, 2018). Although the effect was originally discovered about fifty years ago (Darley & Murdock, 1971; Tulving & Watkins, 1974), research on the FT effect has accelerated only since a publication by Szpunar et al. (2008). In their landmark study, Szpunar et al. had participants study five lists of words in anticipation of a final cumulative recall test. Immediately after study of lists 1 through 4, participants were either asked to recall the words of the preceding list (retrieval-practice condition), study the words once again (restudy condition), or solve simple arithmetic tasks (distractor condition). After study of list 5, all participants took an immediate recall test on this critical final list. Results showed that subjects in the retrieval-practice condition recalled more words from critical list 5 and showed fewer intrusions of words from lists 1 through 4 than subjects in the other two conditions. This

beneficial effect of interim retrieval practice, be it relative to restudy or other no-retrieval-practice conditions, has often been termed the FT effect. Although the FT effect can be reflected in both correct recall and the intrusion data, results from the two outcome variables have sometimes diverged and shown enhancements in recall without reductions in intrusions (e.g., Ahn & Chan, 2022; Pierce et al., 2017), and reductions in intrusions without recall enhancements (e.g., Kliegl & Bäuml, 2021b; Weinstein et al., 2014).

Like the backward testing effect, the FT effect has turned out to be a very robust and general effect (for reviews, see Chan, Meissner, & Davis, 2018; Pastötter & Bäuml, 2014; Yang et al., 2018). It has been found for different numbers of study lists (e.g., two lists, three lists, five lists), and has been shown to arise across a variety of study materials, including lists of single words (Chan, Manley, et al., 2018; Pastötter et al., 2011; Szpunar et al., 2008), paired associates (e.g., foreign-translation pairs, Cho et al., 2017; face-name pairs, Weinstein et al., 2011), and more complex study materials like prose passages (Wissman et al., 2011) or video lectures (Jing et al., 2016; Szpunar et al., 2013; Yue et al., 2015). The FT effect has also been shown to arise with nonverbal materials such as motor sequences (Tempel & Frings, 2019), spatial information (Bufe & Aslan, 2018), or artists' painting styles (Lee & Ahn, 2018; Yang & Shanks, 2018), and the effect was even found when the types of materials were changed across the study session (e.g., from factual statements to paintings; Yang et al., 2019). Although the FT effect has mostly been explored in young and healthy college students, some studies provided evidence that the effect generalizes to older adults (Pastötter & Bäuml, 2019), elementary school children (Aslan & Bäuml, 2016; Dang et al., 2021), and patients suffering from traumatic brain injury (Pastötter et al., 2013). Finally, the FT effect has not only been found in carefully controlled laboratory settings, but also in the classroom (Todd et al., 2021) and in other settings providing less experimental control, like online learning (Weinstein et al., 2014) or self-paced learning procedures (Bufe & Aslan, 2018; Yang et al., 2017).

The literature reviewed above suggests that the FT effect may be applied to educational practice. Returning to the introductory question about how to optimize new learning when, during exam preparation, students face the study of rather unrelated course materials or the study of highly related course materials, research to date has provided us with fragmentary information, however. Studies on the FT effect have mostly employed fairly simple study materials like word lists or word pairs, and studies on the effect with

more complex study materials are still relatively scarce in the literature. However, since interim retrieval practice is increasingly being recommended to students as an effective learning tool (e.g., Pastötter & Bäuml, 2014; Yang et al., 2021), high confidence is requisite that the FT effect also holds for the kinds of materials commonly found in educational contexts, such as prose passages and video lectures. A primary goal of the present dissertation was to address the issue, examining whether retrieval practice can induce an FT effect with both related and unrelated complex materials. The following section shortly summarizes the relevant literature.

1.1 The Forward Testing Effect With More Complex Study Material

Long before the FT effect has been extensively researched across the past 15 years, a number of studies from almost 60 years ago provided evidence that test-like questions interspersed between the study of prose materials can facilitate learning (e.g., Bruning, 1968; Frase, 1968a, 1968b; Hershberger & Terry, 1965; Rothkopf, 1966; Rothkopf & Bisbicos, 1967). In one of these studies, for instance, Rothkopf (1966) had subjects study a book chapter for an upcoming memory test. In the retrieval-practice condition, subjects were asked to answer questions on the contents after reading each of several passages of the chapter, while in the control condition, subjects were asked to read through the chapter in an uninterrupted fashion. Results showed that subjects in the retrieval-practice condition outperformed those in the control condition in a final test covering both practiced and unpracticed contents of the chapter. However, and in contrast to the typical FT-effect paradigm, these early studies employed procedures in which all of the presented passages were practiced before memory of the final passage was assessed, making it difficult to disentangle FT effects from backward testing effects and other putative memory effects.

To the best of my knowledge, only three experiments published in three separate studies addressed the FT effect to date employing unrelated complex materials. One study used two unrelated video lectures (“star life cycle” and “lightning formation”) and found an FT effect for the second lecture when subjects engaged in retrieval practice compared

to restudy practice of the first lecture (Yue et al., 2015, Experiment 2). However, only a single lecture was presented before the critical final lecture, and only half of the contents of the first lecture were practiced. Whether results generalize to larger numbers of preceding lectures and nonselective interim practice thus remains unclear. In the second study, subjects were presented with three prose passages on a common topic (“forms of government intervention in the U.S. labor market”), but the three passages (“benefit mandates”, “labor laws”, and “job training programs”) were not directly related to one another and had no overlapping information (Wissman et al., 2011, Experiment 1a). Results revealed an FT effect for the third passage when subjects were asked to retrieve passages 1 and 2 following initial study relative to when all three passages were studied in an uninterrupted fashion. The results leave it open, however, whether the FT effect with unrelated passages also shows up relative to a restudy control condition (for a more detailed discussion on the issue, see below). The third study employed completely unrelated prose passages as study materials (“coyotes”, “porcupines”, “seals”, and “chronic wasting disease”) but did not directly address the effects of retrieval practice. Rather, it examined the presence of an FT-like effect when subjects between study of the single passages engaged in semantic generation, a retrieval activity in which subjects generated exemplars from semantic categories unrelated to the study passages (e.g., professions; Divis & Benjamin, 2014, Experiment 2). Results indeed showed an FT-like effect in response to semantic generation relative to a distractor control condition. However, it is far from clear whether the finding generalizes to retrieval practice, given that retrieval practice and semantic generation can well differ in their effects on recall of subsequently studied material (e.g., Kliegl & Bäuml, 2021b, 2023).

The FT effect with related complex materials has been observed in at least four studies to date. Three of the studies divided a coherent video lecture on a single topic (i.e., “introduction to statistics” or “public health in the 20th century”) into several integrable sections and found an FT effect for the critical final section when subjects engaged in interim retrieval practice between the single sections, both relative to restudy and distractor control conditions (Jing et al., 2016; Szpunar et al., 2013, 2014). Similarly, Wissman et al. (2011) reported an FT effect, relative to no-retrieval-practice conditions, in the learning of prose passages that were directly related to one another and contained information intended to be integrated (i.e., on “capturing and storing atmospheric

greenhouse gases” or “inconsistencies between Hollywood’s depiction of history and factual history”). In this study, recall levels were also compared to a condition in which only the critical final passage was presented for study, with recall in this condition being lower than recall in the retrieval-practice condition, indicating that interim retrieval practice induced proactive facilitation. The finding of proactive facilitation is particularly striking from a traditional interference point of view, because recall of recently studied material is usually lower in the presence than absence of prior study of further material, a finding that has been termed proactive interference (e.g., Underwood, 1957; for reviews, see Anderson & Neely, 1996; Crowder, 1976; Kliegl & Bäuml, 2021a).

Studies on the FT effect with more complex study material thus are still scarce in the literature. Moreover, just a few of them employed a restudy control condition, leaving it largely open whether the FT effect with more complex materials is retrieval specific. Retrieval specificity relates to the question whether the FT effect is induced by retrieval practice *per se*, or is rather induced by reexposure to study materials inherent to retrieval practice (see also Roediger & Karpicke, 2006b; Szpunar et al., 2008). If the FT effect was induced by reexposure – and thus was not retrieval specific – the effect should disappear or even be reversed relative to a restudy control condition, because subjects in this condition are reexposed to all previously studied materials, whereas subjects in the retrieval-practice condition are reexposed to only those materials they have been able to retrieve. Although some prior work suggests that the FT effect with more complex materials is retrieval specific (Jing et al., 2016; Szpunar et al., 2013, Experiment 2; Yue et al., 2015, Experiment 2), the restudy conditions employed for comparison involved merely plain reexposure to materials. The question thus arises whether retrieval specificity also holds for other restudy conditions, such as those involving elaborative processing of the materials. In educational settings, restudy practice is often recommended to students (for discussions, see Rawson & Kintsch, 2005; Roediger & Karpicke, 2006b), and is typically applied much more frequently than retrieval practice by students and educators alike (Miyatsu et al., 2018; Roediger & Karpicke, 2006a). The issue of whether the FT effect is retrieval specific for materials and practice types found in educational contexts is therefore of high relevance for applied purposes.

Research on the FT effect with more complex study material obviously shows a number of empirical gaps. Thus, providing clues for students on how to optimize new

learning when, during exam preparation, they face the study of quite unrelated course materials or the study of highly related course materials may be premature. First, it is largely unclear to date whether retrieval practice induces an FT effect not only with related complex materials but also with completely unrelated complex materials. In practice, this raises the issue of whether the relatedness of course materials can influence whether in-between retrieval practice optimizes new learning. Second, is the finding of an FT-like effect in response to semantic generation restricted to unrelated complex materials, or does it generalize to related complex materials? The question relates to the issue of whether, in practice, future learning is also improved if, between study of related materials, students engage in retrieval of other information than the just studied material – for instance, by setting up the evening grocery list. Third, does the finding of proactive facilitation – i.e., higher recall in the retrieval-practice than in a final-material-only condition – generalize from related to unrelated complex materials? More practically, the question asks whether costs in recall of the final materials do also arise if students were skipping the preceding study of unrelated materials. And fourth, is the FT effect retrieval specific with more complex study material? In practice, it is important to know whether learning and memory of future material can also be aided when, instead of interim retrieval practice, elaborative restudy practice is applied in-between the study, and whether such FT-like effects do arise regardless of whether students are facing the study of highly related materials or unrelated materials. Addressing these issues can help to provide some useful clues with regard to the putative courses scenarios entertained above. In addition, it can help to improve our theoretical understanding of the FT effect, on which is elaborated in the following sections.

1.2 Theoretical Accounts of the Forward Testing Effect and Empirical Support

Various explanations have been put forth to account for the FT effect with more complex study material, which are often similar to those of the FT effect with simple word lists (for reviews, see Chan, Meissner, & Davis, 2018; Pastötter & Bäuml, 2014; Yang et al., 2018). The present dissertation is focused on three explanations suggested in the

literature: the context-change account, the strategy-change account, and the facilitated-comprehension account of the FT effect (for attentional accounts, see chapter 5.2.4). According to the context-change account¹, the FT effect reflects release from proactive interference, with retrieval practice insulating the new material against buildup of proactive interference from the preceding materials via an enhancement in “list” discrimination (Divis & Benjamin, 2014; Pastötter et al., 2011; Szpunar et al., 2008). The account is based on the widespread idea that during encoding of material, the material is stored along with details about the mental context currently present in the learner’s mind (Estes, 1955; Mensink & Raaijmakers, 1988). The critical idea is that retrieval between study of materials drives mental context change, and thus creates distinct context cues for the preceding and the newly studied materials (Jang & Huber, 2008; Shiffrin, 1970). At test, the use of such distinct context cues is assumed to afford a more focused memory search for the new material, reducing proactive interference from the preceding materials and thus creating an FT effect (Divis & Benjamin, 2014; Pastötter et al., 2011; Szpunar et al., 2008).

Evidence in favor of the context-change account of the FT effect, for instance, comes from studies demonstrating that the FT effect is accompanied by shorter response latencies (Bäuml & Kliegl, 2013; Lehman et al., 2014). In one study, Bäuml and Kliegl (2013) had subjects study three lists of words, with critical list 3 being tested immediately after study. Lists 1 and 2 were either retrieved or restudied following initial study. In addition, the experiment included a condition in which subjects were asked to just study list 3, with recall in this condition serving as a baseline to measure amount of proactive interference as induced by prior study and practice of lists 1 and 2 in the retrieval-practice and restudy conditions. Both recall rates and response latencies of list-3 words were

¹ The context-change account is often referred to as the release-from-proactive-interference account in the literature on the FT effect (e.g., Ahn & Chan, 2022; Szpunar et al., 2008; Yang et al., 2018, 2022). However, because release from proactive interference is not necessarily driven by context change, but may also be driven by other cognitive mechanisms (see e.g., Kliegl & Bäuml, 2021a), the account is referred to as the context-change account in the present dissertation.

analyzed. Recall in the restudy condition was lower than in the list-3-only condition, reflecting buildup of proactive interference, and response latencies in the restudy condition were longer than in the list-3-only condition, indicating a slowing of the retrieval process. Critically, retrieval practice between study of the lists did not only increase recall to the level in the list-3-only condition, reflecting a perfect release from proactive interference, but did also decrease response latencies to the level in the list-3-only condition. Because shorter response latency is assumed to indicate a smaller size of the mental search set (Rohrer, 1996; Wixted & Rohrer, 1993), the findings provide support for the context-change account's assumption that the FT effect reflects release from proactive interference caused by more focused memory search at test.

According to the strategy-change account of the FT effect, retrieval practice may enable subjects to optimize their strategies for further learning, assuming that, during recall testing, subjects gain performance-relevant insights into the learning task at hand (Chan, Manley, et al., 2018; Soderstrom & Bjork, 2014; Wissman et al., 2011). Indeed, during recall testing, subjects can build expectations about the test format they will encounter (e.g., free recall), and also about the type of information needed for successful recall (e.g., free recall requires one to generate retrieval cues; Finley & Benjamin, 2012; Thiede et al., 2011). Similarly, recall testing can alert subjects to retrieval failures and thus serve to counteract overestimation of their (poor) prior learning (Little & McDaniel, 2015; Pyc & Rawson, 2012). The critical idea is that test expectations and retrieval failures may induce subjects to switch their strategies for subsequent learning, resulting in potentially more efficient encoding and retrieval of material studied after recall testing (Chan, Manley, et al., 2018; Soderstrom & Bjork, 2014; Wissman et al., 2011).

Support for the strategy-change account of the FT effect, for instance, stems from studies demonstrating that the FT effect is characterized by superior recall organization, with superior organization pertaining to semantic characteristics of the study material (Chan, Manley, et al., 2018; Chan et al., 2020; Kliegl & Bäuml, 2021b) or pertaining to the temporal order of its initial presentation (Yang et al., 2022). In one study, Chan, Manley et al. (2018) had subjects study four lists of words in which different exemplars from the same semantic categories were presented both within and across the lists. Between study of the lists, subjects were either asked to recall the immediately preceding list, restudy the list, or engage in unrelated distractor activities. After study of list 4,

subjects were asked to freely recall the words of this critical final list. Results did not only reveal an FT effect, with higher recall after retrieval practice compared to the other two conditions, but also that subjects in the retrieval-practice condition showed a stronger propensity to cluster their recall on the basis of the words' category membership. Because recall clustering may reflect strategic processing during encoding and/or retrieval of the recalled material, the findings provide evidence for the strategy-change account's assumption that the FT effect is caused by strategy optimization.

While the context-change and strategy-change accounts of the FT effect have been put forth to explain the FT effect with both simple word lists and more complex study material, the facilitated-comprehension account of the FT effect is dedicated to complex materials only. The facilitated-comprehension account suggests that retrieval practice enhances accessibility of the preceding materials during subsequent study of the new material, which may foster the integration of knowledge deriving from different parts of the study material and thus facilitate comprehension of the new material (Jing et al., 2016; Wissman et al., 2011). The account is based on the idea that retrieval practice serves to reactivate the practiced material, increasing the likelihood that the material is covertly retrieved during subsequent encoding of related material, which may aid the integration of the practiced and the new material into a unified memory representation (Hintzman, 2004, 2010; Wahlheim, 2015; Wahlheim & Jacoby, 2013).

To the best of my knowledge, no study has yet provided specific evidence for the facilitated-comprehension account of the FT effect (for a discussion, see chapter 4.2). However, one study provided support for the account's more basic idea that interim retrieval practice can promote new learning via an integration-based mechanism. In this study, Wahlheim (2015) presented subjects with two lists of word pairs, and it was manipulated whether list-1 pairs were restudied (e.g., *pearl-harbor*) or retrieved (e.g., *pearl-?*) before list 2 was presented. List 2 contained control pairs that had not appeared in list 1 (e.g., *soup-bowl*), but also changed pairs for which the first word had already appeared in list 1 but the second word was new (e.g., *pearl-jewelry*). After study of list 2, subjects were tested on list 2 by showing them the first word of each pair (e.g., *pearl-?*) and asking them to recall the second word of the pair that had appeared in the second list (e.g., *jewelry*). Results revealed an FT effect for the changed pairs, with higher recall of list-2 responses after retrieval practice compared to restudy of the list-1 responses. More

important, retrieval practice was also found to increase the likelihood that subjects were able to recollect the changes in responses at test, and the FT effect for the changed pairs disappeared when recall of list-2 responses was conditionalized on whether or not the change in responses was recollected. Critically, change recollection is assumed to indicate integration of the changed responses into a unified memory representation (Jacoby et al., 2015; Wahlheim & Jacoby, 2013), and such integration has been attributed to covert retrieval of a previously studied response during subsequent encoding of a new but related response (Hintzman, 2004, 2010). The findings therefore suggest that interim retrieval practice can promote new learning by fostering study-phase retrieval and integration, an idea that has also been incorporated in the facilitated-comprehension account of the FT effect with more complex materials.

1.3 The Forward Testing Effect With Complex Material: Empirical Findings and Theoretical Implications

The context-change account, the strategy-change account, and the facilitated-comprehension account of the FT effect are not mutually exclusive, but their logic sometimes offers different predictions about basic characteristics of the FT effect. The key predictions of the three accounts have hardly been tested employing complex study materials, however, which is why definitive answers on exactly which mechanisms mediate the FT effect with these types of materials are still missing. Moreover, on the basis of the prior research, fairly mixed answers result concerning the question of whether one of the three accounts may best explain the FT effect with complex materials. Indeed, the finding of proactive facilitation that was reported for the FT effect with related prose materials (Wissman et al., 2011) aligns with both the strategy-change and facilitated-comprehension accounts, but challenges the context-change explanation. In contrast, the finding of an FT-like effect in response to semantic generation that was reported for unrelated prose materials (Divis & Benjamin, 2014) is consistent with the context-change account, but is covered by neither the strategy-change nor the facilitated-comprehension explanation.

According to the context-change account of the FT effect, which assumes that the FT effect is mediated by mental context change, the FT effect should not be tied to (episodic) retrieval involving practice of the previously studied material, but generalize to an FT-like effect in response to various forms of memory retrieval, such as (semantic) retrieval of information unrelated to the studied material. Consistently, findings from a number of studies indicate that various forms of retrieval can drive mental context change, enhancing list discrimination and inducing release from interference at test (e.g., Divis & Benjamin, 2014; Jang & Huber, 2008; Pastötter et al., 2008, 2011; Sahakyan & Kelley, 2002; Shiffrin, 1970; Tulving & Watkins, 1974). In particular, semantic-generation tasks have been shown to accelerate contextual drift, resulting in context-dependent memory effects that can be similar to those as induced by retrieval practice (Divis & Benjamin, 2014; Jang & Huber, 2008; Kliegl & Bäuml, 2021b, 2023; Pastötter et al., 2011). However, on the basis of the context-change account, the FT effect should not reflect proactive facilitation. The assumption is that, in the absence of context change, subjects are unable to focus their memory search on the final material at test, but rather search the entire set of materials that have previously been exposed, leading to the buildup of proactive interference (Baddeley, 1990; Crowder, 1976; Wixted & Rohrer, 1993). Context change between study of materials is assumed to reduce such proactive interference by eliminating the preceding materials from the mental search set, leading to increased recall of the final material at the time of test (Bäuml & Kliegl, 2013; Sahakyan & Kelley, 2002; Smith & Vela, 2001). If so, the upper limit for the recall enhancement as induced by context change should be marked by the level of recall when only the final material has been exposed and the preceding materials have not been presented. The context-change account therefore predicts that interim retrieval practice can enhance recall of the final material to a level similar but not superior to the level in a final-material-only condition, resulting in an FT effect that reflects proactive-interference release but not proactive facilitation.

In contrast to the context-change account, both the strategy-change and facilitated-comprehension accounts of the FT effect may explain why the FT effect can reflect proactive facilitation. According to the two accounts, retrieval practice may increase the effectiveness of subsequent learning, assuming that it can induce a switch to more efficient encoding and retrieval strategies for subsequent learning (Chan, Manley, et al.,

2018; Soderstrom & Bjork, 2014; Wissman et al., 2011), or assuming that it can induce facilitated comprehension during subsequent encoding (Jing et al., 2016; Wissman et al., 2011). Because study and practice of preceding material are required to induce strategy change or facilitated comprehension, the two accounts suggest that the learning of critical material may be more effective when its study is preceded by additional material. If so, retrieval practice on preceding materials may enhance recall of critical final material even above the recall level in a final-material-only condition, and the FT effect thus reflect proactive facilitation. However, on the basis of the two accounts, the FT effect should not generalize to an FT-like effect in response to semantic generation. Retrieval of semantic information unrelated to the study material does not provide insights into the learning task that would promote the application of new and more efficient strategies during subsequent encoding and retrieval (Kliegl & Bäuml, 2021b). Similarly, semantic generation does not increase the accessibility of previously studied information (Divis & Benjamin, 2014), suggesting that semantic generation does not afford facilitated comprehension during subsequent encoding.

While the strategy-change and facilitated-comprehension accounts of the FT effect thus offer similar predictions on whether semantic generation can induce an FT-like effect and whether the FT effect can reflect proactive facilitation, the two accounts clearly differ in whether or not they predict retrieval specificity of the FT effect. According to the strategy-change account of the FT effect, which assumes that the FT effect is mediated by strategy optimization, the FT effect should be retrieval specific and thus not generalize from retrieval practice to any type of restudy practice. Retrieval practice but not restudy practice can provide performance-relevant insights – for instance, into retrieval failures and the format of forthcoming recall tests – suggesting that retrieval practice but not restudy practice should promote strategy optimization and elicit an FT effect (Chan, Manley, et al., 2018; Pyc & Rawson, 2012; Soderstrom & Bjork, 2014). The facilitated-comprehension account of the FT effect, in contrast, does not allow clearcut a-priori predictions about retrieval specificity of the FT effect. However, because the account assumes that the FT effect is mediated by knowledge integration (Jing et al., 2016; Wissman et al., 2011; see also Wahlheim, 2015), the suggestion may arise that the FT effect is not retrieval specific but rather generalizes to an FT-like effect in response to certain types of restudy practice. Previous research has already established that prior

knowledge in a given domain can potentiate the learning of material related to this domain (e.g., Bransford & Johnson, 1972; Buchin & Mulligan, 2023; McNamara et al., 1996), and that knowledge acquisition can be fostered by various learning activities, including retrieval practice (see Agarwal et al., 2021; Karpicke, 2017) and elaborative studying (see Nesbit & Adesope, 2006; Schroeder et al., 2018). If so, and if the FT effect with more complex study material indeed resembled knowledge effects on learning, not only retrieval practice on the preceding materials may afford facilitated comprehension of the final material, but also other types of interim learning activities, such as, for instance, elaborative restudy practice. The context-change account of the FT effect, however, offers a similar prediction as the strategy-change account of the FT effect, and predicts that the FT effect should be retrieval specific. Retrieval practice but not restudy practice can drive mental context change and enhance “list” discrimination, suggesting that retrieval practice but not restudy practice should reduce the mental search set size at test and induce release from proactive interference (Bäuml & Kliegl, 2013; Pastötter et al., 2011; Szpunar et al., 2008).

In sum, none of the three accounts of the FT effect can explain both the finding of proactive facilitation that was reported for the FT effect with related prose materials (Wissman et al., 2011) and the finding of an FT-like effect in response to semantic generation that was reported for unrelated prose materials (Divis & Benjamin, 2014). With the challenge to all accounts that attribute the FT effect with the two types of materials to a common single mechanism, the two findings raise the possibility that relatedness of study materials may influence the FT effect with more complex materials, and, for instance, context change mediate the effect with unrelated materials and strategy change or facilitated comprehension with related materials. If so, and if facilitated comprehension indeed contributed to the FT effect with related materials, retrieval specificity of the FT effect may also depend on relatedness of study materials, and the effect be retrieval specific with unrelated but not with related materials. In line with such a proposal, knowledge effects on learning have been shown to arise for domain-related but not for domain-unrelated information (e.g., Hunt & Rawson, 2011; Witherby & Carpenter, 2022), suggesting that the prior study and practice of related material is required to afford facilitated comprehension of the final material.

1.4 The Potential Role of Relatedness of Study Materials

The question of whether relatedness of study materials influences the FT effect with more complex study material is a critical gap in research on the FT effect. The issue has recently been addressed in word-list studies on the FT effect, with evidence that relatedness of materials can influence the FT effect with simple word lists. Employing lists of unrelated words as well as lists of words in which different exemplars from the same categories were provided both within and across lists, previous studies have found the FT effect to be present with both types of word lists (e.g., Kliegl & Bäuml, 2021b, 2023; Szpunar et al., 2008). However, only with unrelated lists did FT-like effects arise in response to semantic generation (Divis & Benjamin, 2014; Kliegl & Bäuml, 2021b, 2023; Pastötter et al., 2011); and only with categorially related lists did the FT effect persist across a longer delay interval between study and recall of the critical list, as well as a longer lag between the study of the preceding lists and study of the critical list (Chan, Manley, et al., 2018; Kliegl & Bäuml, 2021b). The differences in findings between the two types of lists were interpreted as evidence for a two-factor account of the FT effect, which suggests a major role of context change in the FT effect with unrelated lists and a major role of strategy change in the FT effect with related lists (Kliegl & Bäuml, 2021b, 2023). Indeed, because context-change effects have been shown to be relatively short-lived (Abel & Bäuml, 2017; Divis & Benjamin, 2014) and to arise in response to both retrieval practice and semantic generation (Jang & Huber, 2008; Shiffrin, 1970), the findings on the FT effect with unrelated lists align with the context-change explanation. In contrast, because changes in encoding and retrieval strategy are assumed to be lasting and should rely on insights into the learning task that are tied to retrieval practice (Chan, Manley, et al., 2018; Soderstrom & Bjork, 2014), the findings on the FT effect with related lists agree with the strategy-change explanation.

Word-list studies thus provided evidence that relatedness of study materials can influence the FT effect with simple word lists, alluding to the possibility that relatedness of materials may also influence the FT effect with more complex materials. However, due to the many empirical gaps in research on the effect with more complex study material, the previous findings do not allow to draw firm conclusions on whether qualitatively

different FT effects also arise for related and unrelated complex materials. Indeed, the presence of an FT-like effect in response to semantic generation was examined for unrelated prose materials only (Divis & Benjamin, 2014), and the study included neither a retrieval-practice nor a final-material-only condition. Similarly, the presence of proactive facilitation in the FT effect was examined for related prose materials only (Wissman et al., 2011), and the study did not include a semantic-generation condition. The differences in comparison conditions do not allow for comparison of results across the two types of materials employed. In particular, no study has yet examined whether the FT effect generalizes to an FT-like effect in response to elaborative restudy practice, and whether retrieval specificity of the FT effect can vary for related versus unrelated complex materials. Prior research is therefore effectively silent on whether relatedness of materials influences the FT effect with more complex study material.

2

Open Questions and Goals of the Present Dissertation

2.1 Open Questions

Research on the FT effect with more complex study material is still fragmentary and shows a number of empirical gaps, which is why definitive answers on exactly which mechanisms mediate the FT effect are still missing. One of the critical gaps is the question whether relatedness of study materials influences the FT effect with more complex materials, and, for instance, context change mediates the effect with unrelated materials and strategy change or facilitated comprehension with related materials. Examination of the following open research questions can help to clarify the issue:

First, does the presence of the FT effect depend on relatedness of study materials? While the FT effect has repeatedly been demonstrated with related complex materials (Jing et al., 2016; Szpunar et al., 2013, 2014; Wissman et al., 2011), prior studies on the FT effect with unrelated complex materials either employed materials that were not directly related to one another but shared a common topic (Wissman et al., 2011), or they differed in critical ways from other studies on the FT effect by employing selective interim practice instead of nonselective interim practice (Yue et al., 2015), or employing interim semantic generation instead of interim retrieval practice (Divis & Benjamin, 2014). No study has yet examined whether the same basic FT effect is present when highly related versus completely unrelated complex materials are employed with procedural details being held constant. Such comparisons can serve to evaluate generalizability of the effect with more complex study material.

Second, does the FT effect generalize to an FT-like effect in response to semantic generation? In other words, is episodic retrieval of previously studied material required to induce the effect, or does the effect generalize to semantic retrieval of information not related to the study material? While the issue has repeatedly been addressed in word-list studies, only a single experiment in the literature addressed the issue to date employing more complex study material (Divis & Benjamin, 2014). In this study, semantic generation was found to induce an FT-like effect with unrelated prose passages. No study has yet examined whether semantic generation induces an FT-like effect also with related complex materials, and no study has yet directly compared the effects of interim retrieval practice and interim semantic generation for more complex study material. While the context-change account predicts that FT(-like) effects arise in response to both retrieval practice and semantic generation, the strategy-change and facilitated-comprehension accounts predict that the FT effect does not generalize to an FT-like effect in response to semantic generation. Comparison of effects of the two types of retrieval can thus indicate whether qualitatively different FT effects arise for related and unrelated complex materials.

Third, does the FT effect reflect proactive facilitation? In other words, does retrieval practice between study of materials create higher recall of the critical final material than when the final material is studied only? Thus far, only a single experiment on the issue has been reported in the literature, with evidence that interim retrieval practice induced proactive facilitation with related prose passages (Wissman et al., 2011). The finding has not yet been replicated, and no study has yet examined whether interim retrieval practice induces proactive facilitation with unrelated complex materials. While the context-change account predicts that the FT effect does not involve proactive facilitation, both the strategy-change and facilitated-comprehension accounts are open to effects of proactive facilitation in the FT effect. The question of whether the FT effect reflects proactive facilitation can thus also help to clarify whether qualitatively different FT effects arise for related and unrelated complex materials.

Fourth, is the FT effect retrieval specific, or does the effect generalize to an FT-like effect in response to elaborative restudy practice? While prior studies on the FT effect with more complex study material suggest retrieval specificity of the effect (Jing et al., 2016; Szpunar et al., 2013; Yue et al., 2015), the restudy conditions employed for

comparison with the retrieval-practice conditions involved merely plain reexposure to the previously studied materials. No study has yet examined whether retrieval specificity of the FT effect still holds when restudy practice is enriched with elaborative reprocessing, and whether the answer to this question differs for related versus unrelated complex materials. While both the context-change and strategy-change accounts predict that the FT effect is retrieval specific and thus does not generalize from retrieval practice to any type of restudy practice, the facilitated-comprehension account is open to FT-like effects in response to elaborative restudy practice. The question of whether retrieval specificity of the FT effect can vary with relatedness of study materials may thus provide further information on whether qualitatively different FT effects arise for related and unrelated complex materials.

2.2 Goals of the Present Dissertation

The goal of the present dissertation was to examine the potential role of relatedness of study material for the FT effect with prose materials. To this end, two research questions were addressed. First, does the basic FT effect arise with both highly related and completely unrelated prose materials, thus addressing the issue of whether relatedness of materials can influence the presence of the FT effect? And second, can basic characteristics of the FT effect vary with relatedness of prose materials, thus addressing the issue of whether qualitatively different FT effects arise for related and unrelated prose materials? The research questions were addressed in a series of six experiments.

The first set of experiments (Experiments 1–3) had the goal to examine whether an FT(-like) effect is present in response to both retrieval practice and semantic generation, and whether the FT effect reflects proactive facilitation. Critically, it was examined whether the answers to these questions can vary with relatedness of study materials. In all three experiments, a typical version of the FT-effect task was applied (e.g., Pastötter et al., 2011; Szpunar et al., 2008), in which participants studied five prose passages in succession and immediately recalled critical final passage 5, with passage-5 recall serving as the dependent measure. The presence of an FT effect was tested by comparing recall between a retrieval-practice and a restudy condition, manipulating

whether participants between study of the single passages were asked to practice the preceding passage by virtue of retrieval or restudy (cf. Szpunar et al., 2008). The presence of an FT-like effect was tested by comparing recall between a semantic-generation and an arithmetic-tasks condition, manipulating whether participants between study of the single passages were prompted to name exemplars from semantic categories not related to the passages, or to solve simple arithmetic tasks (cf. Divis & Benjamin, 2014). Proactive facilitation was examined by testing whether recall in the retrieval-practice condition is higher than in a condition in which the final passage 5 was studied only (cf. Wissman et al., 2011). To address the potential role of relatedness of study material, the research questions were examined for both completely unrelated prose passages (Experiments 1 and 3) and highly related prose passages (Experiments 2 and 3), keeping procedural details constant to allow for direct comparison of results across the two types of materials.

On the basis of the context-change account of the FT effect, FT(-like) effects should arise in response to both retrieval practice and semantic generation, but the FT effect should not reflect proactive facilitation. In contrast, on the basis of the strategy-change and facilitated-comprehension accounts of the FT effect, the FT effect should not generalize to an FT-like effect in response to semantic generation, but the FT effect may indeed reflect proactive facilitation. In consequence, Experiments 1–3 can serve to contrast the context-change account from the strategy-change and facilitated-comprehension accounts of the FT effect. The first set of experiments was therefore primarily focused on the potential role of context change in the FT effect with prose materials, and whether this role may vary depending on relatedness of study materials.

The second set of experiments (Experiments 4–6) had the goal to examine whether the FT effect is retrieval specific or generalizes to an FT-like effect in response to elaborative restudy practice, and whether the answer to this question can vary with relatedness of study materials. In all three experiments, another typical version of the FT-effect task was applied (e.g., Bäuml & Kliegl, 2013; Kliegl & Bäuml, 2021b), in which participants were now asked to study three prose passages in succession and immediately recall critical final passage 3, with passage-3 recall serving as the dependent measure. Retrieval specificity of the FT effect was tested by comparing the effects of retrieval practice and elaborative restudy practice, manipulating whether participants between

study of the single passages were asked to practice the preceding passage by virtue of retrieval or the creation of a concept map. Recall levels in the two conditions were compared against a restudy condition, in which participants between the passages were asked to simply restudy the preceding passage. Recall levels in the two conditions were also compared against a condition in which participants just studied passage 3, which again permitted a test of proactive facilitation (cf. Experiments 1–3). To address the potential role of relatedness of study material, retrieval specificity was examined for both highly related prose passages (Experiments 4 and 5) and completely unrelated prose passages (Experiment 6), again keeping procedural details constant to allow for direct comparison of results across the two types of materials. Experiments 4 and 5 additionally included a final inference test on passages 1–3, which was intended to measure knowledge integration across the to-be-studied information (e.g., Graesser et al., 1994).

On the basis of the strategy-change and context-change accounts of the FT effect, the FT effect should be retrieval specific and thus not generalize from retrieval practice to any type of restudy practice. In contrast, on the basis of the facilitated-comprehension account of the FT effect, the FT effect may not be retrieval specific and generalize to an FT-like effect in response to elaborative restudy practice, at least if one assumes that both types of practice can promote knowledge integration. In consequence, Experiments 4–6 can serve to contrast the facilitated-comprehension account from the strategy-change and context-change accounts of the FT effect. The second set of experiments was therefore primarily focused on the potential role of facilitated comprehension in the FT effect with prose materials, and whether this role may vary depending on relatedness of study materials.

Across the six experiments, one might have expected fairly different patterns of results, considering the many empirical gaps in the literature discussed above. Previous findings on the FT effect, however, raise the possibility that context change mediates the effect with unrelated prose material and strategy change or facilitated comprehension with related prose material (Divis & Benjamin, 2014; Wissman et al., 2011). If so, an FT-like effect in response to semantic generation should be present with unrelated prose passages only and proactive facilitation be present with related prose passages only. In addition, if facilitated comprehension indeed contributed to the effect with related prose

material, retrieval specificity of the FT effect may also vary with relatedness of prose passages, and the effect be retrieval specific with unrelated but not related passages.

3

Examining the Potential Role of Context Change in the Forward Testing Effect With Prose Materials

The first set of experiments was primarily focused on the potential role of context change in the FT effect with prose materials, and whether this role may vary depending on relatedness of study materials. To this end, highly related or completely unrelated prose passages were employed as study material, and it was examined whether an FT(-like) effect is present in response to both retrieval practice and semantic generation, and whether the FT effect reflects proactive facilitation.

According to the context-change account of the FT effect, both retrieval practice and semantic generation should induce an FT(-like) effect, but both should not induce proactive facilitation. The proposal is that retrieval between study of materials drives mental context change, insulating the final material against buildup of proactive interference from the preceding materials via an enhancement in “list” discrimination (Divis & Benjamin, 2014; Pastötter et al., 2011; Szpunar et al., 2008). Because both retrieval practice and semantic generation can trigger mental context change and thus induce “list” discrimination (Jang & Huber, 2008; Shiffrin, 1970), the account predicts that both types of interim retrieval should improve recall of the final material and therefore produce an FT(-like) effect. However, because enhancements in “list” discrimination supposedly improve recall of the final material by (largely) eliminating the preceding materials from the mental search set (Bäuml & Kliegl, 2013; Sahakyan & Kelley, 2002), the account predicts that recall should at best be enhanced to a level similar but not superior to the recall level when the final material was studied only. In consequence, on the basis of the context-change account, the recall enhancements as induced by the two types of interim retrieval should reflect release from proactive interference but should not reflect proactive facilitation.

According to the strategy-change and facilitated-comprehension accounts of the FT effect, in contrast, the FT effect should not generalize to an FT-like effect in response to semantic generation, but the FT effect may indeed reflect proactive facilitation. The proposal is that retrieval practice between study of materials may allow subjects to gain performance-relevant insights on the basis of which subsequent encoding and retrieval strategies can be optimized (strategy change; Chan, Manley, et al., 2018; Soderstrom & Bjork, 2014), or may (alternatively) increase the accessibility of the preceding materials during encoding of the final material which can lead to better comprehension (facilitated comprehension; Jing et al., 2016; Wissman et al., 2011). Because semantic generation between study of materials should not provide such performance-relevant information (Kliegl & Bäuml, 2021b), and should also not make the preceding materials more readily accessible (Divis & Benjamin, 2014), the two accounts predict that interim retrieval practice but not interim semantic generation should improve recall of the final material and thus induce an FT effect. However, because strategy change and facilitated comprehension are assumed to make the learning more effective, and because both require study and practice of preceding material, the two accounts predict that recall of the final material may be enhanced even above the recall level when the final material was studied only, suggesting that the FT effect may reflect proactive facilitation (Wissman et al., 2011).

3.1 Overview of Experiments 1–3

The results of three experiments are reported aimed to examine the FT effect with highly related and completely unrelated prose passages. In particular, it was examined whether, for the two types of materials, (i) an FT(-like) effect is present in response to both retrieval practice and semantic generation, and (ii) the FT effect reflects proactive facilitation. In Experiments 1 and 2, participants studied five prose passages and, immediately after study of passage 5, were asked to freely recall this critical final passage. Experiment 1 employed passages that were unrelated to one another and covered entirely different topics; Experiment 2 employed passages that were directly related to one another, covered a common topic, and were intended to be integrated. In all other aspects, the two

experiments were identical. After study of passages 1 through 4, participants were either asked to recall the immediately preceding passage (retrieval-practice condition), study the immediately preceding passage once again (restudy condition), generate as many exemplars as possible from given semantic categories (semantic-generation condition), or solve simple arithmetic tasks (arithmetic-tasks condition). Both experiments also contained a condition in which only the final passage 5 was studied (passage-5-only condition). Recall in this condition was taken as a baseline to measure the presence or absence of proactive facilitation (cf. Wissman et al., 2011). Using similar procedure as Experiments 1 and 2, Experiment 3 examined the effect of relatedness of prose passages within a single experiment, including retrieval-practice and restudy conditions as well as the condition in which the final passage 5 was studied only.

On the basis of prior work, retrieval practice was expected to induce an FT effect with both related and unrelated prose passages, with higher recall of final passage 5 in the retrieval-practice than the restudy condition (e.g., Szpunar et al., 2013; Wissman et al., 2011; Yue et al., 2015). In addition, with unrelated passages, semantic generation was expected to induce an FT-like effect, i.e., higher recall in the semantic-generation than the arithmetic-tasks condition (Divis & Benjamin, 2014), and, with related passages, retrieval practice was expected to induce proactive facilitation, i.e., higher recall in the retrieval-practice than the passage-5-only condition (Wissman et al., 2011). Critically, if context change mediated the FT effect with unrelated but not related material, the FT effect should generalize to an FT-like effect in response to semantic generation with unrelated passages only and proactive facilitation in response to retrieval practice be present with related passages only.

3.2 Experiment 1

The goal of Experiment 1 was to examine whether, with unrelated prose passages, an FT(-like) effect is present in response to both retrieval practice and semantic generation, and whether the FT effect for this type of material reflects proactive facilitation. If context change mediated the FT effect with unrelated passages, both retrieval practice and semantic generation should enhance recall of the final passage relative to their (restudy

or arithmetic-tasks) control conditions, indicating an FT effect in response to retrieval practice and an FT-like effect in response to semantic generation. Both types of retrieval, however, should enhance recall to a level (inferior or) similar but not superior to the level in the passage-5-only condition, a finding that would indicate (partial) release from proactive interference but would indicate absence of proactive facilitation.

3.2.1 Method

Ethical Considerations

All reported experiments were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Each participant in Experiments 1, 2, and 3 provided informed consent prior to participation.

Participants

Sample size in Experiments 1, 2, and 3 was determined on the basis of a power analysis using the G*Power program (Version 3.1.9.7; Faul et al., 2007). Setting α to .05 and β to .20, a sample of 26 subjects per condition was recommended to detect the meta-analytic effect size of the FT effect with prose materials ($g = 0.80$; Chan, Meissner, & Davis, 2018). Closely following this recommendation, a total of 140 students (mean age = 23.5 years; 118 female, 22 male) of different German universities were enrolled to take part in Experiment 1, with 28 subjects in each of the five experimental conditions. All participants spoke German as their native language and received monetary reward or course credit for participation. Participants in all three experiments were tested individually in an online video conference with the experimenter, who was present for the entire experimental session and monitored compliance with instructions.

Materials

Five prose passages were employed as study material that were unrelated to one another and covered entirely different topics. Four of the passages were adapted from prior studies: the neandertals (Fritz & Morris, 2015; passage 1), the porcupines (Divis & Benjamin, 2014; passage 2), the chronic wasting disease (Divis & Benjamin, 2014; passage 4), and the history and uses of garlic (Fritz & Morris, 2015, passage 5). Another passage described the story of Johann Beringer's "lying-stones" (passage 3) and was

based on information in Taylor (2017). Each passage was presented in German language and approximately 140 words in length (for the English versions, see [osf.io/dg5n3]). Presentation order of the five passages was the same for all participants. Thus, passage 5 always served as the final critical passage, and passages 1–4 always served as the preceding passages.

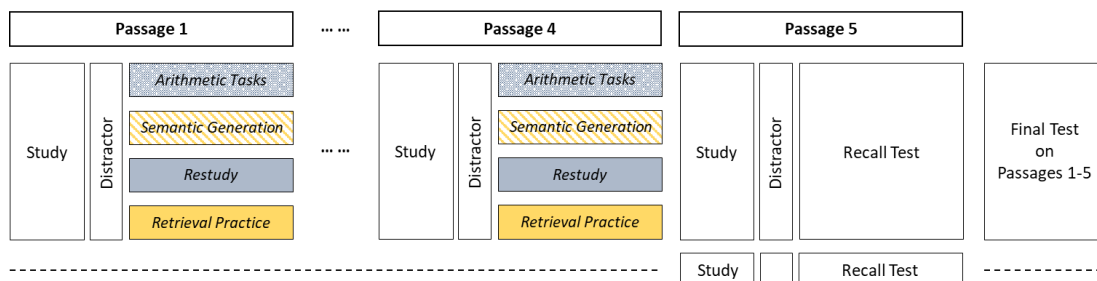
Design

Four of the five conditions followed a 2 x 2 design with the between-subjects factors of practice (present vs. absent) and retrieval (present vs. absent). Practice was present in the restudy and retrieval-practice conditions, in which immediately after study of passages 1–4, participants practiced the preceding passage by virtue of restudy or retrieval practice; practice was absent in the arithmetic-tasks and semantic-generation conditions, in which immediately after study of passages 1–4, participants engaged in arithmetic or semantic generation tasks that did not require practice of the preceding passages. Retrieval was present in the retrieval-practice and semantic-generation conditions, and absent in the restudy and arithmetic-tasks conditions.

Besides these four conditions in which all five passages were presented for study, an additional passage-5-only condition was included in the experiment in which passages 1–4 were not presented prior to study of the critical passage 5. Recall in this condition was used as a baseline to measure whether prior study (and practice) of passages 1–4 improved or impaired passage-5 recall in the other four conditions.

Procedure

Figure 1 depicts the experiment's procedure and conditions. In the restudy, retrieval-practice, arithmetic-tasks, and semantic-generation conditions, participants studied five prose passages in anticipation of a final cumulative test. Prior to the start of the experiment, participants were told to expect various activities that might follow the presentation of each single passage, which could include restudy of a passage that they had just studied, free recall of a passage that they had just studied, solving simple arithmetic tasks, or generating exemplars from semantic categories unrelated to the presented passages. We pretended that activities following each passage were determined on a random basis. In fact, the activities varied between conditions, and participants

Figure 1*Procedure and conditions employed in Experiments 1 and 2*

Note. In four experimental conditions, five prose passages were studied in succession. Each passage was followed by a short distractor task. After study of passages 1 through 4, participants either solved simple arithmetic tasks (arithmetic-tasks condition), generated exemplars from semantic categories unrelated to the passages (semantic-generation condition), restudied the preceding passage (restudy condition), or freely recalled the preceding passage (retrieval-practice condition). Passage 5 was recalled immediately after study. All five passages were then tested in a final recall test. In an additional passage-5-only condition, passage 5 was studied and recalled without prior study of passages 1 through 4.

passed through the same activities after study of passages 1–4 within each experimental condition (cf. Szpunar et al., 2008). Participants were also told that, regardless of the activities that followed the presentation of each single passage, there would definitely be a final recall test on all the studied passages.

All five passages were presented in succession for 3 min each at the center of the computer screen. After presentation of each single passage, participants counted backwards in steps of threes from a three-digit number for 30 s. The type of activity that followed the backward counting after passages 1–4 was conducted for 3 min each and varied with condition. Participants were either asked to (i) study the preceding passage once again, which was reexposed on the computer screen (restudy condition), (ii) type in all the information they could remember from the preceding passage in an empty response box on the screen (retrieval-practice condition), (iii) solve simple arithmetic tasks (arithmetic-tasks condition), or (iv) type in as many words as possible from a given semantic category (semantic-generation condition). In the semantic generation task,

participants after study of each passage were first given 90 s to generate exemplars from one of eight categories (body parts, food, four-legged animals, male names, means of transport, musical instruments, office supplies, or sports), and then were given another 90 s to generate exemplars from another category. Assignment of the eight categories to activities following passages 1–4 was random for each participant.

After presentation of passage 5 and the backward counting, participants in all four conditions were given 3 min to type in all the information they could remember from passage 5 in an empty response box on the screen. After completion of the passage-5 recall test and another 3 min filled with Tetris as distractor task, participants passed through the final recall test in which they had 15 min to type in all the information they could remember from all five passages they had previously studied.

In the passage-5-only condition, passages 1 through 4 were not presented for study. Rather, participants in this condition were asked to just study passage 5 for an upcoming recall test. Presentation of passage 5, backward counting, and the passage-5 recall test were identical in procedure to the other four conditions. There was no final recall test in the passage-5-only condition.

Scoring

To assess recall performance for passages 1–5, all five passages were divided into idea units, and the number of ideas reproduced on the recall protocols was counted to calculate the scores of the single participants. Participants were given 1 point for correctly recalling an entire idea unit, or 0.5 points for correctly recalling an idea unit's critical part. Credit was assigned for verbatim responses and for correct paraphrases of idea units. For example, referring to the passage-5 idea unit "*Sanskrit writings from 5000 years ago refer to garlic and 3000-year-old Chinese manuscripts mention garlic*", the response "*Sanskrit writings from 5000 years ago and Chinese manuscripts from 3000 years ago allude to garlic*" was scored to be worth 1 point, as the response correctly reproduced the entire idea unit. However, the responses "*Sanskrit writings from 5000 years ago allude to garlic*", and "*Chinese manuscripts from 3000 years ago allude to garlic*" were both scored to be worth 0.5 points only, as both responses correctly reproduced critical parts of the idea unit but failed to mention certain aspects of it. Performance on the passage-5 recall test was of greatest interest, and two raters separately scored 70 of the passage-5

protocols. High interrater reliability was obtained, with a Pearson correlation of .99 between their scores. The remaining protocols on passages 1–5 were scored by one rater. Recall performance during practice and on the final test was coded using the identical coding scheme.

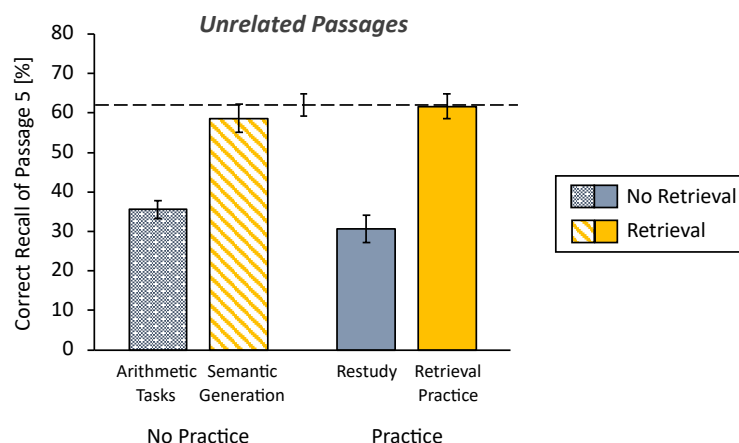
3.2.2 Results

For all experiments, we provided Bayes factors (B_{01}) – which reflect the odds for the null hypothesis over the alternative hypothesis – when the results did not reach conventional level of statistical significance (i.e., $\alpha = .05$) (Masson, 2011; Raftery, 1995). All data are available at [osf.io/dg5n3].

Passage-5 Recall Test

Correct Recall. Figure 2 shows the percentage of correctly recalled passage-5 idea units for each of the five experimental conditions. Regarding the four conditions in which all five passages were presented for study, a 2 x 2 ANOVA with the between-subjects factors of practice (present vs. absent) and retrieval (present vs. absent) showed no significant main effect of practice, $F(1, 108) = 0.082$, $MSE = 0.03$, $p = .775$, $\eta^2 < .01$, $B_{01} = 10.191$, but a main effect of retrieval, $F(1, 108) = 74.431$, $MSE = 0.03$, $p < .001$, $\eta^2 = .41$, reflecting that recall was higher in the presence than the absence of retrieval (60.1% vs. 33.1%). Critically, there was no significant interaction between the two factors, $F(1, 108) = 1.614$, $MSE = 0.03$, $p = .207$, $\eta^2 = .02$, $B_{01} = 4.639$, suggesting that the effect of retrieval did not depend on practice. Consistently, planned comparisons revealed a reliable difference in recall between the retrieval-practice and restudy conditions (61.7% vs. 30.7%), $t(54) = 6.745$, $p < .001$, $d = 1.80$, 95% CI of the difference = [21.8, 40.2], as well as between the semantic-generation and arithmetic-tasks conditions (58.6% vs. 35.6%), $t(54) = 5.413$, $p < .001$, $d = 1.45$, 95% CI of the difference = [14.5, 31.6], thus demonstrating an FT effect in response to retrieval practice and an FT-like effect in response to semantic generation.

Additional comparisons to the passage-5-only condition revealed that recall in this condition was similar to recall in the retrieval-practice and semantic-generation conditions (passage-5 only vs. retrieval practice: 61.8% vs. 61.7%, $t(54) = 0.039$, $p_{adj} =$

Figure 2*Passage-5 Recall in Experiment 1*

Note. Percentage of correctly recalled passage-5 idea units for unrelated passages as a function of condition: arithmetic tasks (no practice, no retrieval), semantic generation (no practice, retrieval), restudy (practice, no retrieval), retrieval practice (practice, retrieval). The dashed line represents mean recall in the passage-5-only condition. Error bars represent standard errors.

.969, $d = 0.01$, 95% CI of the difference = [-8.0, 8.3], $B_{01} = 7.478$; passage-5 only vs. semantic generation: 61.8% vs. 58.6%, $t(54) = 0.726$, $p_{\text{adj}} = .942$, $d = 0.19$, 95% CI of the difference = [-5.7, 12.2], $B_{01} = 5.667$), but higher than in the restudy and arithmetic-tasks conditions (passage-5 only vs. restudy: 61.8% vs. 30.7%, $t(54) = 7.163$, $p_{\text{adj}} = .003$, $d = 1.91$, 95% CI of the difference = [22.4, 39.9]; passage-5 only vs. arithmetic tasks: 61.8% vs. 35.6%, $t(54) = 7.408$, $p_{\text{adj}} = .004$, $d = 1.98$, 95% CI of the difference = [19.2, 33.4]). To control the familywise error rate across the four comparisons, p -values were adjusted by employing the sequential Bonferroni procedure.

Intrusions. Intrusions were defined as idea units from passages 1, 2, 3, or 4 that participants erroneously produced on the passage-5 recall test. Critically, not a single participant had an intrusion from a preceding passage, irrespective of condition.

Table 1

Percentage of Correctly Recalled Idea Units During Practice and on the Final Test for Experiment 1 (Standard Errors in Parenthesis)

Condition	Arithmetic tasks	Semantic generation	Restudy	Retrieval practice
Passage 1				
Practice	---	---	---	42.3 (3.50)
Final test	24.5 (2.82)	28.8 (3.80)	30.0 (3.54)	36.1 (3.32)
Passage 2				
Practice	---	---	---	48.2 (2.86)
Final test	22.9 (2.48)	30.7 (3.15)	30.0 (3.93)	40.4 (3.07)
Passage 3				
Practice	---	---	---	40.3 (3.53)
Final test	17.6 (1.97)	26.0 (2.80)	29.1 (3.20)	35.9 (3.23)
Passage 4				
Practice	---	---	---	54.5 (3.26)
Final test	22.8 (1.76)	32.7 (3.62)	29.5 (3.38)	45.1 (3.93)
Passage 5				
Practice	---	---	---	---
Final test	29.2 (3.02)	46.4 (3.40)	23.7 (3.15)	51.1 (3.82)

Practice Success for Passages 1 to 4

Table 1 shows the percentage of correctly recalled idea units during practice of passages 1–4 in the retrieval-practice condition. Because presentation order of passages was not manipulated in this study, no statistical analyses were carried out to compare practice performance across passages.

Final Test

Table 1 shows recall of passages 1–5 on the final test in the restudy, retrieval-practice, arithmetic-tasks, and semantic-generation conditions. Because presentation order of passages was not manipulated in this study, no statistical analyses were carried out to compare recall across passages, and final test performance was thus pooled for passages 1 through 5 to compare recall across conditions. Pairwise comparisons revealed higher performance in the retrieval-practice than the restudy condition (41.7% vs. 28.4%), $t(54) = 3.191$, $p = .002$, $d = 0.85$, 95% CI of the difference = [4.93, 21.62], and higher performance in the semantic-generation than the arithmetic-tasks condition (32.9% vs. 23.4%), $t(54) = 3.221$, $p = .002$, $d = 0.86$, 95% CI of the difference = [3.60, 15.46]. Thus, with unrelated passages, interim retrieval practice induced a beneficial effect on final recall relative to restudy, and interim semantic generation a beneficial effect relative to the arithmetic-tasks condition.

3.2.3 Discussion

Using unrelated prose passages as study material, the FT effect was present in response to retrieval practice and an FT-like effect was present in response to semantic generation. These results are consistent with Divis and Benjamin's (2014, Experiment 2) finding on the effects of semantic generation and extend on Wissman et al.'s (2011, Experiment 1a) and Yue et al.'s (2015, Experiment 2) work by demonstrating that, also with completely unrelated passages and nonselective interim practice, retrieval practice can induce an FT effect relative to a restudy condition. Recall in the restudy condition was lower than recall in the passage-5-only condition, indicating buildup of proactive interference in the restudy condition, but recall in the retrieval-practice condition was similar to recall in the passage-5-only condition, indicating release from such proactive interference but absence of proactive facilitation in the retrieval-practice condition. The results for the semantic-generation condition mimicked those for the retrieval-practice condition, with recall in this condition being higher than recall in the arithmetic-tasks condition but similar to recall in the passage-5-only condition. These findings are in line with the context-change account of the FT effect, but challenge both the strategy-change and facilitated-comprehension explanations. The results from the final cumulative test mimicked those

from the passage-5 recall test, with both retrieval practice and semantic generation inducing beneficial effects on final recall performance relative to their (restudy or arithmetic-tasks) control conditions (for similar results using word lists as study material, see Pastötter et al., 2011).

Intrusions during recall of the final passage were absent irrespective of condition, leaving no room for reductions in intrusions, be it in response to retrieval practice or semantic generation. Buildup of proactive interference and release from proactive interference were thus reflected in correct recall of the final passage but not in the intrusion data. A possible reason for this pattern may be that, during recall of the final passage, intrusions from the preceding passages came to mind and caused recall impairment. However, because likely such intrusions could quickly be identified as incorrect candidates stemming from the (unrelated) preceding passages, they were not reported in the recall protocols (see Pierce et al., 2017). The results thus add to the list of findings demonstrating that the FT effect in correct recall can be present even if the FT effect in intrusions is absent (for similar results, see Ahn & Chan, 2022; Pierce et al., 2017; Wissman et al., 2011).

3.3 Experiment 2

The goal of Experiment 2 was to examine whether the findings of Experiment 1 with unrelated prose passages generalize to related prose passages. If context change mediated the FT effect with unrelated passages but not with related passages, there is reason to expect that the findings for related passages will differ from the findings for the unrelated passages. In particular, retrieval practice but not semantic generation may now enhance recall of the final passage relative to the respective (restudy or arithmetic-tasks) control condition, indicating that, with related passages, the FT effect does not generalize to an FT-like effect in response to semantic generation. In addition, with this type of material, retrieval practice may enhance recall to a level superior to the level in the passage-5-only condition, a finding that would indicate the presence of proactive facilitation.

3.3.1 Method

Participants

Another 140 students (mean age = 23.8 years; 115 female, 25 male) were enrolled from different German universities to take part in Experiment 2, with 28 subjects in each of the five experimental conditions. Sample size was identical to Experiment 1. All participants spoke German as their native language and received monetary reward or course credit for participation.

Materials, Design, and Procedure

Experimental details were identical to Experiment 1 (see Figure 1), with the only exception of study material. In Experiment 2, the study material comprised five prose passages that were directly related to one another, covered a common topic, and contained information that was intended to be integrated. Specifically, the passages were consecutive sections of an expository text about the big bang theory, which was taken from prior work (Chan et al., 2006) and translated into German (see also Bäuml & Schlichting, 2014, Experiment 2; Wallner & Bäuml, 2017, Experiment 4). Again, each passage was approximately 140 words in length (for the English versions, see [osf.io/dg5n3]). Like in Experiment 1, presentation order was the same for all participants, with passage 5 always serving as the final critical passage and passages 1–4 always serving as the preceding passages.

Scoring

For scoring purposes, all five passages were parsed into idea units, and recall performance for passages 1–5 was assessed following the same criteria that had been used for scoring in Experiment 1. For example, regarding the passage-5 idea unit “*By late 1989 NASA’s Cosmic Background Explorer was ready for action*”, the response “*In 1989 NASA’s Cosmic Background Explorer was ready for use*” was scored to be worth 1 point, as the response correctly reproduced the entire idea unit. However, the responses “*In 1989 NASA’s satellite was ready for use*”, and “*NASA’s Cosmic Background Explorer was ready for use*” were both scored to be worth 0.5 points only, as both responses correctly reproduced only parts of the idea unit. Again, two raters separately scored 70 of the

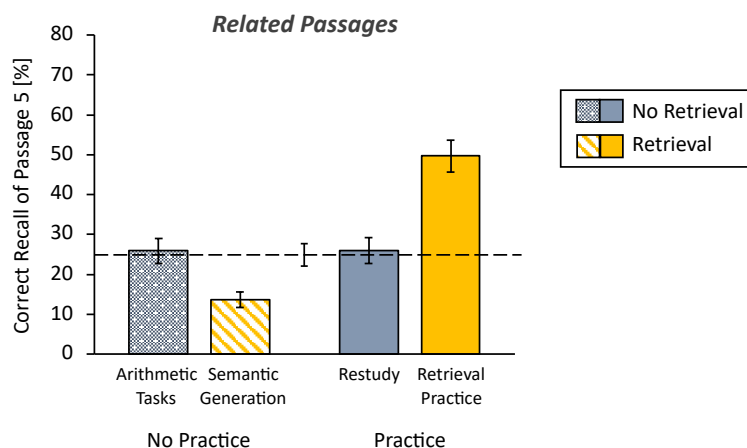
passage-5 protocols, and a Pearson correlation of .99 between their scores was obtained. The remaining protocols on passages 1–5 were scored by one rater. Like in Experiment 1, recall performance during practice and on the final test was coded using the identical coding scheme.

3.3.2 Results

Passage-5 Recall Test

Correct Recall. Figure 3 shows the percentage of correctly recalled passage-5 idea units for each of the five experimental conditions. Regarding the four conditions in which all five passages were studied, a 2 x 2 ANOVA with the between-subjects factors of practice (present vs. absent) and retrieval (present vs. absent) showed a main effect of practice, $F(1, 108) = 32.058$, $MSE = 0.03$, $p < .001$, $\eta^2 = .23$, but no significant main effect of retrieval, $F(1, 108) = 3.221$, $MSE = 0.03$, $p = .075$, $\eta^2 = .03$, $B_{01} = 2.035$, reflecting that, in general, recall was higher when practice was present than when it was absent (37.9% vs. 19.8%), whereas no such difference arose between the presence and absence of retrieval (31.7% vs. 26.0%). Critically, there was also a significant interaction between the two factors, $F(1, 108) = 31.591$, $MSE = 0.03$, $p < .001$, $\eta^2 = .23$, suggesting that the effect of retrieval depended on practice. Indeed, while planned comparisons revealed that recall in the retrieval-practice condition was higher than in the restudy condition (49.7% vs. 26.0%), $t(54) = 4.575$, $p < .001$, $d = 1.22$, 95% CI of the difference = [13.3, 34.0], recall in the semantic-generation condition was even lower than in the arithmetic-tasks condition (13.7% vs. 25.9%), $t(54) = 3.266$, $p = .002$, $d = 0.87$, 95% CI of the difference = [4.7, 19.7]. Thus, there was an FT effect in the retrieval-practice condition, but there was no FT-like effect in the semantic-generation condition.

Additional comparisons to the passage-5-only condition revealed that recall in this condition was similar to recall in the restudy and arithmetic-tasks conditions (passage-5 only vs. restudy: 24.7% vs. 26.0%, $t(54) = 0.300$, $p_{\text{adj}} > .999$, $d = 0.08$, 95% CI of the difference = [-10.2, 7.5], $B_{01} = 7.204$; passage-5 only vs. arithmetic tasks: 24.7% vs. 25.9%, $t(54) = 0.272$, $p_{\text{adj}} = .786$, $d = 0.07$, 95% CI of the difference = [-10.0, 7.6], $B_{01} = 7.200$), but lower than in the retrieval-practice condition (24.7% vs. 49.7%), $t(54) = 4.928$, $p_{\text{adj}} = .004$, $d = 1.32$, 95% CI of the difference = [14.8, 35.2], and higher than in the

Figure 3*Passage-5 Recall in Experiment 2*

Note. Percentage of correctly recalled passage-5 idea units for related passages as a function of condition: arithmetic tasks (no practice, no retrieval), semantic generation (no practice, retrieval), restudy (practice, no retrieval), retrieval practice (practice, retrieval). The dashed line represents mean recall in the passage-5-only condition. Error bars represent standard errors.

semantic-generation condition (24.7% vs. 13.7%), $t(54) = 3.030$, $p_{\text{adj}} = .012$, $d = 0.81$, 95% CI of the difference = [3.7, 18.3]. Like in Experiment 1, the p -values for all four comparisons were adjusted by using the sequential Bonferroni procedure.

Intrusions. Regarding the number of intrusions, a 2 x 2 ANOVA with the factors of practice and retrieval showed no significant main effect of practice, $F(1, 108) = 1.200$, $MSE = 0.60$, $p = .276$, $\eta^2 = .01$, $B_{01} = 5.702$, but a main effect of retrieval, $F(1, 108) = 9.257$, $MSE = 0.60$, $p = .003$, $\eta^2 = .08$, reflecting that participants produced fewer intrusions in the presence than the absence of retrieval (0.13 vs. 0.57). There was also a marginally significant interaction between practice and retrieval, $F(1, 108) = 3.332$, $MSE = 0.60$, $p = .071$, $\eta^2 = .03$, $B_{01} = 1.930$. Planned comparisons revealed a reliable difference in intrusions between the retrieval-practice and restudy conditions (0.07 vs. 0.79), $t(54) = 2.942$, $p = .005$, $d = 0.79$, 95% CI of the difference = [0.23, 1.20], but no reliable difference in intrusions between the semantic-generation and arithmetic-tasks conditions

Table 2

Percentage of Correctly Recalled Idea Units During Practice and on the Final Test for Experiment 2 (Standard Errors in Parenthesis)

Condition	Arithmetic tasks	Semantic generation	Restudy	Retrieval practice
Passage 1				
Practice	---	---	---	33.0 (3.19)
Final test	15.6 (2.16)	14.4 (1.83)	20.9 (3.09)	28.2 (3.69)
Passage 2				
Practice	---	---	---	46.7 (3.76)
Final test	11.2 (2.02)	6.2 (1.56)	17.9 (3.17)	30.4 (3.55)
Passage 3				
Practice	---	---	---	38.3 (3.27)
Final test	14.3 (3.59)	9.9 (1.97)	18.0 (2.81)	34.9 (3.60)
Passage 4				
Practice	---	---	---	43.4 (3.49)
Final test	13.8 (2.45)	9.4 (1.87)	22.9 (3.02)	35.9 (3.38)
Passage 5				
Practice	---	---	---	---
Final test	21.3 (2.91)	12.2 (2.49)	19.3 (2.99)	38.3 (3.11)

(0.18 vs. 0.36), $t(54) = 1.083$, $p = .283$, $d = 0.29$, 95% CI of the difference = [-0.51, 0.15], $B_{01} = 4.100$.

Practice Success for Passages 1 to 4

Table 2 shows the percentage of correctly recalled idea units during practice of passages 1–4 in the retrieval-practice condition. Like in Experiment 1, comparisons of performance

across passages were not carried out because presentation order of passages was not varied between participants.

Final Test

Table 2 shows recall of passages 1–5 on the final test in the restudy, retrieval-practice, arithmetic-tasks, and semantic-generation conditions. Like in Experiment 1, final test performance was pooled for passages 1 through 5 because presentation order of passages was not varied between participants. Pairwise comparisons revealed higher performance in the retrieval-practice than the restudy condition (33.5% vs. 19.8%), $t(54) = 3.715$, $p < .001$, $d = 0.99$, 95% CI of the difference = [6.31, 21.12], but lower performance in the semantic-generation than the arithmetic-tasks condition (10.4% vs. 15.2%), $t(54) = 2.031$, $p = .047$, $d = 0.54$, 95% CI of the difference = [-9.55, -0.06]. Thus, with related passages, interim retrieval practice again induced a beneficial effect on final recall relative to restudy, while interim semantic generation did no longer induce such an effect relative to the arithmetic-tasks condition.

3.3.3 Discussion

Using related prose passages as study material, the FT effect was still present and recall higher in the retrieval-practice than the restudy condition, which is consistent with prior work (Jing et al., 2016; Szpunar et al., 2013, 2014; Wissman et al., 2011). Unlike in Experiment 1, however, recall in the restudy condition was similar to recall in the passage-5-only condition, and recall in the retrieval-practice condition was higher than recall in the passage-5-only condition, indicating the presence of proactive facilitation in the retrieval-practice condition (for a similar finding, see Wissman et al., 2011, Experiment 4). Again unlike in Experiment 1, the results for the semantic-generation condition did not mimic those for the retrieval-practice condition, with recall in this condition being even lower than recall in the arithmetic-tasks and passage-5-only conditions. These results extend on prior findings by indicating that, with highly related passages, the FT effect does not generalize to an FT-like effect in response to semantic generation. These findings disagree with the context-change account and rather point to strategy change or facilitated comprehension as possible mechanisms underlying the FT

effect. The results from the final cumulative test again mimicked those from the passage-5 recall test, with retrieval practice inducing a beneficial effect on final recall performance relative to the restudy condition and semantic generation inducing a detrimental effect relative to the arithmetic-tasks condition.

Intrusions during recall of the final passage were lower in the retrieval-practice than the restudy condition, which is consistent with prior work demonstrating an FT effect in intrusions with related complex material (Szpunar et al., 2013; Wissman et al., 2011). Intrusions, however, were not significantly reduced in the semantic-generation compared to the arithmetic-tasks condition, which may be due to the fact that the number of intrusions was already quite low in the arithmetic-tasks condition, leaving not much room for a reliable reduction in intrusions in response to retrieval activity.

Experiments 1 and 2 differed considerably in recall levels in the passage-5-only condition, which points to major differences in recall difficulty between the two experiments' final materials. To exclude that the observed differences in results between the two experiments were driven by differences in passage-5 recall difficulty, an additional Experiment 3 was conducted in which relatedness of prose passages was now manipulated within a single experiment and final passage 5 held constant across the relatedness conditions.

3.4 Experiment 3

Experiment 3 was conducted to examine whether the differences in results observed between Experiments 1 and 2 were due to differences in passage-5 recall difficulty rather than differences in relatedness of prose passages. To address this possibility, the role of relatedness of passages for the FT effect was now investigated within a single experiment, comparing recall of final passage 5 between a retrieval-practice condition, a restudy condition, and a condition in which passage 5 was studied only. If relatedness of prose passages was indeed the driving force behind the observed differences in results, the findings of Experiments 1 and 2 should generalize to Experiment 3. If so, the FT effect should be present in both the related and unrelated conditions, while proactive facilitation should be present in the related condition but absent in the unrelated condition.

3.4.1 Method

Participants

Another 140 students (mean age = 24.1 years; 106 female, 33 male, 1 diverse) were enrolled from different German universities to take part in Experiment 3, with 28 subjects in each of the five experimental conditions. Sample size followed Experiments 1 and 2. All participants spoke German as their native language and received monetary reward or course credit for participation.

Materials

Like in Experiments 1 and 2, the study material comprised five prose passages. In all experimental conditions, passage 5 used in Experiment 2 (“big bang theory”) was employed as the final critical passage, while passages 1–4 varied between conditions. In the unrelated condition, passages 1–4 used in Experiment 1 (“neandertals”, “porcupines”, “lying-stones”, and “chronic wasting disease”) were employed as the preceding passages; in the related condition, passages 1–4 used in Experiment 2 (“big bang theory”) were employed. Again, presentation order of the five passages was held constant for all participants.

Design

Four of the five conditions followed a 2 x 2 design with the between-subjects factors of type of practice (restudy vs. retrieval practice) and relatedness of passages (unrelated vs. related). In the unrelated condition, participants were presented with five passages that were unrelated to one another and covered different topics, while in the related condition, five passages that were directly related to one another and covered a common topic were presented. Participants were asked after study of passages 1–4 to practice the just presented passage by virtue of restudy or retrieval practice. Besides these four conditions, the experiment again included a passage-5-only condition in which passages 1–4 were not presented and participants were asked to just study passage 5.

Procedure

The procedure differed in two aspects from Experiments 1 and 2. First, in order to avoid floor effects, time for both study and practice were increased to 4 min per passage.

Second, because passage-5 recall was of primary interest, the final recall test was no longer carried out at the end of the experiment.

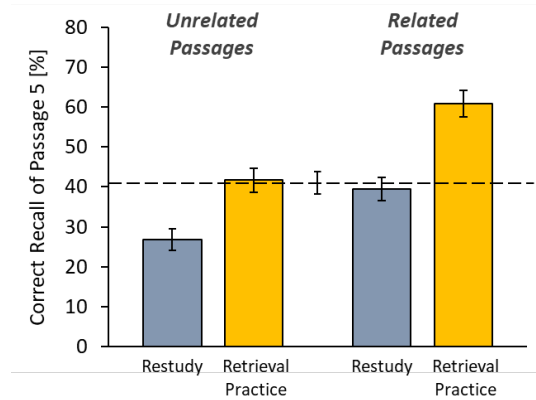
Scoring

Scoring of passage-5 recall responses followed the same criteria that had been used for scoring in Experiment 2. Again, two raters separately scored 70 of the passage-5 protocols, and a Pearson correlation of .99 between their scores was obtained. Given the high level of interrater agreement, the remaining protocols were scored by one rater. Recall success during practice of passages 1–4 was no longer scored and coded in Experiment 3.

3.4.2 Results

Passage-5 Recall Test

Correct Recall. Figure 4 shows the percentage of correctly recalled passage-5 idea units for each of the five experimental conditions. Regarding the four conditions in which five passages were presented for study, a 2 x 2 ANOVA with the between-subjects factors of type of practice (restudy vs. retrieval practice) and relatedness of passages (unrelated vs. related) showed main effects of type of practice, $F(1, 108) = 38.178$, $MSE = 241.60$, $p < .001$, $\eta^2 = .26$, and relatedness of passages, $F(1, 108) = 29.391$, $MSE = 241.60$, $p < .001$, $\eta^2 = .21$, reflecting that recall was higher in the retrieval-practice than the restudy condition (51.3% vs. 33.1%), and higher in the related than the unrelated condition (50.2% vs. 34.2%). Critically, there was no significant interaction between the two factors, $F(1, 108) = 1.240$, $MSE = 241.60$, $p = .268$, $\eta^2 = .01$, $B_{01} = 5.583$, suggesting that the FT effect did not depend on relatedness of passages. Consistently, planned comparisons between the retrieval-practice and restudy conditions revealed reliable FT effects with both the related and unrelated passages (related: 60.9% vs. 39.4%, $t(54) = 4.943$, $p < .001$, $d = 1.32$, 95% CI of the difference = [12.7, 30.1]; unrelated: 41.7% vs. 26.8%, $t(54) = 3.751$, $p < .001$, $d = 1.00$, 95% CI of the difference = [6.9, 22.8]). For both the restudy and retrieval-practice conditions, recall was higher after study and practice of the related than the unrelated passages (restudy: 39.4% vs. 26.8%, $t(54) = 3.264$, $p = .002$,

Figure 4*Passage-5 Recall in Experiment 3*

Note. Percentage of correctly recalled passage-5 idea units as a function of type of practice (restudy, retrieval practice) and relatedness of passages (unrelated passages, related passages). The dashed line represents mean recall in the passage-5-only condition. Error bars represent standard errors.

$d = 0.87$, 95% CI of the difference = [4.8, 20.4]; retrieval practice: 60.9% vs. 41.7%, $t(54) = 4.348$, $p < .001$, $d = 1.16$, 95% CI of the difference = [10.3, 28.0]).

Additional analyses showed that, for the unrelated passages, recall in the passage-5-only condition was similar to recall in the retrieval-practice condition (40.6% vs. 41.7%), $t(54) = 0.244$, $p_{\text{adj}} = .808$, $d = 0.07$, 95% CI of the difference = [-9.6, 7.5], $B_{01} = 7.256$, but higher than in the restudy condition (40.6% vs. 26.8%), $t(54) = 3.422$, $p_{\text{adj}} = .003$, $d = 0.92$, 95% CI of the difference = [5.7, 21.9]. For the related passages, in contrast, recall in the passage-5-only condition was similar to recall in the restudy condition (40.6% vs. 39.4%), $t(54) = 0.284$, $p_{\text{adj}} > .999$, $d = 0.08$, 95% CI of the difference = [-7.2, 9.6], $B_{01} = 7.178$, but lower than in the retrieval-practice condition (40.6% vs. 60.9%), $t(54) = 4.512$, $p_{\text{adj}} = .004$, $d = 1.21$, 95% CI of the difference = [-29.2, -11.2]. For all four comparisons to the passage-5-only condition, p -values were again adjusted following the sequential Bonferroni procedure.

Intrusions. With unrelated passages, none of the participants had an intrusion from a preceding passage, irrespective of practice condition. With related passages, in

contrast, intrusion rates were significantly greater than zero in the two practice conditions ($M = 0.52$), $t(55) = 4.334$, $p < .001$, $d = 0.58$, 95% CI of the difference = [0.28, 0.76], and participants produced fewer intrusions in the retrieval-practice than the restudy condition (0.18 vs. 0.86), $t(54) = 3.045$, $p = .004$, $d = 0.81$, 95% CI of the difference = [0.23, 1.13].

3.4.3 Discussion

The results of Experiment 3 replicated those of Experiments 1 and 2. First, the FT effect was present with both related and unrelated prose passages, as is reflected by the finding that retrieval practice between study of passages enhanced recall of final passage 5 relative to restudy in both the related and unrelated conditions. And second, proactive facilitation was present with related passages but absent with unrelated passages, as is indicated by the finding that only in the related condition, retrieval practice induced recall levels for passage 5 that exceeded those in the passage-5-only condition, while in the unrelated condition, retrieval practice induced recall levels that were indistinguishable from those in the passage-5-only condition. Critically, Experiment 3 was able to replicate the findings of Experiments 1 and 2 even though relatedness of prose passages was now manipulated within a single experiment and final passage 5 held constant across the relatedness conditions. The results thus suggest that the differences in findings observed between Experiments 1 and 2 were driven by differences in relatedness of prose passages rather than differences in passage-5 recall difficulty. Consistent with the findings on intrusions in Experiments 1 and 2, intrusions during recall of final passage 5 were absent in the unrelated condition, while an FT effect in intrusions was present in the related condition, reflecting fewer intrusions in the retrieval-practice than the restudy condition.

3.5 Discussion of Experiments 1–3

The first set of experiments demonstrates that retrieval practice can induce an FT effect with both related and unrelated prose passages. In fact, for both types of passages, retrieval practice on the previously encoded materials created higher recall levels on the final critical material than restudy of the preceding materials did (Experiments 1–3). At

the same time, two differences in findings emerged between the two types of materials. The one difference was that the FT effect generalized to an FT-like effect in response to semantic generation with unrelated passages but not with related passages (Experiments 1 and 2). The other difference was that retrieval practice induced a recall level for the final critical passage that was similar to the recall level in the passage-5-only condition with unrelated prose materials (Experiments 1 and 3) but above this level with related prose materials (Experiments 2 and 3), with only the latter indicating the presence of proactive facilitation.

The differences in findings between related and unrelated materials suggest that the role of context change in the FT effect with prose materials may critically depend on relatedness of study materials. According to the context-change account of the FT effect, the FT effect reflects an enhancement in “list” discrimination, assuming that retrieval between study of materials drives mental context change and thus can insulate the final material against buildup of proactive interference from the preceding materials at test (Divis & Benjamin, 2014; Pastötter et al., 2011; Szpunar et al., 2008). Because both retrieval practice and semantic generation can trigger mental context change and thus induce “list” discrimination (Jang & Huber, 2008; Shiffrin, 1970), the account predicts that an FT(-like) effect should be present in response to both retrieval practice and semantic generation. However, because enhancements in “list” discrimination supposedly improve recall of the final material by (largely) eliminating the previously studied materials from the mental search set (Bäuml & Kliegl, 2013; Sahakyan & Kelley, 2002), the account predicts that FT(-like) effects should reflect release from proactive interference but should not reflect proactive facilitation. The results for the unrelated passages are in line with these predictions, revealing recall improvements for the final material in response to both retrieval practice and semantic generation and recall levels in the two conditions that were similar to the level in the passage-5-only condition. In contrast, the results for the related passages disagree with these predictions, showing that the recall improvement for the final material did not generalize from retrieval practice to semantic generation and the recall level in the retrieval-practice condition exceeded that in the passage-5-only condition. The pattern of results observed across the present three experiments thus may indicate a critical role of context change in the FT effect with unrelated prose materials but not with related prose materials.

The differences in findings do not only suggest a role of context change in the FT effect with unrelated prose materials, but may also point to a role of strategy change or facilitated comprehension in the FT effect with related prose materials. The strategy-change account of the FT effect assumes that retrieval practice between study of materials can provide subjects with performance-relevant information, on the basis of which they may switch to more efficient strategies for subsequent encoding and retrieval (Chan, Manley, et al., 2018; Soderstrom & Bjork, 2014). Alternatively, the facilitated-comprehension account of the FT effect assumes that interim retrieval practice can make the preceding materials more readily accessible during encoding of the final material which may allow for better comprehension (Jing et al., 2016; Wissman et al., 2011). Because semantic generation should not provide performance-relevant insights into the learning task (Kliegl & Bäuml, 2021b), and should also not make the previously studied material more readily accessible (Divis & Benjamin, 2014), both accounts predict that the FT effect should not generalize to an FT-like effect in response to semantic generation. In addition, because strategy change and facilitated comprehension supposedly increase the effectiveness of learning, and because both require the study and practice of preceding material, both accounts are open to effects of proactive facilitation (Wissman et al., 2011). The results for the related passages are in line with these predictions, showing a difference in results between retrieval practice and semantic generation and higher recall of the final material in the retrieval-practice than the passage-5-only condition. In contrast, the results for the unrelated passages do not align with these predictions, revealing an equivalence of effects of retrieval practice and semantic generation and a recall level in the retrieval-practice condition that did not exceed the level of recall in the passage-5-only condition. The present results thus may point to a critical role of strategy change or facilitated comprehension in the FT effect with related prose materials but not with unrelated prose materials. Since the results observed for the related passages can be reconciled with both the strategy-change and facilitated-comprehension explanations, they do not provide specific support for one of the two accounts of the FT effect, however.

4

Examining the Potential Role of Facilitated Comprehension in the Forward Testing Effect With Prose Materials

The second set of experiments was primarily focused on the potential role of facilitated comprehension in the FT effect with prose materials, and whether this role may vary depending on relatedness of study materials. To this end, highly related or completely unrelated prose passages were employed as study material, and it was examined whether the FT effect is retrieval specific or generalizes to an FT-like effect in response to elaborative restudy practice.

According to the strategy-change and context-change accounts of the FT effect, the FT effect should be retrieval specific and thus not generalize from retrieval practice to any type of restudy practice. The one proposal is that retrieval practice between study of materials allows subjects to gain performance-relevant insights – for instance, into retrieval failures and the format of forthcoming recall tests, which may induce them to apply new and more efficient strategies during subsequent encoding and retrieval (strategy change; Chan, Manley, et al., 2018; Soderstrom & Bjork, 2014; Wissman et al., 2011). Because restudy practice does not provide such performance-relevant information on the learning task at hand, retrieval practice but not restudy practice should promote optimization of subsequent encoding and retrieval strategies (e.g., Pyc & Rawson, 2012). The other proposal is that retrieval between study of materials drives mental context change and thus enhances “list” discrimination, which may induce release from proactive interference due to a more focused memory search at test (context change; Divis & Benjamin, 2014; Pastötter et al., 2011; Szpunar et al., 2008). Because restudy practice does not speed mental context change and thus does also not induce “list” discrimination, retrieval practice but not restudy practice should reduce the mental search set size at test

and induce release from proactive interference (e.g., Bäuml & Kliegl, 2013). In consequence, both the strategy-change and context-change accounts of the FT effect predict that retrieval practice but not restudy practice in-between the study of materials should improve recall of the final material and thus induce an FT effect.

The facilitated-comprehension account of the FT effect, in contrast, does not allow clearcut a-priori predictions about retrieval specificity of the FT effect. However, because the account assumes that the FT effect is mediated by knowledge integration (e.g., Jing et al., 2016), the facilitated-comprehension account is open to FT-like effects in response to certain types of restudy practice. The proposal is that interim retrieval practice increases the accessibility of previously studied material during subsequent encoding of new material, which may aid the integration of knowledge deriving from different parts of the study material and thus facilitate comprehension of the new material (Jing et al., 2016; Wissman et al., 2011; see also Wahlheim, 2015). Following this rationale, the FT effect as induced by facilitated comprehension may resemble knowledge effects on learning, and thus not arise in response to retrieval practice only but generalize to an FT-like effect in response to elaborative restudy practice. If so, not only retrieval practice on preceding material may improve recall of the final material, but also elaborative restudy practice between study of the materials. Such a prediction, however, would need support from two lines of findings in the memory literature: the one indicating that both retrieval practice and elaborative studying can promote knowledge acquisition from previously studied information, and the other indicating that prior knowledge is indeed effective at potentiating new learning. The following paragraphs summarize the literature on these issues.

4.1 The Potentiating Effects of Knowledge on New Learning

A wealth of research has already demonstrated that prior knowledge can have potentiating effects on new learning (for reviews, see Ericsson et al., 2006; Gobet, 2015; Vicente & Wang, 1998). For instance, previous studies employing educational texts as study material have found that subjects with high prior knowledge in a given text's domain more readily acquire information from this text than subjects with low prior knowledge

(e.g., Alexander et al., 1994a, 1994b; Boscolo & Mason, 2003; Buchin & Mulligan, 2023; McNamara et al., 1996) – a finding that has been termed the rich-get-richer effect (Witherby & Carpenter, 2022). To illustrate, McNamara et al. (1996) had students study a prose passage on heart disease. The students were classified as high- versus low-knowledge learners based on their performance on a prior-knowledge pretest about the human heart. Learning was measured by free recall as well as by inference questions assessing deeper comprehension of the studied passage. Results showed that high-knowledge learners outperformed low-knowledge learners on the two types of measures (for similar results, see Alexander et al., 1994a, 1994b; Boscolo & Mason, 2003; Buchin & Mulligan, 2023), but the effects of prior knowledge were more pronounced when the students studied less versus more coherent versions of the prose passage – with coherence referring to the extent to which a text leaves the learner with gaps in comprehension (for similar results, see McNamara, 2001; McNamara & Kintsch, 1996).

The potentiating effects of knowledge on new learning have been attributed to an integration-based mechanism, assuming that prior knowledge in a certain text's domain allows the learner to fill comprehension gaps during reading. According to the text-comprehension model by Kintsch (1988, 1998), a learner's mental representation of a text comprises different levels of understanding, including the textbase and the situation model. The proposal is that the textbase relies on information that can be directly derived from the text, while the situation model relies on integration of text information with the learner's prior knowledge base. Given that texts rarely provide all the information needed for ready comprehension, the active use of prior knowledge is required, allowing the learner to make inferences on how the single parts of a text relate together. To illustrate, a text on heart disease might state: "When a baby has a septal defect, the blood cannot get rid of enough carbon dioxide through the lungs." (cf. Kintsch, 1994). Lacking any situation model, a learner knows that having a septal defect causes problems with removing the carbon dioxide – this is what can be derived from the text. But in order to understand how the two parts of the sentence relate together, the learner would need to know that (i) blood travels through the heart just before it goes to the lungs, (ii) blood travels back through the heart after it has been cleansed of carbon dioxide in the lungs, and (iii) the septum is a wall in the heart that separates the blood travelling to the lungs from blood that has already gone through the lungs. Relating this knowledge to the

sentence stated in the text may allow the inference that some blood entering the heart will, rather than travelling to the lungs, pass instead through the defective septum and mix with the blood that has already gone through the lungs, which causes the blood to not be fully ribbed of carbon dioxide. The result of such inferencing is a situation model, which would enable the learner to bridge the comprehension gap encountered in the text's discourse (for a similar illustration, see Bobbitt, 2018, pp. 6–7; see also Kintsch, 1994).

One caveat regarding the potentiating effects of knowledge on new learning is that subjects naturally differ in their prior knowledge in a certain text's domain. As the gap-filling process can only be successful for the knowledgeable learners, poor learning outcomes may result when requisite prior knowledge lacks (e.g., McNamara et al., 1996). Of relevance for present purposes, knowledge requisite for text comprehension may not only derive from a subject's prior knowledge base. Contents discussed in textbook chapters, for instance, are often directly related and build off one another, such that information discussed in earlier parts of a chapter may enable learners to fill gaps in comprehension encountered in the later parts of the chapter. Thus, whenever a learning activity is effective in promoting one's knowledge on the earlier parts, it may also be effective in potentiating learning of the subsequently studied, later parts. If so, and if the FT effect with prose materials was indeed mediated by facilitated comprehension, the effect for this type of material may not be retrieval specific but, in principle, arise after any kind of learning activity that effectively boosts knowledge acquisition from the previously studied information. The primary goal here was to test this hypothesis by directly comparing the effects of interim retrieval practice to the effects of interim elaborative studying with concept mapping, both of which have been prescribed as effective learning tools when faced with the study of educational materials.

4.2 Retrieval Practice and Elaborative Studying With Concept Mapping as Knowledge-Acquisition Tools

Prior research on the backward testing effect has repeatedly demonstrated that retrieval practice, compared to restudy, of previously studied material can yield mnemonic benefits for the practiced material (for reviews, see Agarwal et al., 2021; Karpicke, 2017; Yang et

al., 2021). With more complex materials, like prose passages or video lectures, the effect has not only been found on measures of direct retention (e.g., free recall; Roediger & Karpicke, 2006b), but also on measures of deeper comprehension (e.g., inference questions; Butler, 2010; Hinze et al., 2013; Karpicke & Blunt, 2011; McDaniel et al., 2009), suggesting that retrieval practice can indeed promote one's knowledge on the previously studied material. Previous research on the FT effect, in contrast, has examined the effect on measures of direct retention but rarely on measures of deeper comprehension (for reviews, see Chan, Meissner, & Davis, 2018; Pastötter & Bäuml, 2014; Yang et al., 2018). In consequence, it is far from clear whether retrieval-practice-induced knowledge on previously studied material can facilitate comprehension of the subsequently studied material.

In one study, Jing et al. (2016) had subjects study a video lecture on public health, which was divided into several parts containing highly related information. Retrieval practice between study of the single parts was not only found to improve recall of the critical final part relative to a restudy condition, reflecting an FT effect, but also to improve integration of related information within and across the different parts of the lecture (for similar results using prose passages, see Wissman & Rawson, 2015). The authors proposed that interim retrieval practice may make the practiced material more readily accessible during study of the new material, which may foster knowledge integration across the different parts of the study material and thus facilitate comprehension of the final material (see also Wahlheim, 2015). However, because knowledge integration in this study was (partially) measured by the degree to which recall during a final test was clustered by lecture parts – and increased clustering in recall has been suggested to be indicative for strategy change (Chan, Manley, et al., 2018; Chan et al., 2020; Kliegl & Bäuml, 2021b; Yang et al., 2022) – the Jing et al.'s findings appear also consistent with the strategy-change account of the FT effect and thus do not provide specific support for a critical role of facilitated comprehension in the FT effect. In another study, Wissman et al. (2011) demonstrated that the FT effect with related prose passages improved recall of a critical final passage relative to a condition in which the final passage was studied only (for similar results, see present Experiments 2 and 3), suggesting that retrieval-practice-induced knowledge on the preceding passages potentiated the learning of the final passage. Again, however, the finding of such proactive facilitation can be

reconciled with both the facilitated-comprehension and strategy-change explanations of the FT effect, both of which suggest that learning of the final material is more effective in the presence than absence of prior study and practice of preceding materials.

Besides retrieval practice, another learning activity frequently used in educational settings is elaborative studying with concept mapping. Concept mapping refers to the creation of node-link-node diagrams that represent key terms and relations among the terms within a set of materials (Chevron, 2014; Novak & Cañas, 2008; Novak & Gowin, 1984). Concept mapping can be used for a variety of purposes, including brainstorming, outlining, and as a knowledge-assessment tool. When applied as a tool for elaborative studying, concept mapping is thought to aid knowledge integration, requiring relational processing of the to-be-studied material. In line with this suggestion, prior research employing more complex study material has shown that concept mapping can promote knowledge acquisition, compared to, for instance, simply listening to the study material (for reviews, see Nesbit & Adesope, 2006; Schroeder et al., 2018). To the best of my knowledge, no study has yet examined whether concept mapping can also promote the learning of material studied after concept mapping on the preceding materials has taken place. Because in educational practice, concept mapping seems to be much more popular than retrieval practice (e.g., Hay et al., 2008; Quinn et al., 2003), and students are frequently reluctant to use learning activities they do not prefer (Miyatsu et al., 2018; Rea et al., 2022), the issue of whether the FT effect generalizes to an FT-like effect in response to concept mapping may not only be of theoretical interest, but also have important practical implications.

4.3 The Potential Role of Relatedness of Study Materials

Returning to the textbook-chapter example raised above, practice-induced knowledge on a chapter's earlier parts may enable learners to fill in the information necessary for comprehending subsequently studied, later parts. If so, not only retrieval practice on previously studied material may afford facilitated comprehension of subsequently studied material, but also elaborative restudy practice with concept mapping, which may result in FT(-like) effects that resemble effects of prior knowledge on new learning. Critically,

because knowledge effects on learning are typically restricted to information related to the domain of prior knowledge (e.g., Hunt & Rawson, 2011; Witherby & Carpenter, 2022), practice-induced knowledge on related material should be required to enable facilitated comprehension of the subsequent material. Following this rationale, changing study material from related to unrelated material should eliminate any effects of facilitated comprehension, regardless of whether they were induced by retrieval practice or concept mapping. In consequence, if the FT-like effect induced by concept mapping was tied to facilitated comprehension, but the FT effect with unrelated materials was mediated by context change (cf. present Experiments 1 and 3), retrieval specificity of the FT effect may depend on relatedness of study materials, and the effect be retrieval specific with unrelated but not with related prose materials.

4.4 Overview of Experiments 4–6

The results of three experiments are reported aimed to test retrieval specificity of the FT effect with highly related and completely unrelated prose passages. In particular, it was examined whether, for the two types of materials, similar effects on learning can arise when, instead of interim retrieval practice, elaborative restudy practice with concept mapping is applied as interim learning activity. In all three experiments, participants were asked to study three prose passages and, immediately after study of passage 3, took a free-recall test on this critical final passage. Experiments 4 and 5 employed different materials, but within each experiment the passages were highly related in content and built off one another. Experiment 6 employed passages that were unrelated to one another and had no overlapping content. After study of passages 1 and 2, participants in each experiment were either asked to study the preceding passage again (restudy condition), freely recall the preceding passage (retrieval-practice condition), or create a concept map about the preceding passage (concept-mapping condition). In Experiments 4 and 5, participants in all three conditions took a final inference test on passages 1–3, which is thought to measure knowledge integration (e.g., Graesser et al., 1994; Kintsch, 1994; Wiley et al., 2005). All three experiments additionally included a condition in which participants were asked to just study passage 3 (passage-3-only condition). Inclusion of such condition

permits a test of proactive facilitation (cf. Wissman et al., 2011) and can be further indicative of induced facilitated comprehension.

On the basis of previous findings, retrieval practice was expected to induce an FT effect with both related and unrelated prose passages, with higher recall of final passage 3 in the retrieval-practice than the restudy condition (Wissman et al., 2011; present Experiments 1–3). Only with related passages, however, was retrieval practice expected to induce proactive facilitation – i.e., higher recall in the retrieval-practice than the passage-3-only condition (Wissman et al., 2011; present Experiments 1–3). In order to more directly address the role of facilitated comprehension in the FT effect, it was examined whether interim retrieval practice enhances knowledge integration across related passages, as measured by the final inference test. Because interim retrieval practice has previously been shown to foster integration across the to-be-studied information (Jing et al., 2016; Wahlheim, 2015), improved performance on the final inference test was expected in the retrieval-practice relative to the restudy condition.

Of central interest here, it was examined whether retrieval specificity of the FT effect can vary with relatedness of prose passages. Two expectations stand out, depending on whether facilitated comprehension or strategy change is assumed to underlie the FT effect with related prose materials. If strategy change mediated the FT effect with related material and context change with unrelated material, retrieval specificity of the FT effect should not vary with relatedness of prose passages, and the effect be retrieval specific with both related and unrelated passages. Indeed, retrieval practice but not restudy practice should induce strategy optimization (e.g., Chan, Manley, et al., 2018; Pyc & Rawson, 2012), and retrieval practice but not restudy practice should induce mental context change (e.g., Bäuml & Kliegl, 2013; Pastötter et al., 2011), suggesting that the FT effect should not generalize to an FT-like effect in response to concept mapping with the two types of materials.

In contrast, if facilitated comprehension mediated the FT effect with related material and context change with unrelated material, retrieval specificity of the FT effect may vary with relatedness of prose passages, and the effect be retrieval specific with unrelated but not with related passages. Indeed, because concept mapping is thought to aid knowledge integration by promoting relational processing of practiced material (e.g., Novak & Cañas, 2008; Novak & Gowin, 1984), concept mapping on previously studied

material may afford facilitated comprehension of subsequently studied, related material, very similar to what retrieval practice may achieve by making the practiced material more accessible. If so, with related passages, the two learning activities may induce parallel memorial effects, and interim concept mapping thus not only improve performance on the final inference test, but also induce an FT-like effect and proactive facilitation. However, because knowledge effects on learning typically arise for domain-related material only (e.g., Hunt & Rawson, 2011; Witherby & Carpenter, 2022), neither retrieval practice nor concept mapping on preceding material should afford facilitated comprehension of subsequent material if the study material is changed to unrelated material. In consequence, if the FT-like effect induced by concept mapping was tied to facilitated comprehension, but the FT effect with unrelated passages was mediated by context change, concept mapping may no longer induce an FT-like effect with unrelated passages, and only retrieval practice induce an FT effect with this type of material.

4.5 Experiment 4

The goal of Experiment 4 was to examine whether the FT effect is retrieval specific with related prose passages. To this end, it was examined whether, for this type of material, recall of the final passage is enhanced to a similar degree when, instead of retrieval practice, elaborative restudy practice with concept mapping is applied as interim learning activity. If strategy change mediated the FT effect with related passages, only retrieval practice should enhance recall of the final passage, while concept mapping should leave recall of the final passage unaffected. Alternatively, if facilitated comprehension mediated the FT effect with related passages, both types of learning activity may not only improve performance on the final inference test, but may also enhance recall of the final passage relative to the restudy and passage-3-only conditions, with the former indicating FT(-like) effects and the latter indicating proactive facilitation.

4.5.1 Method

Ethical Considerations

All reported experiments were carried out in accordance with the provisions of the World Medical Association's Declaration of Helsinki. Each participant in Experiments 4, 5, and 6 provided informed consent prior to participation.

Participants

Sample size in Experiments 4, 5, and 6 was determined on the basis of a power analysis using the G*Power program (Version 3.1.9.7; Faul et al., 2007). Setting α to .05 and β to .20, a sample of 26 subjects per condition was recommended to detect the meta-analytic effect size of the FT effect with prose materials ($g = 0.80$; Chan, Meissner, & Davis, 2018). Closely following this recommendation, a total of 112 students (mean age = 24.4 years; 85 female, 27 male) of different German universities were enrolled to take part in Experiment 4, with 28 subjects in each of the four experimental conditions. All participants spoke German as their native language and received monetary reward or course credit for participation. Participants in all three experiments were tested individually in an online video conference with the experimenter, who was present for the entire experimental session and monitored compliance with instructions.

Materials

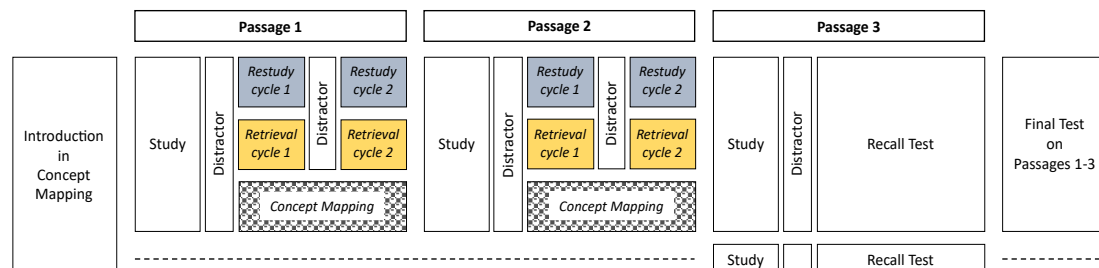
The study material comprised three prose passages that were directly related to one another, covered a common topic, and contained information that was intended to be integrated. The passages were consecutive sections of an expository text about the Earth's internal structure (e.g., crust, mantle, core) and tectonic processes (e.g., continental drift, subduction, earthquakes), which was developed based on information described in Frisch and Meschede (2021) (for the English version of the text, see [osf.io/bckyn]). Each passage was presented in German language and approximately 180 words in length. Presentation order of the three passages was the same for all participants. Thus, passage 3 always served as the final critical passage, and passages 1 and 2 always served as the preceding passages.

The final test was composed of 10 short-answer questions referring to the three prose passages' content (for the English version of the questions, see [osf.io/bckyn]). All of the questions were bridging-inference questions that required subjects to integrate information stated in two or more sentences, within or across passages, to infer the correct answer. One example of an inference question is "*Where does the oceanic crust show the highest ages?*", with "*In the subduction zone*" being the correct answer. In order to answer the question, the following sentences from the text must be understood and linked together: "*Oceanic crust is constantly being formed. This occurs along the mid-ocean ridges [...] In the process, the older crust is pushed away from the ridges on either side. Around the Pacific, oceanic crust submerges beneath the adjacent continental crust [...] Due to the compression, mountain ranges emerge above the subduction zone.*" The Cronbach's alpha estimate of internal consistency reached .66 for the 10 inference questions, reflecting an acceptable reliability (cf. Schmitt, 1996).

Design and Procedure

The experiment comprised four conditions which were manipulated between participants: restudy, retrieval practice, concept mapping, and passage-3 only (for an illustration of the experiment's procedure and conditions, see Figure 5). In all four conditions, participants received a brief instruction about the nature of concept mapping prior to the start of the experiment, adopting the procedure outlined in previous research (Blunt & Karpicke, 2014; Karpicke & Blunt, 2011). Participants were told that concept maps are diagrams where the terms in a set of material are represented as nodes and relations among the terms as lines linking the nodes together. Participants were also shown an example of a concept map depicting properties of water (e.g., liquid, solid, gaseous) and stages of the water cycle (e.g., evaporation, condensation). The example concept map was chosen to represent scientific content of common knowledge without directly relating to information contained in the study text (for the English version of the example concept map, see [osf.io/bckyn]). After completion of this initial phase of instruction, all participants started to pass through the experiment depending on the experimental condition.

In the restudy, retrieval-practice, and concept-mapping conditions, participants studied three prose passages in anticipation of a final test. Prior to the study phase,

Figure 5*Procedure and conditions employed in Experiments 4 and 5*

Note. The experimental session started with an introduction in concept mapping, irrespective of condition. Afterwards, in three conditions, three prose passages were studied in succession. Each passage was followed by a short distractor task. After study of passages 1 and 2, participants either engaged in two restudy cycles on the preceding passage (restudy condition), two retrieval cycles on the preceding passage (retrieval-practice condition), or constructed a concept map about the preceding passage (concept-mapping condition). Passage 3 was recalled immediately after study. Participants in all three conditions then took a final test on the three passages. In an additional passage-3-only condition, passage 3 was studied and recalled without prior study of passages 1 and 2.

participants were told to expect various activities that might follow the presentation of each of the three passages: restudy of a passage that they had just studied, free recall of a passage that they had just studied, or creating a concept map about a just studied passage. It was pretended that the activities following the passages were determined on a random basis when, in fact, the activities varied between conditions. In particular, participants passed through the same activities after study of passages 1 and 2 within each experimental condition. Participants were also informed that, regardless of the activities that followed the presentation of the three passages, there would definitely be a final test on all the studied passages.

The three passages were presented one after another for 6 min each at the center of the computer screen. After presentation of each single passage, participants counted backwards in steps of threes from a three-digit number for 30 s. The activities that followed the backward counting after passages 1 and 2 lasted a total of 12 min each and

depended on the experimental condition. Participants were either asked to (i) restudy the preceding passage during two 5-min cycles of repeated presentation, separated by 2 min filled with simple arithmetic tasks (restudy condition), (ii) recall the preceding passage during two 5-min cycles by typing in all the information they could remember in an empty response box on the screen, separated by 2 min filled with simple arithmetic tasks (retrieval-practice condition), or (iii) create a concept map about the preceding passage (concept-mapping condition). In the concept mapping task, participants were given 12 min to construct a concept map on a blank sheet of paper including as much information as possible from the text passage. Participants were allowed to refer to the passage and the example concept map from the initial phase, which were both provided on the screen for the entire 12-min period. Once the time was up, participants were asked to turn their sheet of paper over so that the created map could no longer be seen.

After presentation of passage 3 and the backward counting task, participants in all three conditions were given 5 min to type in all the information they could remember from passage 3 in an empty response box on the screen. After completion of the passage-3 recall test and another 3 min filled with Tetris as distractor task, participants passed through the final test comprising 10 short-answer questions on passages 1–3. The questions were presented one at a time on the screen in random order, and participants were asked to type in their answers and work through the questions in a self-paced manner. Participants were not allowed to go back to a previous question once they proceeded to the next question. All participants were able to complete the final test within a 15-min duration.

In the passage-3-only condition, participants were asked to just study passage 3 for an upcoming recall test, without prior study of passages 1 and 2. Presentation of passage 3, backward counting, and the passage-3 recall test were identical in procedure to the other three conditions. Recall in this condition was used as a baseline to measure whether prior study and practice of passages 1 and 2 improved or impaired recall of passage 3 in the restudy, retrieval-practice, and concept-mapping conditions. There was no final test on passages 1–3 in the passage-3-only condition.

Scoring

To assess performance during practice and on the passage-3 recall test, passages 1–3 were parsed into idea units, and the number of ideas reproduced on the recall protocols and concept maps was counted to calculate the scores of the single participants. Participants were given 1 point for correctly producing an entire idea unit, or 0.5 points for correctly producing an idea unit's critical part. Credit was assigned for verbatim responses and for correct paraphrases of idea units. For example, at passage-1 practice, the response "*With growing depth, the pressure and temperature in the Earth's interior increase*" was scored to be worth 1 point, as the response correctly reproduced the idea unit "*With the increase in depth, the pressure and temperature in the Earth's interior rise*". However, the responses "*With the increase in depth, the pressure in the Earth's interior rises*", and "*The pressure and temperature in the Earth's interior rise*" were both scored to be worth 0.5 points only, because both responses failed to mention certain details of the idea unit but correctly incorporated the idea's critical information. Passage-3 recall was of greatest interest, and two raters separately scored 56 of the passage-3 protocols. High interrater reliability was obtained, with a Pearson correlation of .99 between their scores. The remaining protocols on passages 1–3 and the concept maps were scored by one rater. Recall protocols and concept maps were coded using the identical coding scheme.

To assess performance on the final test, the number of questions correctly answered was counted for each participant. All 10 questions were scored by awarding 1 point for completely correct answers and 0.5 points for partially correct answers, resulting in a maximum score of 10 possible points. Raters were provided with a coding scheme that identified the key information a response had to contain to be scored as completely or partially correct. For example, referring to the question "*Where does the oceanic crust show the highest ages?*", the responses "*In the subduction zone*" and "*Around the Pacific*" were both considered completely correct and assigned 1 point, whereas the response "*Furthest from the mid-ocean ridges*" was considered partially correct and assigned 0.5 points only. Two raters were blind to condition and separately scored all 84 final test protocols. High interrater reliability was obtained, with a Pearson correlation of .99 between their scores.

4.5.2 Results

For all experiments, we again provided Bayes factors (B_{01}) – which reflect the odds for the null hypothesis over the alternative hypothesis – when the results did not reach conventional level of statistical significance (i.e., $\alpha = .05$) (Masson, 2011; Raftery, 1995). All data are available at [osf.io/bckyn].

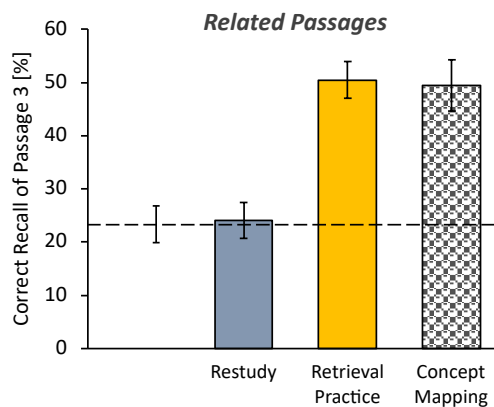
Practice Success for Passages 1 and 2

Table 3 shows the percentage of idea units from passages 1 and 2 that were either recalled during the first and second cycle of retrieval practice or represented on the concept maps. Because presentation order of passages was not manipulated in this study, we did not carry out comparisons across the single passages, and practice performance for passages 1 and 2 was thus pooled for statistical analyses. Regarding performance in the retrieval-practice condition, pairwise comparisons between the first and second cycles of retrieval practice revealed that recall remained relatively stable across cycles (54.5% vs. 56.1%), $t(27) = 1.349$, $p = .189$, $d = 0.26$, 95% CI of the difference = [-3.9, 0.8], $B_{01} = 0.462$. Practice performance, however, differed reliably between the retrieval-practice and concept-mapping conditions, with the concept-mapping condition producing more idea units in the presence of the to-be-practiced material than the retrieval-practice condition in its absence (66.9% vs. 55.3%), $t(54) = 3.324$, $p = .002$, $d = 0.89$, 95% CI of the difference = [4.6, 18.6].

Table 3

Percentage of Correctly Produced Idea Units During Practice of Passages 1 and 2 for Experiment 4 (Standard Errors in Parenthesis)

Condition	Retrieval practice		Concept mapping
	Cycle 1	Cycle 2	
Passage 1	58.7 (2.86)	59.3 (2.97)	70.7 (2.40)
Passage 2	50.3 (3.21)	52.9 (2.94)	63.2 (2.79)

Figure 6*Passage-3 Recall in Experiment 4*

Note. Percentage of correctly recalled passage-3 idea units for related passages as a function of condition (restudy, retrieval practice, concept mapping). The dashed line represents mean recall in the passage-3-only condition. Error bars represent standard errors.

Passage-3 Recall Test

Correct Recall. Figure 6 shows the percentage of correctly recalled passage-3 idea units for each of the four experimental conditions. A one-way ANOVA of the three conditions in which all three passages were presented for study (restudy, retrieval practice, concept mapping) showed a significant effect of condition, $F(2, 81) = 14.341$, $MSE = 438.53$, $p < .001$, $\eta_p^2 = .26$. Pairwise comparisons revealed that, relative to the restudy condition, recall was higher in both the retrieval-practice and concept-mapping conditions (retrieval practice vs. restudy: 50.5% vs. 24.1%, $t(54) = 5.385$, $p < .001$, $d = 1.44$, 95% CI of the difference = [16.6, 36.3]; concept mapping vs. restudy: 49.5% vs. 24.1%, $t(54) = 4.328$, $p < .001$, $d = 1.16$, 95% CI of the difference = [13.7, 37.3]), thus demonstrating an FT effect in response to retrieval practice and an FT-like effect in response to concept mapping. There was no reliable difference in recall between the retrieval-practice and concept-mapping conditions (50.5% vs. 49.5%), $t(54) = 0.161$, $p = .873$, $d = 0.04$, 95% CI of the difference = [-11.0, 12.9], $B_{01} = 7.384$.

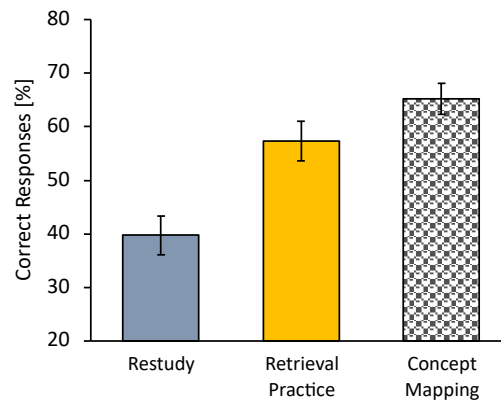
Additional comparisons to the passage-3-only condition revealed that recall in this condition was similar to recall in the restudy condition (23.3% vs. 24.1%), $t(54) = 0.147$,

$p_{\text{adj}} = .884$, $d = 0.04$, 95% CI of the difference = [-10.5, 9.0], $B_{01} = 7.400$, but lower than in the retrieval-practice and concept-mapping conditions (passage-3 only vs. retrieval practice: 23.3% vs. 50.5%, $t(54) = 5.502$, $p_{\text{adj}} = .003$, $d = 1.47$, 95% CI of the difference = [-37.0, -17.2]; passage-3 only vs. concept mapping: 23.3% vs. 49.5%, $t(54) = 4.433$, $p_{\text{adj}} = .002$, $d = 1.19$, 95% CI of the difference = [-38.0, -14.3]). For all three comparisons to the passage-3-only condition, familywise error rates were controlled by adjusting p -values according to the sequential Bonferroni procedure.

Intrusions. Intrusions were defined as idea units from passages 1 and 2 that participants erroneously produced on the passage-3 recall test. Regarding the number of intrusions, a one-way ANOVA of the restudy, retrieval-practice, and concept-mapping conditions showed a significant effect of condition, $F(2, 81) = 6.426$, $MSE = 0.23$, $p = .003$, $\eta_p^2 = .14$. Pairwise comparisons revealed that, relative to the restudy condition, participants produced fewer intrusions in both the retrieval-practice and concept-mapping conditions (retrieval practice vs. restudy: 0.04 vs. 0.46, $t(54) = 2.953$, $p = .005$, $d = 0.79$, 95% CI of the difference = [-0.7, -0.1]; concept mapping vs. restudy: 0.11 vs. 0.46, $t(54) = 2.338$, $p = .023$, $d = 0.63$, 95% CI of the difference = [-0.7, -0.1]). There was no reliable difference in intrusions between the retrieval-practice and concept-mapping conditions (0.04 vs. 0.11), $t(54) = 1.029$, $p = .308$, $d = 0.28$, 95% CI of the difference = [-0.2, 0.1], $B_{01} = 4.359$.

Final Test

Figure 7 shows the percentage of correctly answered questions on the final test in the restudy, retrieval-practice, and concept-mapping conditions. A one-way ANOVA of the three conditions showed a significant effect of condition, $F(2, 81) = 14.301$, $MSE = 329.91$, $p < .001$, $\eta_p^2 = .26$. Pairwise comparisons revealed that both the retrieval-practice and concept-mapping conditions outperformed the restudy condition (retrieval practice vs. restudy: 57.3% vs. 39.8%, $t(54) = 3.377$, $p = .001$, $d = 0.90$, 95% CI of the difference = [7.1, 27.9]; concept mapping vs. restudy: 65.2% vs. 39.8%, $t(54) = 5.453$, $p < .001$, $d = 1.46$, 95% CI of the difference = [16.0, 34.7]). There was no reliable difference in performance between the retrieval-practice and concept-mapping conditions (57.3% vs. 65.2%), $t(54) = 1.667$, $p = .101$, $d = 0.45$, 95% CI of the difference = [-17.3, 1.6], $B_{01} = 1.837$.

Figure 7*Final Test Performance in Experiment 4*

Note. Percentage of correctly answered inference questions on the final test as a function of condition (restudy, retrieval practice, concept mapping). Error bars represent standard errors.

4.5.3 Discussion

The results replicate prior findings by showing that retrieval practice between the study of related prose passages enhances recall of a final passage relative to both a restudy and a final-passage-only condition, indicating that, for related prose materials, retrieval practice can not only induce an FT effect but also proactive facilitation (Wissman et al., 2011; present Experiments 2 and 3). In addition, interim retrieval practice seems to aid the integration of related information from across a study session (Jing et al., 2016; Wahlheim, 2015), as is reflected by the finding that retrieval practice between study of related passages improved performance on a final inference test relative to the restudy condition. Going beyond the prior work, the results revealed parallel effects on learning for related passages when, instead of interim retrieval practice, interim elaborative restudy practice with concept mapping was applied. Concept mapping between study of related passages did not only improve performance on the final inference test relative to the restudy condition, but also enhanced recall of the final passage relative to the restudy and passage-3-only conditions, indicating an FT-like effect and proactive facilitation. These findings are in line with the facilitated-comprehension account of the FT effect,

which suggests that the FT effect may generalize from retrieval practice to elaborative restudy practice. The findings, however, challenge the strategy-change account of the FT effect, which predicts that the FT effect should be retrieval specific.

Regarding intrusions during recall of the final passage, the results again replicate prior findings by showing that retrieval practice between study of related passages can reduce intrusions relative to a restudy condition (cf. present Experiments 2 and 3; for similar results using a video lecture, see Szpunar et al., 2013). Extending on the prior findings, the results revealed that intrusions were reduced to a similar degree in the concept-mapping relative to the restudy condition. These findings indicate that concept mapping between study of related passages did not only induce an FT-like effect in correct recall, but also elicited an FT-like effect in the intrusion data.

4.6 Experiment 5

The results of Experiment 4 indicate that the FT effect is not retrieval specific with related prose passages, but can generalize to an FT-like effect in response to elaborative restudy practice with concept mapping. However, given that these findings may have important implications for educational practice, high confidence is requisite that concept mapping between study of related passages does potentiate new learning not only for a narrow scope of prose materials. For this reason, Experiment 5 was conducted to examine whether the results of Experiment 4 would generalize if a different set of related passages were employed as study material. In particular, the passages employed in Experiment 5 were chosen to deal with an entirely different knowledge domain than those employed in Experiment 4, aiming to reflect another example of materials students may be confronted with in everyday educational settings.

4.6.1 Method

Participants

Another 112 students (mean age = 24.0 years; 69 female, 42 male, 1 diverse) were enrolled from different German universities to take part in Experiment 5, with 28 subjects

in each of the four experimental conditions. Sample size was identical to Experiment 4. Considering the topic of the study material (i.e., retrieval competition in memory), psychology students were excluded from participation to avoid ceiling effects of prior knowledge. All participants spoke German as their native language and received monetary reward or course credit for participation.

Materials

Like in Experiment 4, three prose passages were employed as study material that were directly related to one another, covered a common topic, and contained information that was intended to be integrated. In Experiment 5, however, the topic of the study material was changed, and the passages were now consecutive sections of an expository text about mechanisms of episodic forgetting (for the English version, see [osf.io/bckyn]). Collectively, the three passages discussed the problem of retrieval competition between memory entries, and how this competition is assumed to be overcome by two cognitive mechanisms, blocking and inhibition. The text was developed on the basis of a textbook chapter (Bäuml et al., 2020) and covered educationally relevant content in the field of cognitive psychology. Again, each passage was presented in German language and about 180 words in length. Like in Experiment 4, presentation order was the same for all participants, with passage 3 always serving as the final critical passage and passages 1 and 2 always serving as the preceding passages.

Again, the final test was composed of 10 short-answer questions referring to the three prose passages' content (for the English version, see [osf.io/bckyn]). Five of the questions were bridging-inference questions, and the other five were application questions. Both types of questions required subjects to integrate information stated in two or more sentences, within or across passages, to infer the answer. Application questions additionally required subjects to apply the information to a novel situation. One example of an application question is: *“Two students are studying the names of 50 snakes for an upcoming exam. One of them studies the names in random order. The other divides the snakes into 5 groups according to their origin and not only studies the snakes' names, but also whether they come from North America, South America, Africa, Europe, or Asia. Which of the students is likely to do better, and why?”* (Answer: *“The second, as fewer snakes share the same retrieval cue”*) To answer the question, the following sentences

from the text must be integrated and correctly applied to the question: “*Retrieval competition occurs when memories share certain features with each other. When a common feature is provided as a retrieval cue, all associated memories are reactivated and compete for conscious recall [...] For example, if a group of children share the feature that they all belong to the same school class, this common feature can cause problems for the teacher’s memory. When asked for the names of the children in his class, the larger the number of children in the class, the more difficult it becomes to remember each single name.*” The Cronbach’s alpha estimate of internal consistency reached .62 for the 10 inference questions, again reflecting an acceptable reliability (cf. Schmitt, 1996).

Design, Procedure, and Scoring

Experimental design and procedure were identical to Experiment 4 (see Figure 5). For scoring purposes, passages 1–3 were parsed into idea units, and performance during practice and on the passage-3 recall test was assessed following the same criteria that had been used for scoring in Experiment 4. For example, referring to the passage-3 idea unit “*Inhibition can be triggered by retrieval practice of the relevant information, but not by relearning it*”, the response “*Retrieval practice, but not relearning can lead to inhibition*” was scored to be worth 1 point, as the response correctly reproduced the entire idea unit. However, the responses “*Retrieval practice can lead to inhibition*”, and “*Relearning does not lead to inhibition*” were both scored to be worth 0.5 points only, as both responses correctly incorporated critical parts of the idea unit but failed to mention certain aspects of it. Again, two raters separately scored 56 of the passage-3 protocols, and a Pearson correlation of .99 between their scores was obtained. The remaining protocols on passages 1–3 and the concept maps were scored by one rater. Like in Experiment 4, recall protocols and concept maps were coded using the identical coding scheme.

To assess performance on the final test, all 10 questions were scored by awarding 1 point for completely correct answers and 0.5 points for partially correct answers, again resulting in a maximum score of 10 points. Like in Experiment 4, raters were provided with a coding scheme that identified the key information a response had to contain to be scored as completely or partially correct. For example, regarding the above-mentioned question on the two students studying snake names for an exam, the responses “*The second, as fewer snakes share the same retrieval cue*” and “*The second, as fewer snakes*

compete for conscious recall” were both considered completely correct and assigned 1 point. However, the response “*The second*” was considered to be partially correct and assigned 0.5 points only. Again, two raters were blind to condition and separately scored all 84 final test protocols. High interrater reliability was obtained, with a Pearson correlation of .99 between their scores.

4.6.2 Results

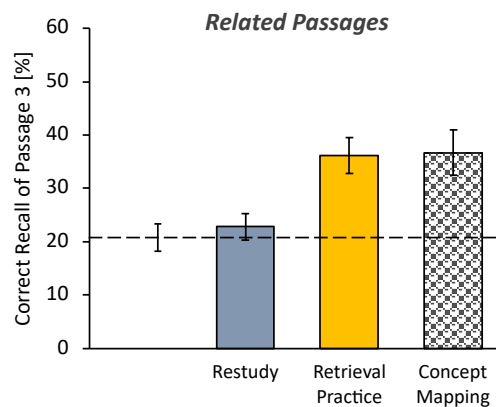
Practice Success for Passages 1 and 2

Like in Experiment 4, practice performance was pooled for passages 1 and 2 because presentation order of passages was not varied between participants. Regarding performance in the retrieval-practice condition, pairwise comparisons between the first and second cycles of retrieval practice again revealed that recall remained relatively stable across cycles (42.4% vs. 41.4%), $t(27) = 1.082$, $p = .289$, $d = 0.21$, 95% CI of the difference = $[-0.9, 2.8]$, $B_{01} = 0.339$. Like in Experiment 4, practice performance differed reliably between the retrieval-practice and concept-mapping conditions, with the concept-mapping condition producing more idea units during practice than the retrieval-practice condition (51.0% vs. 41.9%), $t(54) = 2.119$, $p = .039$, $d = 0.57$, 95% CI of the difference = $[0.5, 17.7]$ (see Table 4).

Table 4

Percentage of Correctly Produced Idea Units During Practice of Passages 1 and 2 for Experiment 5 (Standard Errors in Parenthesis)

Condition	Retrieval practice		Concept mapping
	Cycle 1	Cycle 2	
Passage 1	39.9 (2.82)	39.7 (3.03)	50.3 (3.63)
Passage 2	44.9 (3.43)	43.2 (2.98)	51.8 (3.50)

Figure 8*Passage-3 Recall in Experiment 5*

Note. Percentage of correctly recalled passage-3 idea units for related passages as a function of condition (restudy, retrieval practice, concept mapping). The dashed line represents mean recall in the passage-3-only condition. Error bars represent standard errors.

Passage-3 Recall Test

Correct Recall. Figure 8 shows the percentage of correctly recalled passage-3 idea units for each of the four experimental conditions. Like in Experiment 4, a one-way ANOVA of the restudy, retrieval-practice, and concept-mapping conditions showed a significant effect of condition, $F(2, 81) = 5.209$, $MSE = 333.70$, $p = .007$, $\eta_p^2 = .11$. Pairwise comparisons revealed that, relative to the restudy condition, recall was higher in both the retrieval-practice and concept-mapping conditions (retrieval practice vs. restudy: 36.2% vs. 22.8%, $t(54) = 3.173$, $p = .002$, $d = 0.85$, 95% CI of the difference = [4.9, 21.8]; concept mapping vs. restudy: 36.7% vs. 22.8%, $t(54) = 2.820$, $p = .007$, $d = 0.75$, 95% CI of the difference = [4.0, 23.8]), again demonstrating an FT effect in response to retrieval practice and an FT-like effect in response to concept mapping. Like in Experiment 4, there was no reliable difference in recall between the retrieval-practice and concept-mapping conditions (36.2% vs. 36.7%), $t(54) = 0.095$, $p = .925$, $d = 0.03$, 95% CI of the difference = [-11.4, 10.4], $B_{01} = 7.449$.

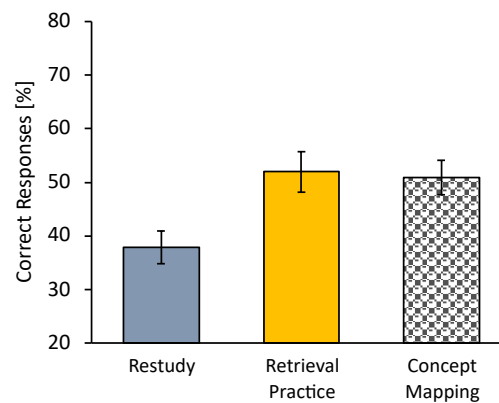
Additional comparisons to the passage-3-only condition revealed that recall in this condition was again similar to recall in the restudy condition (20.8% vs. 22.8%), $t(54) =$

0.560, $p_{\text{adj}} = .578$, $d = 0.15$, 95% CI of the difference = [-9.4, 5.3], $B_{01} = 6.365$, but lower than in the retrieval-practice and concept-mapping conditions (passage-3 only vs. retrieval practice: 20.8% vs. 36.2%, $t(54) = 3.585$, $p_{\text{adj}} = .003$, $d = 0.96$, 95% CI of the difference = [-24.1, -6.8]; passage-3 only vs. concept mapping: 20.8% vs. 36.7%, $t(54) = 3.187$, $p_{\text{adj}} = .004$, $d = 0.85$, 95% CI of the difference = [-26.0, -5.9]). Like in Experiment 4, the p -values for the three comparisons were adjusted by employing the sequential Bonferroni procedure.

Intrusions. Like in Experiment 4, a one-way ANOVA of the restudy, retrieval-practice, and concept-mapping conditions showed a significant effect of condition, $F(2, 81) = 8.278$, $MSE = 0.40$, $p < .001$, $\eta_p^2 = .17$. Pairwise comparisons revealed that, relative to the restudy condition, participants again produced fewer intrusions in both the retrieval-practice and concept-mapping conditions (retrieval practice vs. restudy: 0.04 vs. 0.68, $t(54) = 3.278$, $p = .002$, $d = 0.88$, 95% CI of the difference = [-1.0, -0.3]; concept mapping vs. restudy: 0.14 vs. 0.68, $t(54) = 2.623$, $p = .011$, $d = 0.70$, 95% CI of the difference = [-0.9, -0.1]). Again, there was no reliable difference in intrusions between the retrieval-practice and concept-mapping conditions (0.04 vs. 0.14), $t(54) = 1.406$, $p = .166$, $d = 0.38$, 95% CI of the difference = [-0.3, 0.1], $B_{01} = 2.732$.

Final Test

Figure 9 shows the percentage of correctly answered questions on the final test in the restudy, retrieval-practice, and concept-mapping conditions. Like in Experiment 4, a one-way ANOVA of the three conditions showed a significant effect of condition, $F(2, 81) = 5.510$, $MSE = 313.47$, $p = .006$, $\eta_p^2 = .12$. Pairwise comparisons revealed that both the retrieval-practice and concept-mapping conditions again outperformed the restudy condition (retrieval practice vs. restudy: 52.0% vs. 37.9%, $t(54) = 2.916$, $p = .005$, $d = 0.78$, 95% CI of the difference = [4.4, 23.8]; concept mapping vs. restudy: 50.9% vs. 37.9%, $t(54) = 2.984$, $p = .004$, $d = 0.80$, 95% CI of the difference = [4.3, 21.8]). Like in Experiment 4, there was no reliable difference in performance between the retrieval-practice and concept-mapping conditions (52.0% vs. 50.9%), $t(54) = 0.216$, $p = .830$, $d = 0.06$, 95% CI of the difference = [-8.9, 11.0], $B_{01} = 7.305$.

Figure 9*Final Test Performance in Experiment 5*

Note. Percentage of correctly answered inference questions on the final test as a function of condition (restudy, retrieval practice, concept mapping). Error bars represent standard errors.

4.6.3 Discussion

Experiment 5 was able to replicate the results of Experiment 4 employing a new set of related prose passages from a different knowledge domain. Again, the results for the concept-mapping condition mimicked those for the retrieval-practice condition, showing not only improved performance on the final inference test in both conditions, but also enhanced recall of the final passage relative to the restudy and passage-3-only conditions, reflecting FT(-like) effects and proactive facilitation. The experiment thus provides another demonstration that the FT effect is not retrieval specific with related prose passages, but can generalize to an FT-like effect in response to elaborative restudy practice with concept mapping. These findings again disagree with the strategy-change account and rather point to facilitated comprehension as possible mechanism underlying the FT effect. Like in Experiment 4, intrusions during recall of the final passage were reduced relative to the restudy condition in both the retrieval-practice and concept-mapping conditions, reflecting the presence of FT(-like) effects in the intrusion data.

4.7 Experiment 6

The goal of Experiment 6 was to examine whether the FT effect is retrieval specific with unrelated prose passages. In particular, Experiment 6 aimed to address the question of whether elaborative restudy practice with concept mapping still induces an FT-like effect if the study material is changed to unrelated passages. If facilitated comprehension indeed mediated the FT(-like) effects observed with the related passages, but the FT effect with unrelated passages was mediated by context change, retrieval specificity of the FT effect should vary with relatedness of prose passages, and the effect be retrieval specific with unrelated passages. If so, with unrelated passages, only retrieval practice between study of passages should enhance recall of the final passage relative to the restudy condition, but recall should be enhanced to a level not superior to the level in the passage-3-only condition, reflecting the presence of an FT effect but absence of proactive facilitation. Concept mapping, however, should no longer enhance recall of the final passage when applied in-between the study of unrelated passages, and thus induce neither an FT-like effect nor proactive facilitation.

4.7.1 Method

Participants

Another 112 students (mean age = 22.5 years; 79 female, 33 male) were enrolled from different German universities to take part in Experiment 6, with 28 subjects in each of the four experimental conditions. Sample size was identical to Experiments 4 and 5. All participants spoke German as their native language and received monetary reward or course credit for participation.

Materials

Like in Experiments 4 and 5, three prose passages were employed as study material. In Experiment 6, however, the three passages were unrelated to one another and covered entirely different topics: passage 1 was an expository text about former U.S. President Woodrow Wilson, passage 2 was an expository text about the chronic wasting disease, and passage 3 was an expository text about the history and uses of garlic (for the English

versions, see [osf.io/bckyn]). The text material was adapted from prior studies (passage 1: Kliegl et al., 2024; passage 2: Divis & Benjamin, 2014; passage 3: Fritz & Morris, 2015), translated into German, and shortened to approximately 180 words per passage. Like in Experiments 4 and 5, passage 3 was always presented as the final critical passage and passages 1 and 2 as the preceding passages.

Design and Procedure

Experimental design and procedure were identical to Experiments 4 and 5, with the only exception of the final test. In Experiment 6, the final test on passages 1–3 was not carried out at the end of the experiment because integration of information across passages should have been rare due to the unrelated nature of the three study texts.

Scoring

Again, passages 1–3 were parsed into idea units, and scoring of performance during practice and on the passage-3 recall test followed the same criteria as in Experiments 4 and 5. For example, regarding the passage-3 idea unit “*The Crusaders are believed to have brought garlic to Europe and Britain*”, the response “*The Crusaders are thought to have taken garlic to Europe and Britain*” was scored with 1 point, whereas the responses “*The Crusaders are thought to have taken garlic to Europe*” and “*The Crusaders are thought to have taken garlic to Britain*” were both scored with 0.5 points only. Again, two raters separately scored 56 of the passage-3 protocols, and a Pearson correlation of .99 between their scores was obtained. The remaining protocols on passages 1–3 and the concept maps were scored by one rater. Like in Experiments 4 and 5, recall protocols and concept maps were coded using the identical coding scheme.

4.7.2 Results

Practice Success for Passages 1 and 2

Like in Experiments 4 and 5, practice performance was pooled for passages 1 and 2 because presentation order of passages was not varied between participants. Regarding performance in the retrieval-practice condition, pairwise comparisons between the first and second cycles of retrieval practice again revealed that recall remained relatively stable across cycles (51.3% vs. 51.2%), $t(27) = 0.055$, $p = .956$, $d = 0.01$, 95% CI of the

Table 5

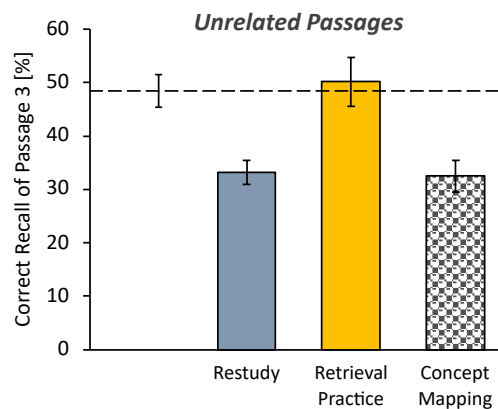
Percentage of Correctly Produced Idea Units During Practice of Passages 1 and 2 for Experiment 6 (Standard Errors in Parenthesis)

Condition	Retrieval practice		Concept mapping
	Cycle 1	Cycle 2	
Passage 1	50.9 (4.10)	52.3 (4.37)	64.7 (3.16)
Passage 2	51.7 (3.35)	50.1 (3.23)	74.6 (2.78)

difference = [-2.2, 2.4], $B_{01} = 0.187$. Like in Experiments 4 and 5, practice performance differed reliably between the retrieval-practice and concept-mapping conditions, with the concept-mapping condition producing more idea units during practice than the retrieval-practice condition (69.6% vs. 51.2%), $t(54) = 4.230$, $p < .001$, $d = 1.13$, 95% CI of the difference = [9.7, 27.1] (see Table 5).

Passage-3 Recall Test

Correct Recall. Figure 10 shows the percentage of correctly recalled passage-3 idea units for each of the four experimental conditions. A one-way ANOVA of the restudy, retrieval-practice, and concept-mapping conditions showed a significant effect of condition, $F(2, 81) = 8.696$, $MSE = 324.77$, $p < .001$, $\eta_p^2 = .18$. Critically, pairwise comparisons revealed that while recall in the retrieval-practice condition was higher than in the restudy condition (50.2% vs. 33.2%), $t(54) = 3.325$, $p = .002$, $d = 0.89$, 95% CI of the difference = [6.8, 27.3], recall in the concept-mapping condition was similar to recall in the restudy condition (32.5% vs. 33.2%), $t(54) = 0.191$, $p = .849$, $d = 0.05$, 95% CI of the difference = [-8.2, 6.7], $B_{01} = 7.343$. Thus, there was an FT effect in the retrieval-practice condition, but there was no FT-like effect in the concept-mapping condition. The difference in recall between the retrieval-practice and concept-mapping conditions was significant (50.2% vs. 32.5%), $t(54) = 3.263$, $p = .002$, $d = 0.87$, 95% CI of the difference = [6.8, 28.6].

Figure 10*Passage-3 Recall in Experiment 6*

Note. Percentage of correctly recalled passage-3 idea units for unrelated passages as a function of condition (restudy, retrieval practice, concept mapping). The dashed line represents mean recall in the passage-3-only condition. Error bars represent standard errors.

Additional comparisons to the passage-3-only condition revealed that recall in this condition was similar to recall in the retrieval-practice condition (48.4% vs. 50.2%), $t(54) = 0.324$, $p_{\text{adj}} = .747$, $d = 0.09$, 95% CI of the difference = [-12.9, 9.3], $B_{01} = 7.087$, but higher than in the restudy and concept-mapping conditions (passage-3 only vs. restudy: 48.4% vs. 33.2%, $t(54) = 3.950$, $p_{\text{adj}} = .003$, $d = 1.06$, 95% CI of the difference = [7.5, 23.0]; passage-3 only vs. concept mapping: 48.4% vs. 32.5%, $t(54) = 3.738$, $p_{\text{adj}} = .002$, $d = 1.00$, 95% CI of the difference = [7.4, 24.5]). For all three comparisons to the passage-3-only condition, p -values were again adjusted by using the sequential Bonferroni procedure.

Intrusions. Not a single participant had an intrusion from a preceding passage, irrespective of condition.

4.7.3 Discussion

The results again replicate prior findings by showing that retrieval practice between the study of unrelated prose passages enhances recall of a final passage relative to restudy

but not relative to when only the final passage has been studied, indicating that, for unrelated prose materials, retrieval practice can induce an FT effect but does not induce proactive facilitation (cf. present Experiments 1 and 3). Extending on the prior findings, the results revealed that concept mapping between study of passages did no longer enhance recall of the final passage when the study material was changed to unrelated passages, indicating that neither an FT-like effect nor proactive facilitation was present in response to elaborative restudy practice with concept mapping. These results, together with those of Experiments 4 and 5, indicate that retrieval specificity of the FT effect can vary with relatedness of prose passages, and the effect be retrieval specific with unrelated passages. These findings are in line with the idea that the FT effect with unrelated passages is not mediated by facilitated comprehension, but rather mediated by context change. Replicating prior findings on intrusions for unrelated passages, intrusions during recall of the final passage did not emerge irrespective of condition, reflecting the absence of an FT effect in the intrusion data (cf. present Experiments 1 and 3).

4.8 Discussion of Experiments 4–6

The second set of experiments demonstrates that retrieval specificity of the FT effect can vary with relatedness of prose passages. With related prose passages, not only retrieval practice on the previously studied passages induced an FT effect relative to the restudy condition, but recall of the final critical passage was enhanced to a similar degree in the concept-mapping condition, indicating that the FT effect generalized to an FT-like effect in response to elaborative restudy practice with concept mapping (Experiments 4 and 5). In addition, with related passages, both types of learning activity induced proactive facilitation relative to the passage-3-only condition, and both improved performance on a final inference test relative to the restudy condition (Experiments 4 and 5). With unrelated prose passages, in contrast, no proactive facilitation was observed, and only retrieval practice induced an FT effect, while concept mapping was found to leave recall of the final passage unaffected relative to the restudy condition (Experiment 6). The findings thus indicate that the FT effect is retrieval specific with unrelated but not with related prose passages.

The findings on retrieval specificity suggest that the role of facilitated comprehension in the FT effect with prose materials may critically depend on relatedness of study materials. According to the facilitated-comprehension account of the FT effect, the FT effect is mediated by knowledge integration, assuming that interim retrieval practice increases the accessibility of preceding materials during subsequent encoding of the final material, which can aid knowledge integration across the different parts of the study material and thus facilitate comprehension of the final material (Jing et al., 2016; Wissman et al., 2011; see also Wahlheim, 2015). Following this rationale, the FT effect as induced by facilitated comprehension may resemble effects of prior knowledge on learning, and thus be not retrieval specific but generalize to an FT-like effect in response to elaborative restudy practice with concept mapping. Indeed, because concept mapping is assumed to aid knowledge integration, requiring relational processing of the study material (Chevron, 2014; Novak & Cañas, 2008; Novak & Gowin, 1984), interim concept mapping may also promote one's knowledge on the preceding materials and thus facilitate comprehension of the final material, very similar to interim retrieval practice.

In line with this suggestion, the present results revealed parallel effects on learning for related passages when, instead of interim retrieval practice, interim concept mapping was applied for practice on the preceding passages. In particular, with related passages, the two types of learning activity did not only promote performance on the final inference test, which is thought to measure knowledge integration (e.g., Graesser et al., 1994; Kintsch, 1994; Wiley et al., 2005), but did also enhance recall of the final passage relative to both the restudy and passage-3-only conditions, indicating that practice-induced knowledge on the preceding passages indeed potentiated learning of the final passage. However, because knowledge effects on learning are typically restricted to domain-related information (e.g., Hunt & Rawson, 2011; Witherby & Carpenter, 2022), changing study material from related to unrelated materials should eliminate any effects of facilitated comprehension, irrespective of whether they were induced by retrieval practice or concept mapping. Consistently, the present results for unrelated passages revealed that concept mapping did no longer induce an FT-like effect relative to the restudy condition, nor did retrieval practice induce proactive facilitation relative to the passage-3-only condition. The pattern of results observed across the present three experiments thus may

indicate a critical role of facilitated comprehension in the FT effect with related prose materials but not with unrelated prose materials.

The findings on retrieval specificity are not easily captured by the strategy-change account of the FT effect. According to this account, the FT effect is mediated by strategy optimization, assuming that retrieval practice on preceding materials provides subjects with performance-relevant information – like, for instance, information about retrieval failures and the format of forthcoming recall tests – which can induce them to apply new and more efficient strategies during subsequent encoding and retrieval (Chan, Manley, et al., 2018; Soderstrom & Bjork, 2014; Wissman et al., 2011). Because restudy practice should be less effective than retrieval practice in alerting subjects to such information, and thus should also be less effective in inducing strategy change (e.g., Chan, Manley, et al., 2018; Pyc & Rawson, 2012), the account predicts that the FT effect should be retrieval specific and thus not generalize to restudy practice. The present results for related passages contrast with this prediction, however, revealing that recall of the final passage was enhanced to a similar degree in the retrieval-practice and concept-mapping conditions.

At first glance, the results may be reconciled with a variant of the strategy-change account, which assumes that concept mapping can also induce strategy change by alerting subjects to their (poor) prior learning. Indeed, during concept-map creation, subjects may realize that they do not have a good grasp of the previously studied material, having difficulties extracting the key concepts out of the text and the relations they have to one another. If so, one may argue that the FT-like effect observed in response to concept mapping was not mediated by facilitated comprehension, but rather mediated by strategy change. However, because effects mediated by strategy change should not be restricted to related study materials (e.g., Soderstrom & Bjork, 2014; Yang et al., 2022), one would expect concept mapping to also induce an FT-like effect with unrelated passages, which is not what the present results showed. The present results thus do not easily align with the strategy-change account of the FT effect. In particular, the findings on retrieval specificity challenge the idea of a critical role of strategy change in the FT effect with related prose materials.

Finally, the findings on retrieval specificity are consistent with the idea that context change plays a critical role in the FT effect with unrelated prose materials but not

with related prose materials. According to the context-change account of the FT effect, the FT effect reflects an enhancement in “list” discrimination, assuming that retrieval between study of materials drives mental context change and thus creates distinct context cues for previously and subsequently studied materials, which can induce release from proactive interference (Divis & Benjamin, 2014; Pastötter et al., 2011; Szpunar et al., 2008). Because retrieval practice but not restudy practice can speed mental context change and thus induce “list” discrimination (e.g., Pastötter et al., 2011), the account predicts that the FT effect should be retrieval specific and thus not generalize to any type of restudy practice. The results for the unrelated passages are in line with this prediction, showing that recall of the final passage was similar in the restudy and concept-mapping conditions, both of which created lower recall levels than the retrieval-practice condition. In contrast, the results for the related passages disagree with this prediction, revealing similar recall levels in the retrieval-practice and concept-mapping conditions, both of which enhanced recall of the final passage relative to the restudy condition. The findings on retrieval specificity thus fit with the idea that context change primarily contributes to the FT effect with unrelated but not with related prose materials. Consistent with such a proposal, the present three experiments once again demonstrated that proactive facilitation arises with related but not with unrelated passages (see also Experiments 1–3). Indeed, because context change supposedly induces release from proactive interference by eliminating the previously studied materials from the mental search set (e.g., Bäuml & Kliegl, 2013), the context-change account predicts that the FT effect should create recall levels not superior to the recall level when the final material was studied only, which is what the present results for unrelated passages showed.

5

General Discussion

Goal of the present dissertation was to examine the potential role of relatedness of study material for the FT effect with prose materials. To this end, two research questions were addressed in a series of six experiments. First, does the basic FT effect arise with both highly related and completely unrelated prose materials, thus addressing the issue of whether relatedness of materials can influence the presence of the FT effect? And second, can basic characteristics of the FT effect vary with relatedness of prose materials, thus addressing the issue of whether qualitatively different FT effects arise for related and unrelated prose materials? Across the six experiments, retrieval practice on previously studied materials was found to enhance recall of critical final material compared to restudy of the preceding materials, both when completely unrelated prose passages (Experiments 1, 3, and 6) and when highly related prose passages (Experiments 2, 3, 4, and 5) were employed as the study material. The present dissertation thus demonstrates that the same basic FT effect can be present for the two types of prose materials. Across the six experiments, however, basic characteristics of the FT effect were found to vary with relatedness of prose materials. With unrelated prose passages, the FT effect generalized to an FT-like effect in response to semantic generation (Experiment 1), and the FT effect led to recall levels for the critical final material that were similar to the recall level when the final material was studied only (Experiments 1, 3, and 6). In contrast, with related prose passages, no FT-like effect was observed in response to semantic generation (Experiment 2), and the FT effect led to recall levels for the final material that were superior to the level in a final-material-only condition, reflecting the presence of proactive facilitation (Experiments 2, 3, 4, and 5). As another difference between materials, retrieval specificity of the FT effect was found to vary with relatedness of prose materials. In particular, the FT effect generalized to an FT-like effect in response to elaborative

restudy practice when related prose passages (Experiments 4 and 5) but not when unrelated prose passages (Experiment 6) were studied, indicating that the FT effect is retrieval specific with unrelated but not with related passages. The differences in findings between the two types of materials suggest that qualitatively different FT effects arise for related and unrelated prose materials.

5.1 Implications for Theoretical Accounts of the Forward Testing Effect

5.1.1 A Two-Factor Account of the Forward Testing Effect With Prose Materials

The differences in findings between related and unrelated study materials are consistent with a two-factor account of the FT effect with prose materials. According to this account, in general, both context change and facilitated comprehension can contribute to the FT effect with prose material, but the relative contributions of the two cognitive mechanisms vary depending on relatedness of study materials. First, the account assumes that retrieval activities, like retrieval practice or semantic generation, induce context change and thus enhance “list” discrimination, which can improve recall of critical final material by enabling a more focused memory search at test. Critically, such context change should primarily come into play when unrelated materials are studied and play much less of a role for related materials. The reason is that, with related materials, encoding of the subsequent material may trigger covert retrieval of the preceding materials (Hintzman, 2004, 2010). Because retrieval of previous material may reinstate the context that was present when that material was studied (Bäuml & Schlichting, 2014; Bäuml & Trißl, 2022), context may be kept relatively stable during study of related materials and the induced context change thus be eliminated. In contrast, with unrelated materials, the preceding materials should hardly come to mind during encoding of the subsequent material and context reinstatement thus be rare. Context change induced by interim retrieval should thus be effective with unrelated materials. Second, the account further assumes that retrieval practice can also induce facilitated comprehension, because retrieval practice can promote one’s knowledge on the preceding materials and thus

enable subjects to fill gaps in comprehension encountered during encoding of the subsequent material via knowledge integration. Such facilitated comprehension should take effect mainly with related materials and play a minor role, if any, for unrelated materials. The reason is that facilitative effects of knowledge on learning are typically restricted to material related to the domain of knowledge (Hunt & Rawson, 2011; Witherby & Carpenter, 2022). Practice-induced knowledge on related preceding materials should thus be required to facilitate comprehension of the subsequent material, and study and practice of unrelated preceding materials hardly afford such facilitated comprehension.

According to the two-factor account, the FT effect should therefore be primarily mediated by context change with unrelated prose materials and by facilitated comprehension with related prose materials. As a result, the FT effect should arise with both related and unrelated prose materials, but basic characteristics of the FT effect should vary with relatedness of study materials. The FT effect with unrelated prose materials should be retrieval specific and generalize to an FT-like effect in response to semantic generation, but the effect should not reflect proactive facilitation. Retrieval specificity should arise because restudy practice typically does not induce context change (Bäuml & Kliegl, 2013; Pastötter et al., 2011). Context-change effects have been shown to arise in response to various forms of retrieval, however, such as retrieval practice or semantic generation (Jang & Huber, 2008; Shiffrin, 1970). Proactive facilitation should be absent because context change supposedly improves recall by eliminating the preceding materials from the mental search set at test (Bäuml & Kliegl, 2013; Sahakyan & Kelley, 2002); context change thus should not enhance recall to a level superior to the recall level when no preceding materials were presented.

In contrast, the FT effect with related prose materials should not be retrieval specific and also not generalize to semantic generation, but the effect may well reflect proactive facilitation. Retrieval specificity should not apply because knowledge on preceding materials can induce facilitated comprehension (Jing et al., 2016; Wissman et al., 2011), and knowledge on such materials can be fostered by various types of practice, including retrieval practice (see Agarwal et al., 2021; Karpicke, 2017) and elaborative restudy practice (see Nesbit & Adesope, 2006; Schroeder et al., 2018). Knowledge on preceding materials, however, is typically not promoted by semantic generation (Divis &

Benjamin, 2014), suggesting that semantic generation does not induce facilitated comprehension. Proactive facilitation may be present because facilitated comprehension can increase the effectiveness of encoding (Jing et al., 2016; Wissman et al., 2011); facilitated comprehension thus may lead to recall levels on critical final material that exceed the recall level when only the final material was studied and knowledge on the preceding materials is absent. The pattern of results observed across the present six experiments is in line with these predictions.

5.1.2 Further Implications for Accounts of the Forward Testing Effect With Prose Materials

Whereas the present results suggest critical roles of context change and facilitated comprehension in the FT effect with prose materials, they do not easily align with the idea of a critical role of strategy change in the effect with these types of materials. According to the strategy-change account, the FT effect is mediated by strategy optimization, assuming that retrieval practice can induce subjects to switch to more efficient encoding and retrieval strategies, because retrieval practice can provide insights into the learning task at hand – for instance, into retrieval failures and the format of forthcoming recall tests (Chan, Manley, et al., 2018; Soderstrom & Bjork, 2014; Wissman et al., 2011). Because restudy practice or semantic generation do not provide such insights that would promote strategy optimization (Kliegl & Bäuml, 2021b; Pyc & Rawson, 2012), the account predicts that the FT effect should arise exclusively in response to retrieval practice and thus not generalize to restudy practice or semantic generation. However, because study and retrieval of preceding material is supposed to be required to induce strategy change, which in turn may increase the effectiveness of subsequent learning, the account suggests that the FT effect may indeed reflect proactive facilitation (Wissman et al., 2011). Regarding the present results for unrelated prose passages, the account can explain why retrieval practice, but not elaborative restudy practice induced an FT effect (Experiment 6), but cannot explain why an FT-like effect was observed in response to semantic generation (Experiment 1). Similarly, concerning the present results for related prose passages, the account can explain why retrieval practice induced proactive facilitation (Experiments 2, 3, 4, and 5) and why semantic generation did not

induce an FT-like effect (Experiment 2), but it has difficulty explaining why an FT(-like) effect was found in response to both retrieval practice and elaborative restudy practice (Experiments 4 and 5). A variant of the strategy-change account may suggest that elaborative restudy practice can also induce strategy change by alerting subjects to their (poor) prior learning, which could explain why this type of practice was able to create an FT-like effect (see also Discussion of Experiments 4–6). If so, however, it remains unclear why elaborative restudy practice did not induce an FT-like effect with unrelated prose passages, given that induced strategy change should also promote the learning of unrelated materials (Soderstrom & Bjork, 2014; Yang et al., 2022). The present results thus do not easily align with the strategy-change account, indicating that it cannot capture the whole range of findings on the FT effect observed in the present experiments, be it for related or unrelated prose materials.

The present results do not only challenge the strategy-change account, but they are also inconsistent with the context-change explanation (Divis & Benjamin, 2014; Pastötter et al., 2011; Szpunar et al., 2008) and the facilitated-comprehension explanation (Jing et al., 2016; Wissman et al., 2011) when applied as single-mechanism accounts of the FT effect. According to the context-change account, the FT effect should show retrieval specificity and generalize to semantic generation, but the effect should not involve proactive facilitation. In particular, the account does not propose a role of relatedness of study material for the FT effect, which contrasts with the differences in findings between related and unrelated prose passages observed in the present experiments. The present results are also inconsistent with the facilitated-comprehension account as a single-mechanism view, which suggests quantitatively different FT effects for related and unrelated prose passages. Indeed, since the study and practice of unrelated preceding materials should hardly afford facilitated comprehension of the subsequent material, the FT effect should disappear when the study material is changed from related to unrelated prose passages, which is not what the present results showed. A variant of the facilitated-comprehension account may suggest that the study and practice of unrelated preceding materials can also induce facilitated comprehension – for instance, by using the preceding information in a compare-contrast manner (cf. Wissman et al., 2011, p. 1146). While such a variant of the account may explain why the same basic FT effect arose for related and unrelated prose passages, it remains unclear why retrieval

specificity of the FT effect was found to vary with relatedness of prose passages. Indeed, if facilitated comprehension was effective with both related and unrelated passages, elaborative restudy practice should affect recall for the two types of materials similarly and either create or not create an FT-like effect, depending on whether it may or may not induce facilitated comprehension. On a more general level, such a variant of the facilitated-comprehension account does not propose a role of relatedness of study material for the FT effect, which again disagrees with the present findings.

5.1.3 The Potential Role of Release From Proactive Interference for the Forward Testing Effect With Prose Materials

The present experiments were not designed to examine the role of release from proactive interference for the FT effect. Nonetheless, some of the results may be of relevance for the question of whether the FT effect with prose materials reflects proactive-interference release – be it through context change, facilitated comprehension, or some other mechanism (for a detailed discussion on proactive-interference release in multiple-list learning, see Kliegl & Bäuml, 2021a). With unrelated prose passages, recall in the restudy condition was reduced and recall in the retrieval-practice condition was unaffected relative to the final-passage-only condition (Experiments 1, 3, and 6), indicating that the encoding of the preceding passages caused proactive interference in the restudy condition and the additional retrieval practice on the preceding passages induced release from such proactive interference (Bäuml & Kliegl, 2013; Szpunar et al., 2008). In contrast, with related prose passages, recall in the restudy condition was unaffected and recall in the retrieval-practice condition was even increased relative to the final-passage-only condition (Experiments 2, 3, 4, and 5), a finding that may indicate that release from proactive interference did not contribute much to the FT effect with related prose passages. Such view might be premature.

Indeed, also with related prose passages, proactive interference may have arisen in the restudy condition, but the induced recall impairment been masked by an improvement effect resulting from better encoding of the critical passage in the restudy than the final-passage-only condition. In the restudy condition, the encoding of the critical passage may have benefited from the encoding of the preceding related passages, which

may have provided background knowledge to deal with comprehension gaps during study (Kintsch, 1988, 1998). Consistent with this suggestion, recall levels in the final-passage-only condition were indeed much lower for related prose passages (24.7%, 23.3%, and 20.8% in Experiments 2, 4, and 5) than for unrelated prose passages (61.8% and 48.4% in Experiments 1 and 6), raising the possibility that, with related passages, the encoding of the final passage was relatively hard when the preceding passages were missing.

Prior work provided evidence that proactive interference can be reduced and even reversed to proactive facilitation by promoting integration of the studied information (Jacoby et al., 2015; Wahlheim, 2015; Wahlheim & Jacoby, 2013). In a similar vein, the present results for related prose passages may indicate that while proactive interference turned to proactive facilitation in the retrieval-practice condition, integration across passages was only partially effective in the restudy condition, resulting in a neutral rather than beneficial effect on recall relative to the final-passage-only condition. In line with such a proposal, performance on the final inference test was found to be lower in the restudy than the retrieval-practice condition (Experiments 4 and 5), suggesting that interim restudy was indeed less effective than interim retrieval practice in promoting integration of knowledge from across the studied information. The finding that intrusion rates reduced from restudy to retrieval practice for related passages (Experiments 2, 3, 4, and 5) basically fits with the suggestion that, also with related prose passages, release from proactive interference may have been less complete in the restudy than the retrieval-practice condition. If so, release from proactive interference may have played a critical role not only for the FT effect with unrelated prose materials but also for the FT effect with related prose materials. Future work may examine this view in more detail.

One finding that warrants further discussion is the observation that, with related prose passages, recall in the semantic-generation condition was reduced relative to the final-passage-only condition (Experiment 2). Although this observation aligns with the suggestion that proactive interference may have arisen also with related passages, it remains unclear why the preceding study of additional passages induced recall impairment in the semantic-generation but not the arithmetic-tasks condition. One potential explanation might be that interim semantic generation impaired access to the preceding passages (see Divis & Benjamin, 2014), which may have prevented facilitated comprehension during subsequent study, resulting in a detrimental rather than neutral

effect of prior study on recall of the critical passage. Consistently, on the final cumulative test, recall of the preceding passages was lower in the semantic-generation than the arithmetic-tasks condition (see Table 2), indicating that memory for the preceding passages was indeed impaired in response to interim semantic generation. It should be stressed, however, that this reasoning is post hoc, and the proposal of reduced facilitated comprehension in the semantic-generation condition is not directly supported by the present results. Moreover, it would be remiss not to mention that the explanation raised cannot easily be reconciled with at least two considerations. First, if semantic generation indeed reduced access to the preceding passages, this should have allowed for a more focused memory search on the critical passage at test and therefore have led to recovery from proactive interference, which is not what the present results suggest. And second, since the impaired access as induced by semantic generation has been attributed to context change (Divis & Benjamin, 2014; Rupperecht & Bäuml, 2017), the explanation appears to be inconsistent with the present two-factor account, according to which induced context change should hardly take effect with related prose materials (see chapter 5.1.1). The present thesis thus cannot provide a fully satisfying explanation for why, with related prose passages, interim semantic generation induced recall impairment for the critical passage. Future work may address the issue in more detail and examine whether, with related prose materials, release from proactive interference is less complete in the semantic-generation than the arithmetic-tasks condition.

5.2 Relation to the Memory Literature

5.2.1 Relation to Prior Studies on the Forward Testing Effect With More Complex Study Material

The present findings are consistent with at least four lines of findings on the FT effect with more complex study material. First, the present results agree with prior work in which the FT effect has been found to be present for more complex materials, such as prose passages or video lectures. While the FT effect has repeatedly been found with related complex materials (Jing et al., 2016; Szpunar et al., 2013, 2014; Wissman et al.,

2011), prior studies on the FT effect with unrelated complex materials either employed materials that were not directly related to one another but shared a common topic (Wissman et al., 2011), or they differed in critical ways from other studies on the FT effect by employing selective interim practice instead of nonselective interim practice (Yue et al., 2015), or by employing interim semantic generation instead of interim retrieval practice (Divis & Benjamin, 2014). The present dissertation replicates the finding of an FT effect for highly related complex materials and demonstrates the same basic FT effect for completely unrelated complex materials, keeping procedural details constant across the two types of materials. Second, the present dissertation also replicates the Divis and Benjamin's (2014) finding of an FT-like effect in response to semantic generation for unrelated prose passages, but shows that it does not generalize to related prose passages. The present results are the first to directly compare the effects of interim retrieval practice and interim semantic generation for more complex study material, and thus are the first to show that the two types of retrieval induce parallel FT(-like) effects with unrelated but not with related complex materials. Third, the present results align with Wissman et al.'s (2011) prior work in which the FT effect has been found to reflect proactive facilitation – i.e., higher recall of the critical final material in the retrieval-practice than in a final-material-only condition. This finding has been reported for related prose passages. The present dissertation replicates the finding and shows that proactive facilitation is absent in the FT effect with unrelated prose passages. Fourth, while the present results are the first to show that retrieval specificity of the FT effect varies with relatedness of study material, single aspects of the results show agreement with prior work. A few prior studies have already indicated that plain restudy practice is not sufficient to create an FT-like effect, be it for related or unrelated complex materials (Jing et al., 2016; Szpunar et al., 2013; Yue et al., 2015). The present results agree with these findings but show that type of restudy practice can be critical for retrieval specificity of the FT effect. In particular, the present results show that, with related but not with unrelated complex material, the FT effect generalizes to an FT-like effect when the restudy practice is enriched with elaborative reprocessing. All of these findings are in line with the two-factor account suggested above.

5.2.2 Relation to Prior Studies on the Forward Testing Effect With Simple Word Lists

The present findings on the role of relatedness of study material for the FT effect are also consistent with prior word-list studies. Employing lists of unrelated words as well as lists of categorially related words, previous studies have found the FT effect to emerge with both types of word lists, suggesting that the presence of the basic FT effect does not critically depend on relatedness of study lists (e.g., Kliegl & Bäuml, 2021b, 2023; Szpunar et al., 2008). In line with these results, the present findings demonstrate that the FT effect arises with both related and unrelated prose passages, indicating that, also for more complex material, the presence of the basic FT effect does not depend much on relatedness of materials. Word-list studies further demonstrated that basic characteristics of the FT effect vary with relatedness of study material. With unrelated lists, the FT effect generalized to semantic generation (Divis & Benjamin, 2014; Kliegl & Bäuml, 2021b, 2023; Pastötter et al., 2011) and disappeared with prolonged delay interval between study and recall of the critical list, as well as with prolonged lag between the study of the preceding lists and study of the critical list (Kliegl & Bäuml, 2021b). In contrast, with categorially related lists, the FT effect did not generalize to semantic generation (Kliegl & Bäuml, 2021b, 2023) and was still present after prolonged delay interval and lag (Chan, Manley, et al., 2018; Kliegl & Bäuml, 2021b). The differences in findings were interpreted as evidence for a two-factor account of the FT effect, which suggests different relative contributions of context change and strategy change to the FT effect with the two types of lists (Kliegl & Bäuml, 2021b, 2023). In line with prior word-list studies, the present findings provide evidence that qualitatively different FT effects also arise for related and unrelated prose passages.

While there is thus consistency with prior word-list studies concerning the role of relatedness of study material for the FT effect, the present findings may additionally point to a role of complexity of study material for the FT effect. Indeed, since basic characteristics of the FT effect have been shown to differ for unrelated and categorially related lists, the question arises whether the FT effect with prose materials may or may not resemble the effect with at least one of the two types of lists. For unrelated prose passages, the finding of an FT-like effect in response to semantic generation (Divis &

Benjamin, 2014; present Experiment 1) agrees with the results of word-list studies reporting similar outcomes for unrelated lists (Divis & Benjamin, 2014; Kliegl & Bäuml, 2021b, 2023; Pastötter et al., 2011). The present and prior findings thus allude to the possibility that the FT effect with unrelated prose passages resembles the effect with unrelated lists. For related prose passages, two lines of findings fit with the results for categorially related lists. The one finding is that semantic generation does not create an FT-like effect for the two types of materials (present Experiment 2; Kliegl & Bäuml, 2021b, 2023); the other finding is that interim retrieval practice improves integration of information across related prose passages (present Experiments 4 and 5; see also Jing et al., 2016), which agrees with the finding that, with categorially related lists, interim retrieval practice increases recall clustering on the basis of the words' category membership (Chan, Manley, et al., 2018; Chan et al., 2020; Kliegl & Bäuml, 2021b). At first glance, the suggestion may thus arise that the FT effect with related prose passages resembles the effect with categorially related lists.

At least two points weigh against this proposal. First, the finding of proactive facilitation for related prose passages (Wissman et al., 2011; present Experiments 2, 3, 4, and 5) contrasts with the results of many word-list studies, in which interim retrieval practice has been found to maintain rather than improve recall from the first to the subsequent study lists (e.g., Chan, Manley, et al., 2018; Chan et al., 2020; Pastötter et al., 2011; Szpunar et al., 2008). And second, because study and practice of preceding lists should not contribute much to better comprehension of subsequent lists, it remains unclear why elaborative restudy practice should create an FT-like effect not only with related prose passages (present Experiments 4 and 5) but also with categorially related lists. Hence, while the FT effect may have similar characteristics for unrelated prose passages and unrelated lists, the present and prior findings point to differences between the FT effect with related prose passages and the FT effect with categorially related lists. If so, context change may play a critical role in the FT effect with unrelated materials, be they word lists or prose passages. In contrast, for related materials, facilitated comprehension may not yet come into play in the FT effect with simple word lists and take effect only with more complex study material like prose passages.

Since for related materials, complexity of study material thus seems to influence the FT effect, one may wonder whether, for these types of materials, there can be a transition from strategy change to facilitated comprehension contributing to the FT effect. One proposal is that interim retrieval practice “always” enables metacognitive insights into the learning task at hand, irrespective of whether simple or more complex study material is employed. Such metacognitive insights may translate into strategy change when the materials lend themselves to relational processing, as should be the case with related materials in general (e.g., Chan, Manley, et al., 2018; Jing et al., 2016; Kliegl & Bäuml, 2021b; present Experiments 4 and 5), and may additionally translate into knowledge-based processing when the materials lend themselves to facilitated comprehension, as should be the case with materials that are directly related in content and build off one another (e.g., Jing et al., 2016; Wissman et al., 2011; present Experiments 2, 3, 4, and 5). If so, with related materials, strategy change may “always” contribute to the FT effect to a certain degree, while facilitated comprehension may contribute to the effect for only those materials that are sufficiently complex. If parsimony demands explaining the FT effect using a streamlined set of cognitive mechanisms, however, facilitated comprehension seems to provide a more comprehensive explanation of the FT effect with more complex related material than strategy change (see chapters 5.1.1 and 5.1.2; for further discussion on the potential role of metacognition for facilitated comprehension, see chapter 5.5.3).

5.2.3 Relation to the Literature on Knowledge-Enhanced Learning

The present findings suggest a link between the FT effect with related prose materials and the potentiating effects of knowledge on new learning. Prior research already pointed to the idea that the FT effect with related complex material may be mediated by knowledge integration. In this study, Jing et al. (2016) employed a video lecture on public health as study material and found that interim retrieval practice not only induced an FT effect for the critical final part of the lecture, as measured by an immediate recall test, but also improved integration of related information within and across the different parts of the lecture, as measured by a final recall test (for similar results using prose passages, see Wissman & Rawson, 2015). The authors proposed that interim retrieval practice may

make the practiced material more readily accessible during study of the subsequent material, which may aid the integration of knowledge deriving from different parts of the study material and thus facilitate its comprehension (for a similar proposal, see Wissman et al., 2011; see also Wahlheim, 2015). Extrapolating from this idea, interim retrieval practice may promote one's knowledge on the preceding materials, enabling subjects to fill gaps in comprehension encountered during encoding of the subsequent material and thus eliciting an FT effect.

Such a claim is supported by two lines of findings in the literature. First, retrieval practice on previously studied material has not only been shown to enhance its direct retention (e.g., as measured by free recall; Roediger & Karpicke, 2006b), but also its deeper comprehension (e.g., as measured by inference questions; Butler, 2010; Hinze et al., 2013; Karpicke & Blunt, 2011; McDaniel et al., 2009), indicating that retrieval practice can indeed promote one's knowledge on the practiced material. And second, prior knowledge in a domain has been shown to potentiate the learning of material related to this domain (e.g., Alexander et al., 1994a, 1994b; Boscolo & Mason, 2003; Bransford & Johnson, 1972; Buchin & Mulligan, 2023; McNamara et al., 1996), and such effects have been attributed to an integration-based mechanism. According to the text-comprehension model by Kintsch (1988, 1998), learners frequently deal with comprehension gaps when reading complex texts. The assumption is that prior knowledge in a text's domain allows the learner to make inferences on how the single parts of the text relate together, enabling the construction of a situation model that entails connections made between the text information and the learner's prior knowledge base (for an illustration, see chapter 4.1). Returning to the proposal of a link with the FT effect with related prose materials, the critical idea is that knowledge requisite for text comprehension may also derive from related material studied prior to the study of the critical material.

The present findings provide support for the idea that the FT effect with related prose materials resembles knowledge effects on learning. First, since knowledge acquisition can be fostered not only by retrieval practice, but also by other types of learning activities, such as elaborative studying with concept mapping (see Nesbit & Adesope, 2006; Schroeder et al., 2018), learning of the critical material may also be improved when practice on the preceding materials occurs by interim concept mapping.

In line with this suggestion, the present results for related prose passages revealed that recall of the critical passage was enhanced to a similar degree in the retrieval-practice and concept-mapping conditions, both of which created FT(-like) effects (Experiments 4 and 5). Second, since high-knowledge learners more readily acquire domain-related information than low-knowledge learners (e.g., Boscolo & Mason, 2003; Buchin & Mulligan, 2023; McNamara et al., 1996), learning of the critical material may be more effective in the presence than absence of practice-induced knowledge on related preceding material. Consistent with this suggestion, the present results for related prose passages showed that recall in the retrieval-practice and concept-mapping conditions was higher than in a condition in which the critical passage was studied only, reflecting the presence of proactive facilitation (Experiments 4 and 5). Third, since background knowledge enables the learner to fill comprehension gaps during reading (Kintsch, 1988, 1998), practice-induced knowledge on related preceding material may promote knowledge integration across the study material. In line with this suggestion, the present results for related prose passages revealed that performance on a final inference test was improved in the retrieval-practice and concept-mapping conditions (Experiments 4 and 5), which was intended to measure the types of inferences necessary for situation-model-level comprehension (e.g., Graesser et al., 1994; Kintsch, 1994; Wiley et al., 2005). Fourth, since knowledge effects on learning are typically restricted to domain-related information (e.g., Hunt & Rawson, 2011; Witherby & Carpenter, 2022), practice-induced knowledge on preceding material should no longer improve the learning of critical material when the study material is unrelated. Consistently, the present results showed that concept mapping did no longer induce an FT-like effect with unrelated prose passages, nor did retrieval practice induce proactive facilitation (Experiment 6). The present findings thus point to parallels between the FT effect with related prose materials and knowledge-enhanced learning.

While the present findings suggest a link between the FT effect with related prose materials and knowledge effects on learning, they remain silent on the cognitive processes underlying the knowledge integration as induced by retrieval practice and concept mapping. Knowledge integration was measured by a final inference test in the present study, and the memory demands involved in answering inference questions have been proposed to comprise three components: recognition, recall, and execution (Barnett &

Ceci, 2002). In the present experiments, there were no memory demands with respect to the recognition component because subjects were explicitly told that they were questioned on the passages they had previously studied. There were significant demands with respect to recall and execution, however, and one may speculate that the two types of learning activities affected the two components differently. On the basis of prior work showing that retrieval practice promotes retention of the practiced information, it has been suggested that retrieval practice improves inference performance by primarily affecting the recall component (e.g., Butler, 2010; Karpicke & Blunt, 2011). Elaborative studying with concept mapping was found to be less effective than retrieval practice in promoting long-term retention (e.g., Karpicke & Blunt, 2011; Lechuga et al., 2015) and may have improved inference performance by primarily affecting the execution component. Indeed, concept-map-creation during learning has been argued to foster well-constructed memory representations (Chevron, 2014; Novak & Cañas, 2008; Novak & Gowin, 1984), enabling subjects to better apply the knowledge they gained about the previously studied information. If so, concept mapping may aid knowledge integration by promoting relational processing of the practiced information, while retrieval practice may achieve improved integration by making the practiced information more accessible. Unfortunately, the final inference test employed in the present study is not able to separate the recall and execution components. Future research is required to clarify whether retrieval practice and concept mapping improve knowledge integration differently when applied in-between the study of related prose materials.

5.2.4 An Additional Role of Attentional Factors in the Forward Testing Effect With Prose Materials?

While the present two-factor account suggests critical roles of context change and facilitated comprehension in the FT effect with prose materials, prior work has proposed a potential role of attentional factors in the FT effect. Indeed, in several recent studies, the proposal has been made that attentional resources may decline the more material is studied in succession, but interim retrieval practice may help sustain attention and thus improve encoding of material encountered later in the study session (e.g., Chan et al.,

2020; Jing et al., 2016; Kliegl & Bäuml, 2023; Pastötter et al., 2011, 2018; Szpunar et al., 2013; Yang et al., 2017). Evidence for this view comes from research on the FT effect for word lists as well as video lectures. In one study, for instance, Pastötter et al. (2011) examined brain oscillations during encoding of five lists of unrelated words. In the absence of interim retrieval practice, an increase in alpha power (8–14 Hz) was observed from the first to the final critical list, a finding that has been linked to memory load (O. Jensen et al., 2002) and inattention (Palva & Palva, 2007). In the presence of interim retrieval practice, however, no such increase in alpha power emerged, a finding that has been attributed to an encoding “reset” that may make the encoding of later lists as effective as the encoding of early lists (Pastötter et al., 2011; see also Kliegl & Bäuml, 2023; Pastötter et al., 2018). Additional evidence for the view, for instance, comes from Szpunar et al. (2013), who found that interim retrieval practice encouraged note-taking and discouraged mind wandering during study of a video lecture about an introductory course to statistics. Similarly, in another study, Jing et al. (2016) showed that interim retrieval practice increased the frequency of lecture-related thoughts (e.g., thoughts related to an earlier part of the lecture) and decreased the frequency of lecture-unrelated thoughts (e.g., zoning out) while subjects were watching a video lecture on public health. Taken together, these studies suggest that retrieval practice may restore attentional resources and redirect them to the encoding task, both when subjects study simple word lists (Kliegl & Bäuml, 2023; Pastötter et al., 2011, 2018) and when they study video lectures (Jing et al., 2016; Szpunar et al., 2013).

To the best of my knowledge, no study has yet examined the role of attentional factors in the FT effect with prose passages, leaving it open whether attention restoration may have contributed to the FT effect in the present studies. The present experiments were not designed to examine attentional accounts of the FT effect. However, attention restoration alone cannot easily account for the present finding that the FT effect with related passages reflected proactive facilitation (Experiments 2, 3, 4, and 5), and it has also difficulty explaining why, with unrelated passages, an FT-like effect arose after semantic generation but not after concept mapping (Experiments 1 and 6), while with related passages, an FT-like effect showed up after concept mapping but not after semantic generation (Experiments 2, 4, and 5). Indeed, if attention restoration alone mediated the FT effect with related passages, recall in the retrieval-practice condition

should have been similar but not superior to the level of recall in the final-passage-only condition, given that attention restoration should have made encoding of the final material as effective as encoding when no prior study has taken place (Pastötter et al., 2011, 2018). Furthermore, if semantic generation and concept mapping induced FT-like effects because they released attentional resources, one should expect the two interim tasks to create such effects for both related and unrelated passages, which is not what the present results showed.

Word-list studies provided evidence that attentional factors may provide an additional contribution to the FT effect that goes beyond context change and strategy change (Chan et al., 2020; Kliegl & Bäuml, 2023). Attentional factors may thus also provide an additional contribution to the FT effect with prose passages, but there is reason to assume that attention restoration and facilitated comprehension may not independently contribute to better encoding of the final critical passage. Prior work indicated that text-based inferences made while reading depend critically on whether the text information needed for such inferencing is accessible in working memory (Wiley & Myers, 2003). If so, knowledge integration across related prose passages may require that attentional resources are restored and refocused to the encoding task, enabling subjects to concurrently consider the preceding information during encoding of the final critical discourse. Consistently, in Jing et al.'s (2016) study, engaging in lecture-related thoughts during study was positively related to the amount of integration across the lecture parts, a finding that may indicate that focusing attention to the lecture contents indeed facilitated their comprehension. Future research may examine the potential role of attentional factors in the FT effect with prose passages, and also the potential interplay of attention restoration and facilitated comprehension.

5.2.5 Relation to Prior Studies on the Role of Release From Proactive Interference for the Forward Testing Effect

The present experiments provide demonstration that absence of intrusions does not necessarily imply the absence of proactive interference. Indeed, with unrelated prose passages, intrusions from the preceding passages did not emerge irrespective of condition, but comparisons of recall rates to the final-passage-only condition clearly indicated

buildup of proactive interference in the restudy condition and release from such proactive interference in the retrieval-practice condition (Experiments 1, 3, and 6; for a discussion, see chapter 3.2.3). Hence, it can be problematic to infer from the absence of intrusions that the FT effect does not reflect recovery from proactive interference. However, conceptually, it is the same conclusion made in other studies on the role of release from proactive interference for the FT effect (e.g., Ahn & Chan, 2022; Dang et al., 2021; Yang et al., 2022).

In one recent study, for instance, Ahn and Chan (2022) aimed to provide a direct test of whether release from proactive interference is at the core, or merely a byproduct, of the FT effect. In their Experiment 2, subjects were asked to study four lists of categorially related words, and it was manipulated whether, between study of the single lists, subjects engaged in restudy or retrieval practice of the preceding list. Critically, it was also manipulated whether different words from the same four categories were presented both within and across the lists, or whether the categories were switched from list to list, such that each list comprised different words from a single category only. Ahn and Chan found the typical FT effect with higher recall of critical list 4 in the retrieval-practice than the restudy condition. More important, the researchers found that prior-list intrusions during recall of list 4 were reduced to near floor levels for subjects in the category-switch condition, but despite of this, the magnitude of the FT effect was unaffected by the category-switch manipulation (for similar procedure and results, see also Ahn & Chan, 2022, Experiment 1). Because a category switch can induce release from proactive interference by enabling a more focused memory search (Gardiner et al., 1972; Wixted & Rohrer, 1993), and such release did likely emerge as indicated by the near absence of intrusions in response to the category-switch manipulation, the authors interpreted their results as evidence against the idea that release from proactive interference is at the core of the FT effect.

While conclusions on the role of release from proactive interference for the FT effect are often drawn from intrusion-based examinations (e.g., Ahn & Chan, 2022; Dang et al., 2021; Yang et al., 2022), caution is warranted because intrusions do not capture the whole range of cognitive processes that have been linked to proactive interference. One may therefore argue, for instance, that the results reported by Ahn and Chan (2022) simply demonstrate that focusing the memory search to only the critical list is not the sole

mechanism contributing to release from proactive interference and the FT effect. To illustrate, prior work indicated that a category switch does not restore attentional resources (Engle, 1975), which have been shown to decrease across successive study lists and thus create proactive interference (Crowder, 1976; Pastötter et al., 2011). If so, in Ahn and Chan's Experiment 2, the category switch may have led to partial release from proactive interference due to a more focused memory search, resulting in the near absence of prior-list intrusions. The additional retrieval practice then may have led to complete elimination of proactive interference due to, for instance, attention restoration, resulting in an FT effect in recall of list 4 (see chapter 5.2.4). Consistent with this line of reasoning, recall of list 4 in the restudy condition was lower than recall of list 1 in the retrieval-practice condition (45% vs. 63%), indicating that the category-switch manipulation did not fully prevent the buildup of proactive interference from encoding of list 1 to encoding of list 4 in the restudy condition.

Ahn and Chan (2022) argued that “adopting this definition of proactive interference for the FT effect would render the release-from-proactive-interference account purely descriptive (rather than explanatory), because whenever an FT effect is found, by definition, participants in the tested [i.e., retrieval-practice] condition must outperform those in the control [i.e., restudy] condition” (p. 1679). The present results weigh against this reasoning, showing that interim retrieval practice can sometimes enhance recall of critical material even relative to a condition without preceding study of further material (Experiments 2, 3, 4, and 5; see also Wissman et al., 2011), an effect that cannot be accounted for by release from proactive interference. Adopting a broad definition of proactive interference is thus essential to evaluate whether the FT effect reflects recovery from memory impairment – as opposed to a “true” beneficial effect on learning, as it is reflected in the finding of proactive facilitation. Future research may examine the role of release from proactive interference for the FT effect adopting a more comprehensive definition of proactive interference and relying not only on intrusions to measure amount of proactive interference.

5.3 Implications for Educational Practice and Research

The present findings are potentially important for educational practice and provide some useful clues with regard to the putative courses scenarios raised in the introduction. First, retrieval practice can benefit learning and memory of future material both when highly related and when unrelated materials are studied, suggesting that, regardless of whether students, during exam preparation, are facing the study of highly related materials from the same course or quite unrelated materials from different courses, in-between retrieval practice can optimize new learning. Second, with unrelated materials, future learning can also benefit from retrieval of other information than the just studied material, so that setting up the evening grocery list between study of the single materials may also improve learning of forthcoming materials. However, and in contrast to interim retrieval practice, doing so may come with costs and not improve memory for the preceding materials (see Divis & Benjamin, 2014). Third, when students are facing the study of highly related materials, it does not pay off to skip the preceding materials, as in such case, learning and memory for the final materials may be much worse than when all preceding materials were attended and retrieval practiced. Finally, with highly related materials, future learning can also be aided when, instead of interim retrieval practice, elaborative restudy practice with concept mapping is applied in-between the study. More generally speaking, learning and memory of future material may benefit from interspersing study with practice tasks that can promote the students' knowledge on preceding related materials. Hence, when courses are dealing with highly related contents, educators are recommended to insert active practice on the preceding discourse, be it by virtue of retrieval practice or concept mapping. When courses are dealing with unrelated contents, however, educators are encouraged to insert retrieval activities, be it retrieval of preceding course materials or retrieval of information unrelated to the educator's discourse.

Although there is an extensive research base on retrieval practice (see Karpicke, 2017; Yang et al., 2021) and concept mapping (see Nesbit & Adesope, 2006; Schroeder et al., 2018), few studies have directly compared the effects of the two types of practice on learning, and, to the best of my knowledge, those studies were focused on learning of the practiced material itself. In most cases, elaborative studying with concept mapping

was found to be less effective than retrieval practice in promoting long-term retention (e.g., Karpicke & Blunt, 2011; Lechuga et al., 2015), unless concept mapping was employed as a retrieval-practice task (e.g., Blunt & Karpicke, 2014; Ortega-Tudela et al., 2019). The present findings provide first evidence that, with related prose materials, concept mapping can promote the learning of subsequently studied, new material to a similar degree to how retrieval practice does, with concept mapping being employed as a pure elaborative-study task without any instructions to retrieve the to-be-studied information. Given the widespread popularity of concept mapping in educational settings (e.g., Hay et al., 2008; Quinn et al., 2003), these findings may have important practical implications, suggesting that one effective way to implement elaborative studying with concept mapping is to apply it in-between the study of related prose materials.

While the present findings suggest that, for related course materials, retrieval practice and concept mapping may both enhance future learning, they leave it open whether the two types of practice may be differently effective with different groups of learners. In fact, retrieval practice can be ineffective, or even counter-effective, for some students, as classroom studies on the backward testing effect indicated (Carpenter et al., 2016; Karpicke et al., 2014). In one study, for instance, Carpenter et al. (2016) found that students who had achieved a high overall performance in an introductory biology course benefited more from retrieving than copying course material at a later knowledge quiz, whereas students who had achieved a low overall performance benefited more from copying than retrieving. In addition, high-performing students recalled more information than low-performing students during initial retrieval practice, and across achievement levels, there was a positive relationship between initial retrieval success and later scores on the knowledge quiz. The findings thus allude to the possibility that the backward testing effect may be absent for low-performing students in classroom settings due to poor initial retrieval success (Carpenter et al., 2016; see also Karpicke et al., 2014).

To the best of my knowledge, no study has yet examined whether the FT effect also varies in size with student achievement. However, if the FT effect with related materials indeed resulted from practice-induced knowledge on preceding materials, one may speculate that the effect for related course materials may be more pronounced for high-performing students than for low-performing students in a course, just like the backward testing effect in classroom settings. Indeed, retrieval of previously studied

course material requires an active memory search, and students with high prior knowledge on the topic of a certain course may be better equipped to successfully retrieve the information than students with low prior knowledge, which may better promote not only memory for the practiced material but also future learning of related discourse. Concept-map-creation on preceding course material, in contrast, may boost future learning even when prior knowledge is low, given that elaborative studying with concept mapping requires processing of information that is already provided to the learner and thus should not depend much on retrieval success. If so, retrieval practice and concept mapping may be differently effective in enhancing future learning with different student achievement levels. Future research may address the issue and examine whether, for related course materials, type of practice inserted in the classroom should be tailored to student achievement in the course.

5.4 Possible Limitations of the Present Experiments

The present experiments employed materials for study that mimicked those employed in prior work on the FT effect. Experiments 1, 3, and 6 employed prose passages that covered different topics and were unrelated to each other, as it was done in Divis and Benjamin's (2014) study. Experiments 2, 3, 4, and 5 employed prose passages that covered the same topic, were directly related to one another, and were intended to be integrated, similar to the materials employed by Wissman et al. (2011) as well as by Szpunar and colleagues (Jing et al., 2016; Szpunar et al., 2013, 2014). On a hypothetical relatedness scale, these two types of materials reflect quite extreme points, raising the question of how results for the FT effect might look like for intermediate levels of relatedness. Wissman et al. (2011, Experiment 1a) employed such intermediate level using prose passages that shared a common topic but were otherwise not directly related to one another and had no overlapping information. Results showed an FT effect relative to a no-retrieval-practice condition, but the study did not include a final-passage-only condition, nor did it include a semantic-generation or a concept-mapping condition, so that a more detailed comparison of results with those from the present experiments is not possible. Still, Wissman et al. (p. 1146) raised doubt on whether topically related but

otherwise separate texts would lead to similar results than highly related texts, suggesting that, for such intermediate relatedness levels, the role of facilitated comprehension might be reduced. Clarifying the exact roles of context change and facilitated comprehension in the FT effect with intermediate relatedness levels in fact is a high priority for future work.

Another possible limitation of the present experiments is that the role of relatedness of study material for the FT effect was investigated via cross-experimental comparisons. Although procedural details were held constant to allow for direct comparison of results across experiments, it is still possible that not only relatedness of prose passages contributed to the differences in findings observed between the single experiments, but also the specific texts that were chosen as the experiments' final materials that may have differed, for instance, in recall difficulty. To address the point, Experiment 3 manipulated relatedness of prose passages within a single experiment, holding final passage 5 constant across the relatedness conditions. Results clearly replicated those of Experiments 1 and 2, suggesting that recall difficulty was not responsible for the difference in results between Experiments 1 and 2. However, because Experiment 3 did not include a semantic-generation or a concept-mapping condition, it cannot rule out the possibility that some of the final materials employed in the present experiments were, for some reason, not amenable to demonstrating FT-like effects in response to semantic generation or concept mapping, which might, in principle, have contributed to the differences in results observed between experiments.

Since prior research on the FT effect with more complex study material has been rather rare, the primary goal of the present experiments was to examine generalizability of the effect with these types of materials. For this reason, we opted to employ various examples of materials students may be confronted with in everyday educational contexts, as opposed to having the same text material throughout the entire set of experiments. Consistently, there are many previous studies in the literature that took a similar methodological approach, studies that addressed the learning of educationally relevant materials and examined generalizability of findings when the study material was changed in a series of experiments (e.g., Karpicke & Blunt, 2011; Roediger & Karpicke, 2006b; Wissman et al., 2011; Wissman & Rawson, 2015). Granted, the present dissertation is only a first attempt to explore the potential role of relatedness of material for the FT effect with prose materials, and to date, we do not know much about the specific preconditions

for the present findings to emerge. Future work is required to examine the role of study material for the observed differences in findings in a more fine-grained experimental approach.

5.5 Directions for Future Research

Most research on the FT effect has focused on the learning of relatively simple materials. Future work, both empirical and theoretical, is needed to explore the FT effect with more complex study material, materials as they were applied in the present experiments and are often found in educational practice. Empirically, such work may examine whether the effect with these types of materials generalizes from the laboratory context to authentic educational environments (for an example, see Todd et al., 2021). Moreover, because such environments typically contain learners with a wide range of achievement levels and prior knowledge on the topic of a certain course, classroom studies may examine whether the size of the FT effect with course materials can vary with student achievement and the effect be, for instance, more pronounced for high-performing students than for low-performing students in a course (see chapter 5.3). Theoretically, future work may investigate in more detail exactly which mechanisms mediate the FT effect with complex, educationally relevant materials. For instance, the present dissertation work suggests a link of facilitated comprehension to the general literature on knowledge effects on learning, but the exact cognitive processes underlying such facilitated comprehension still remain unclear (see chapter 5.2.3). Furthermore, the current two-factor account of the FT effect with prose materials may need revision, for instance, towards a more general three-factor account, considering that prior work pointed to a potential (additional) role of attentional factors in the effect with more complex material (Jing et al., 2016; Szpunar et al., 2013; for a discussion, see chapter 5.2.4). The FT effect with more complex materials is currently poorly understood, and a comprehensive discussion on directions for future research is clearly beyond the scope of the present work. However, since the present dissertation addressed the role of relatedness of study material for the FT effect with prose materials, some research questions are raised in the following that may serve to further

examine whether qualitatively different FT effects arise for related and unrelated prose materials.

5.5.1 The Potential Roles of Retention Interval and Lag

Does the FT effect with prose materials persist across prolonged retention interval and prolonged lag? The present experiments focused on immediate learning outcomes, leaving it open whether the effect changes in size over longer delay intervals between study and recall of the final material, as well as over longer lags between the study of the preceding materials and study of the final material. While the backward testing effect has been shown to be more pronounced when the criterion test is delayed (e.g., Kornell et al., 2011; Roediger & Karpicke, 2006b), the influence of retention interval and lag on the FT effect is largely unclear to date, in particular for more complex study materials. To the best of my knowledge, only three studies exist in the literature that addressed the issue to date. Two of the studies employed simple word lists as study material and found the FT effect to be largely maintained in size after a 25-min delay interval as well as after a 25-min lag with categorially related lists (Chan, Manley, et al., 2018; Kliegl & Bäuml, 2021b), but to disappear with such a delay interval or lag with unrelated lists (Kliegl & Bäuml, 2021b; see also Divis & Benjamin, 2014). In the other study, Wissman and Rawson (2015, Experiment 4) applied related prose passages and reported a substantial reduction of the FT effect from a 0-min to a 15-min delay, which may indicate that the effect with this type of material is relatively transient. In this study, all participants were asked to immediately retrieve the final passage before the delayed criterion test was administered, a procedure that endorses possible backward testing effects on retention of the final material and therefore does not allow to draw firm conclusions on the persistence of the FT effect. Future work may examine whether the FT effect with prose passages dissipates across prolonged retention interval or lag, and whether the answers to these questions vary with the relatedness of study materials.

On the basis of the suggested two-factor account of the FT effect with prose materials, one may expect a difference in persistence of the FT effect depending on relatedness of study material. Because, with unrelated prose materials, the FT effect should be due to context change mainly, the FT effect should be reduced or be even absent

for this material after both prolonged retention interval and prolonged lag. Context-change effects typically disappear with increasing delay (Abel & Bäuml, 2017, 2019; Divis & Benjamin, 2014), indicating that retrieval-induced context change should no longer enhance recall of the final material after prolonged retention interval. Furthermore, lagged learning of the final material should induce context change and thus enhance “list” discrimination also in the absence of interim retrieval (Estes, 1955; Mensink & Raaijmakers, 1988), suggesting that there should barely be any additional recall benefits of retrieval-induced context change after prolonged lag.

In contrast, because mainly facilitated comprehension should mediate the FT effect with related prose materials, there is reason to expect that the effect for this material may still be present after prolonged retention interval and lag. Differences in encoding typically translate into long-lasting differences in memory performance (e.g., Slamecka & McElree, 1983; Wixted, 2022), suggesting that the effects of facilitated comprehension on learning may be maintained across retention intervals. Moreover, since retrieval practice typically promotes the long-term retention of the practiced material (e.g., Kornell et al., 2011; Roediger & Karpicke, 2006b), retrieval-induced knowledge on preceding materials may still afford facilitated comprehension when learning of the final material is lagged. This holds while the exact pattern of results may critically depend on the cognitive processes underlying such facilitated comprehension (e.g., increased accessibility, Wahlheim, 2015; relational processing, Novak & Cañas, 2008), which may be differently effective in promoting knowledge integration with longer retention intervals or lags.

5.5.2 The Effects of Interim Retrieval Practice on Initially Studied Material

How does interim retrieval practice affect initially studied material? The present experiments focused on the effects of retrieval practice on the subsequent material. However, it would be important to know whether interim retrieval practice also induces effects on memory of the preceding material, such as, for instance, an initially studied but unpracticed prose passage. The issue has been addressed in only two prior studies to date. In the one study, Divis and Benjamin (2014) employed lists of unrelated words (Experiment 1) as well as unrelated prose passages (Experiment 2) as study material. The researchers found that, for both types of materials, memory of the last studied material

was improved but memory of the first studied material was impaired in response to an interim semantic-generation task, indicating that retrieval activities between study of materials can not only create benefits for the subsequent material but also come with costs for the initially studied material. In the other study, Kliegl et al. (2022) applied the standard interim retrieval-practice task and employed either lists of unrelated words or categorially related lists for study. Results showed a typical FT effect for the two types of lists, but no effect of interim retrieval practice was found on recall of an initially studied, unpracticed list, irrespective of whether unrelated or categorially related lists were used. No study has yet examined whether the Kliegl et al.'s observations generalize to more complex study materials. Future work may examine whether, with prose materials, the FT effect comes with cost effects on memory of initially studied material, and whether the answer to this question varies with the relatedness of study materials.

According to the two-factor account of the FT effect with prose materials, the effects of interim retrieval practice on initially studied material should depend on relatedness of study material. In particular, if the FT effect was primarily triggered by context change with unrelated material and by facilitated comprehension with related material, the FT effect should come with cost effects on memory of initially studied, unpracticed material with unrelated prose materials but not with related prose materials. Context change between study of materials typically impairs access to earlier material, which has been attributed to a mismatch between the context at encoding of the material and the context at test (Pastötter et al., 2008; Sahakyan & Kelley, 2002). As a result, retrieval-induced context change should not only enhance memory of the subsequent material, but also cause forgetting of the material studied prior to retrieval practice (cf. Divis & Benjamin, 2014). In contrast, knowledge gained from retrieval practice on study materials would only be able to affect encoding for material studied after retrieval practice. Retrieval-induced facilitated comprehension thus should enhance memory of the subsequent material, but it should not take effect on memory of initially studied, unpracticed material.

At first glance, the present results from the final cumulative tests seem to challenge the two-factor account of the FT effect, showing that interim retrieval practice did not induce recall impairment for any of the preceding passages, regardless of whether unrelated (Experiment 1) or related (Experiment 2) prose passages were presented for

study. These results should be interpreted with caution, however, because they likely do not provide pure measures of the effects of interim retrieval practice on initially studied material. First, in the present experiments, there was practice on all the preceding passages in the retrieval-practice and restudy conditions. Because retrieval practice on studied material can enhance its retention due to the act of retrieval (e.g., Roediger & Karpicke, 2006b), possible impairment effects on recall of the preceding passages may have been masked by enhancement effects resulting from retrieval practice of the passages. Second, output order of passages was not controlled in the final cumulative tests, which may have led to differential output order effects in the retrieval-practice and restudy conditions. Prior work on the FT effect recently reported that interim retrieval practice affects output order on the final cumulative recall test, making subjects more likely to organize their recall in the order of encoding (Yang et al., 2020). If so, final recall of passage 1 may reflect weak output interference in the retrieval-practice condition but strong output interference in the restudy condition, which may have biased results towards a higher level of passage-1 recall in the retrieval-practice than the restudy condition. For these reasons, the present results make it hard to draw conclusions on whether the FT effect with prose materials can be accompanied by cost effects. To clarify the issue, fresh experiments would be required in which the initial material is studied but not practiced and subjects at test are asked to recall the initially studied material first (cf. Kliegl et al., 2022).

5.5.3 The Potential Roles of Type of Retrieval Practice and Text Coherence

Is the FT effect with prose materials modulated by difficulty of learning? For instance, is the FT effect larger in size when the retrieval-practice task involves elaboration, or when less coherent versions of the study texts are applied? To the best of my knowledge, no study has yet examined whether the FT effect is affected by type of retrieval practice or text coherence. Nonetheless, some expectations may arise on the basis of research on the backward testing effect as well as on knowledge-enhanced learning.

Prior research on the backward testing effect has demonstrated that knowledge can be promoted to a greater extent with elaborative retrieval practice. In particular, for more complex study materials, retrieval practice has been shown to be more effective

when it entails not only simple recall of the studied material, but also requires subjects to construct explanations for the learning contents (Hinze et al., 2013), to link the recalled information to own experiences (Endres et al., 2017), or to answer higher-order questions that call for application, evaluation, and analysis of the studied information (J. L. Jensen et al., 2014). Because knowledge obtained from practice on studied materials may afford an FT effect for subsequent materials (see present Experiments 4 and 5), not only the backward testing effect but also the FT effect could be more pronounced with more versus less elaborative retrieval practice.

Aside from the type of retrieval practice that is applied, another important factor for the difficulty of learning is level of text coherence. Prior research has shown that the effects of prior knowledge on learning can be modulated by text coherence. Indeed, subjects with high prior knowledge on a given topic show the usual advantage over subjects with low prior knowledge when the learning task requires acquisition of a low-coherence text on that topic. This advantage, however, can be reduced or even disappear when the coherency of the text is increased (e.g., by adding connective links between sentences or replacing pronouns with noun phrases) (McNamara, 2001; McNamara et al., 1996; McNamara & Kintsch, 1996). Because the present findings point to parallels between effects of prior knowledge on learning and the FT effect (see chapter 5.2.3), text coherence could also affect the size of the FT effect and the effect be more pronounced with less versus more coherent text. Future work may examine whether type of retrieval practice and text coherence play modulating roles for the FT effect with prose materials, and whether the answers to these questions vary with the relatedness of study materials.

The present two-factor account of the FT effect suggests that the impact of difficulty of learning on the FT effect may depend on relatedness of study material. If primarily context change mediated the FT effect with unrelated prose materials, the effect may be largely unaffected for this material by type of retrieval practice as well as text coherence. Context change can be triggered by various forms of memory retrieval, including (episodic) retrieval of previously studied material, (semantic) retrieval of other information than the just studied material, and retrieval from autobiographical memory (Jang & Huber, 2008; Sahakyan & Kelley, 2002; Shiffrin, 1970). Substantial context change may thus be induced regardless of whether subjects engaged in more versus less elaborative retrieval, and retrieval-induced context change not depend much on the type

of retrieval practice. Furthermore, there is no obvious reason why retrieval-induced context change should depend on text coherence. In fact, interim retrieval tasks have been shown to be highly effective at inducing context change, irrespective of whether coherent text material or incoherent lists of simple words were employed for study (Divis & Benjamin, 2014).

In contrast, with related prose materials, the FT effect may be larger in size when elaborative retrieval practice or when low-coherence texts were applied, if mainly facilitated comprehension mediated the effect for this material. Asking subjects to engage in elaborative retrieval can induce them to construct a richer situation model of the learning contents (Endres et al., 2017; Hinze et al., 2013). Since situation-model processing is linked to inference performance (Graesser et al., 1994; Kintsch, 1994), subjects may be better equipped to fill the comprehension gaps during encoding of subsequent material when the preceding materials were practiced by elaborative retrieval. Moreover, low-coherence texts require use of prior knowledge to fill in the gaps, while high-coherence texts enable more ready comprehension and thus require less inferencing (McNamara, 2001; McNamara et al., 1996). Knowledge obtained from practice on preceding materials may thus afford a large advantage for encoding of the subsequent material when coherency of the text is low, but may afford a small advantage, if any, with high-coherence text material.

For the sake of completeness, it should be mentioned that the finding of a larger FT effect with more demanding retrieval practice or study texts would also be consistent with a strategy-change perspective. As previously stated, the results of the present series of experiments do not easily align with the strategy-change account of the FT effect (for a discussion, see chapter 5.1.2). To conclude from this that metacognitive insights play less of a role in the FT effect with prose materials may be premature, however. Prior work already indicated that the use of prior knowledge during learning reflects an active and effortful process that learners sometimes skip (McNamara et al., 1996; Nguyen & McDaniel, 2016). Increasing the difficulty of learning – for instance, by presenting poorly written texts or asking for elaborative retrieval – has been shown to encourage a switch in learning strategies and effectively guide subjects toward knowledge-based processing of the learning contents (Endres et al., 2017; Hinze et al., 2013; McNamara et al., 1996). If so, metacognitive insights may be critical for whether or not knowledge is applied for

facilitated comprehension and thus play a modulating role for the FT effect. Future work may address the issue and examine the potential interplay of metacognition and facilitated comprehension.

5.5.4 Potential Implications for Educational Practice

The research questions raised above may not only be of relevance for our theoretical understanding of the FT effect, but also for many applied situations in which students need to learn some course materials for upcoming exams at the end of the semester. First, during exam preparation, students often practice course materials in order to optimize long-term learning of the respective materials for an exam taking place in a few days or weeks. Furthermore, students in such situations often need to master vast amounts of course materials, so that sessions of study and practice on the single materials may be spaced some hours or even days apart. From a practical standpoint, it may thus be important to know whether interim retrieval practice can still promote learning and memory of future material when retention interval and lag are prolonged.

Second, when preparing for the final exams, students often engage in extended study sessions with the goal of optimizing learning for the entire material encountered in the course of the session. For this reason, it may be important to assess whether retrieval practice interspersed in a study session can come with costs and cause forgetting of materials encountered early in the session, or whether interim retrieval practice is still beneficial to learning when not only the later but also the earlier materials of a session are considered.

Third, students may sometimes find learning of course materials demanding, while at other times students may not feel sufficiently challenged by the course materials. Knowing whether interim retrieval practice is “always” beneficial to learning of forthcoming material may thus be an important practical issue, and also whether the type of retrieval practice applied by the students is critical for the benefits on future learning. For instance, it may be useful to know if future learning can be aided to a greater extent when students between study of course materials engage in elaborative retrieval, or whether similar benefits can be obtained with more simple – and potentially less time-consuming – types of retrieval practice. Moreover, course materials may differ in whether

they appear more or less coherent to the students, and it may therefore be of high practical relevance to know whether the benefits of interim retrieval practice depend on text coherence. For instance, retrieval practice may no longer improve learning of future material when some course material is too demanding for a certain student, considering that, in such situations, retrieval success can be very low (see e.g., Carpenter et al., 2016). Similarly, when some course material appears easy to understand because it is close to what a student already knows about the topic, the question arises whether it will still pay off to practice retrieval in-between the study session.

Last but not least, during exam preparation, students sometimes face the study of quite unrelated materials relevant to different courses and at other times the study of highly related materials relevant to a single course. Future research may address the issues raised above for both putative courses scenarios and come up with a more elaborate picture of how and when to apply interim retrieval practice as a valid learning tool.

5.6 Conclusions

The present dissertation work is the first to show that relatedness of study material can play a critical role for the FT effect with prose materials. In a series of six experiments, retrieval practice between study of prose passages was found to enhance recall of a final critical passage compared to interim restudy, regardless of whether the passages were related or whether they were unrelated in content. However, despite the presence of the same basic FT effect for the two types of materials, three differences in findings emerged between related and unrelated study materials. Only with unrelated passages did a similar recall benefit arise when subjects between study of passages engaged in retrieval of other (semantic) information than the just studied material, and only with related passages was recall of the final passage higher when the preceding passages were presented and retrieved than when no preceding passages were presented. Moreover, only with related passages was recall of the final passage enhanced to a similar degree when, instead of interim retrieval practice, elaborative restudy practice with concept mapping was applied as interim learning activity. The observed differences in findings between the two types

of materials point to qualitatively different FT effects for related and unrelated prose materials.

From a theoretical perspective, the findings are consistent with a new two-factor account of the FT effect with prose materials. According to this account, both context change and facilitated comprehension can contribute to the FT effect, but the effect with unrelated prose materials is primarily driven by context change and the effect with related prose materials is primarily driven by facilitated comprehension. From a practical perspective, the findings provide some clues for students on how to optimize new learning during exam preparation. When course materials are dealing with related content, learning of later materials may be aided by interspersing study with practice tasks that can promote the students' knowledge on the earlier materials, such as retrieval practice or concept mapping. In contrast, when course materials are dealing with unrelated content, learning of the later materials may be aided by interim retrieval tasks, be it retrieval of some earlier studied material or retrieval of information unrelated to the course materials.

Literature

- Abbott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, 11(1), 159–177.
- Abel, M., & Bäuml, K.-H. T. (2017). Testing the context-change account of list-method directed forgetting: The role of retention interval. *Journal of Memory and Language*, 92, 170–182.
- Abel, M., & Bäuml, K.-H. T. (2019). List-method directed forgetting after prolonged retention interval: Further challenges to contemporary accounts. *Journal of Memory and Language*, 106, 18–28.
- Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review*, 33(4), 1409–1453.
- Ahn, D., & Chan, J. C. K. (2022). Does testing enhance new learning because it insulates against proactive interference? *Memory & Cognition*, 50(8), 1664–1682.
- Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994a). How subject-matter knowledge affects recall and interest. *American Educational Research Journal*, 31(2), 313–337.
- Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994b). The influence of topic knowledge, domain knowledge, and interest on the comprehension of scientific exposition. *Learning and Individual Differences*, 6(4), 379–397.
- Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 237–313). Academic Press.

- Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 940–945.
- Aslan, A., & Bäuml, K.-H. T. (2016). Testing enhances subsequent learning in older but not in younger elementary school children. *Developmental Science*, 19(6), 992–998.
- Baddeley, A. D. (1990). *Human memory: Theory and practice*. Allyn & Bacon.
- Bäuml, K.-H. T., Abel, M., & Kliegl, O. (2020). Inhibitory processes in episodic memory. In M. W. Eysenck & D. Groome (Eds.), *Forgetting: Explaining Memory Failure* (pp. 125–146). Sage Publishing Ltd.
- Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language*, 68(1), 39–53.
- Bäuml, K.-H. T., & Schlichting, A. (2014). Memory retrieval as a self-propagating process. *Cognition*, 132(1), 16–21.
- Bäuml, K.-H. T., & Trißl, L. (2022). Selective memory retrieval can revive forgotten memories. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 119(8), 1–6.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637.
- Blumenthal, G. B., & Robbins, D. (1977). Delayed release from proactive interference with meaningful material: How much do we remember after reading brief prose passages? *Journal of Experimental Psychology: Human Learning and Memory*, 3(6), 754–761.
- Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, 106(3), 849–858.
- Bobbitt, T. (2018). *The forward effect of testing on comprehending complex texts* [Doctoral dissertation, University of Oklahoma]. ShareOK.

- Boscolo, P., & Mason, L. (2003). Topic knowledge, text coherence, and interest: How they interact in learning from instructional texts. *Journal of Experimental Education*, 71(2), 126–148.
- Bower, G. H. (1972). Mental imagery and associative learning. In L. W. Gregg, *Cognition in learning and memory*. John Wiley & Sons.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, 11(6), 717–726.
- Bruning, R. H. (1968). Effects of review and testlike events within the learning of prose materials. *Journal of Educational Psychology*, 59(1, Pt.1), 16–19.
- Buchin, Z. L., & Mulligan, N. W. (2023). Retrieval-based learning and prior knowledge. *Journal of Educational Psychology*, 115(1), 22–35.
- Bufe, J., & Aslan, A. (2018). Desirable difficulties in spatial learning: Testing enhances subsequent learning of spatial information. *Frontiers in Psychology*, 9, Article 1701.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118–1133.
- Carpenter, S. K. (2023). Encouraging students to use retrieval practice: A review of emerging research from five types of interventions. *Educational Psychology Review*, 35(4), Article 96.
- Carpenter, S. K., Lund, T. J. S., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016). A classroom study on the relationship between student achievement and retrieval-enhanced learning. *Educational Psychology Review*, 28(2), 353–375.

- Chan, J. C. K., Manley, K. D., & Ahn, D. (2020). Does retrieval potentiate new learning when retrieval stops but new learning continues? *Journal of Memory and Language*, 115, Article 104150.
- Chan, J. C. K., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag: a strategy change perspective. *Journal of Memory and Language*, 102, 83–96.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135(4), 553–571.
- Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, 144(11), 1111–1146.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55–81.
- Chevron, M. P. (2014). A metacognitive tool: Theoretical and operational analysis of skills exercised in structured Concept Maps. *Perspectives in Science*, 2, 46–54.
- Cho, K. W., Neely, J. H., Crocco, S., & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *The Quarterly Journal of Experimental Psychology*, 70(7), 1211–1235.
- Crowder, R. G. (1976). *Principles of learning and memory*. Lawrence Erlbaum.
- Dallett, K., & Wilcox, S. G. (1968). Contextual stimuli and proactive inhibition. *Journal of Experimental Psychology*, 78(3, Pt.1), 475–480.
- Dang, X., Yang, C., & Chen, Y. (2021). Age difference in the forward testing effect: The roles of strategy change and release from proactive interference. *Cognitive Development*, 59, 1–11.
- Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, 91(1), 66–73.

- de Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and some recent findings. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory* (pp. 19–50). Wiley.
- Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning, but hinders prior learning. *Memory & Cognition*, 42(7), 1049–1062.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58.
- Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. Teachers College, Columbia University.
- Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction*, 49, 13–20.
- Engle, R. (1975). Pupillary measurement and release from proactive inhibition. *Perceptual and Motor Skills*, 41(3), 835–842.
- Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.
- Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, 62(3), 145–154.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 632–652.

- Frase, L. T. (1968a). Effect of question location, pacing, and mode upon retention of prose material. *Journal of Educational Psychology*, 59(4), 244–249.
- Frase, L. T. (1968b). Some data concerning the mathemagenic hypothesis. *American Educational Research Journal*, 5, 181–189.
- Frisch, W., & Meschede, M. (2021). *Plattentektonik: Kontinentverschiebung und Gebirgsbildung* [Plate tectonics: continental drift and mountain formation] (6th ed.). Wissenschaftliche Buchgesellschaft.
- Fritz, C. O., & Morris, P. E. (2015). Partset cuing of texts, scenes, and matrices. *British Journal of Psychology*, 106(1), 1–21.
- Gardiner, J. M., Craik, F. I., & Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 778–783.
- Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, 6(40), 1–104.
- Gobet, F. (2015). *Understanding expertise: A multi-disciplinary approach*. Palgrave.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371–395.
- Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, 19(1), 126–134.
- Hay, D., Kinchin, I., & Lygo-Baker, S. (2008). Making learning visible: The role of concept mapping in higher education. *Studies in Higher Education*, 33(3), 295–311.
- Hershberger, W. A., & Terry, D. F. (1965). Typographical cuing in conventional and programed texts. *Journal of Applied Psychology*, 49(1), 55–60.

- Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition*, 32(2), 336–350.
- Hintzman, D. L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition*, 38(1), 102–115.
- Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*, 69(2), 151–164.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, 10(5), 562–567.
- Hunt, R. R., & Rawson, K. A. (2011). Knowledge affords distinctive processing in memory. *Journal of Memory and Language*, 65(4), 390–405.
- Hunt, R. R., Smith, R. E., & Dunlap, K. R. (2011). How does distinctive processing reduce false recall? *Journal of Memory and Language*, 65(4), 378–389.
- Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology*, 8(2), 200–224.
- Jacoby, L. L., Wahlheim, C. N., & Kelley, C. M. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1282–1297.
- Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 112–127.
- Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test ... or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, 26(2), 307–329.

- Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. E. (2002). Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex*, 12(8), 877–882.
- Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied*, 22(3), 305–318.
- Johnstone, A. H., & Percival, F. (1976). Attention breaks in lectures. *Education in Chemistry*, 13(2), 49-50.
- Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. H. Byrne (Ed.), *Cognitive psychology of memory, Vol. 2 of learning and memory: A comprehensive reference* (pp. 1–26). Elsevier.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775.
- Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014). Retrieval-based learning: The need for guided retrieval in elementary school children. *Journal of Applied Research in Memory and Cognition*, 3(3), 198–206.
- Karpicke, J. D., Butler, A. C., & Roediger, H. L. III (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, 17(4), 471–479.
- Keppel, G. (1968). Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 172–213). Prentice-Hall.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49(4), 294–303.

- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, *115*, Article 101237.
- Kliegl, O., & Bäuml, K.-H. T. (2021a). Buildup and release from proactive interference – cognitive and neural mechanisms. *Neuroscience and Biobehavioral Reviews*, *120*, 264–278.
- Kliegl, O., & Bäuml, K.-H. T. (2021b). When retrieval practice promotes new learning – the critical role of study material. *Journal of Memory and Language*, *120*, 104253.
- Kliegl, O., & Bäuml, K.-H. T. (2023). How retrieval practice and semantic generation affect subsequently studied material: An analysis of item-level effects. *Memory*, *31*(1), 127–136.
- Kliegl, O., Bartl, J., & Bäuml, K.-H. T. (2024). The pretesting effect comes to full fruition after prolonged retention interval. *Journal of Applied Research in Memory and Cognition*, *13*(1), 63–70.
- Kliegl, O., Kriechbaum, V. M., & Bäuml, K.-H. T. (2022). The effects of interspersed retrieval practice in multiple-list learning on initially studied material. *Frontiers in Psychology*, *13*, Article 889622.
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*(2), 219–224.
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85–97.

- Kriechbaum, V. M., & Bäuml, K.-H. T. (2024, January 29). Retrieval Practice Can Promote New Learning With Both Related and Unrelated Prose Materials. Retrieved from osf.io/dg5n3.
- Kriechbaum, V. M., & Bäuml, K.-H. T. (2025, February 28). Retrieval Practice “versus” Elaborative Studying With Concept Mapping – Both Promote New Learning With Related Prose Materials. Retrieved from osf.io/bckyn.
- Lechuga, M. T., Ortega-Tudela, J. M., & Gómez-Ariza, C. J. (2015). Further evidence that concept mapping is not better than repeated retrieval as a tool for learning from texts. *Learning and Instruction, 40*, 61–68.
- Lee, H. S., & Ahn, D. (2018). Testing prepares students to learn better: The forward effect of testing in category learning. *Journal of Educational Psychology, 110*(2), 203–217.
- Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(6), 1787–1794.
- Lindquist, S. I., & McLean, J. P. (2011). Daydreaming and its correlates in an educational environment. *Learning and Individual Differences, 21*(2), 158–167.
- Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition, 43*(1), 85–98.
- Malis, G. (1970). *The priority effect in the A-B, A-C paradigm* [Unpublished doctoral dissertation]. University of Toronto.
- Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*(3), 679–690.
- McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science, 20*(4), 516–522.

- McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55(1), 51–62.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1–43.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22(3), 247–288.
- Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, 95(4), 434–455.
- Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). Five popular study strategies: Their pitfalls and optimal implementations. *Perspectives on Psychological Science*, 13(3), 390–407.
- Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory*, 24(2), 257–271.
- Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76(3), 413–448.
- Nguyen, K., & McDaniel, M. A. (2016). The JOIs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology: Applied*, 22(1), 59–71.
- Novak, J. D., & Cañas, A. J. (2008). *The theory underlying concept maps and how to construct and use them*. Technical Report IHMC CmapTools 2006–01 Rev 01–2008, Florida Institute for Human and Machine Cognition.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press.

- Ortega-Tudela, J. M., Lechuga, M. T., & Gómez-Ariza, C. J. (2019). A specific benefit of retrieval-based concept mapping to enhance learning from texts. *Instructional Science*, 47(2), 239–255.
- Palva, S., & Palva, J. M. (2007). New vistas for α -frequency band oscillations. *Trends in Neurosciences*, 30(4), 150–158.
- Pastötter, B., & Bäuml, K.-H. T. (2007). The crucial role of postcue encoding in directed forgetting and context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 977–982.
- Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology*, 5, 286.
- Pastötter, B., & Bäuml, K.-H. T. (2019). Testing enhances subsequent learning in older adults. *Psychology and Aging*, 34(2), 242–250.
- Pastötter, B., Bäuml, K.-H. T., & Hanslmayr, S. (2008). Oscillatory brain activity before and after an internal context change: Evidence for a reset of encoding processes. *NeuroImage*, 43(1), 173–181.
- Pastötter, B., Engel, M., & Frings, C. (2018). The forward effect of testing: Behavioral evidence for the reset-of-encoding hypothesis using serial position analysis. *Frontiers in Psychology*, 9, Article 1197.
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), 287–297.
- Pastötter, B., Weber, J., & Bäuml, K.-H. T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology*, 27(2), 280–285.
- Pierce, B. H., Gallo, D. A., & McCain, J. L. (2017). Reduced interference from memory testing: A postretrieval monitoring account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(7), 1063–1072.

- Putnam, A. L., Sungkhasettee, V. W., & Roediger, H. L. III. (2016). Optimizing learning in college: Tips from cognitive psychology. *Perspectives on Psychological Science, 11*(5), 652–660.
- Pyc, M. A., & Rawson, K. A. (2012). Why is test–restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 737–746.
- Quinn, H. J., Mintzes, J. J., & Laws, R. A. (2003). Successive concept mapping: Assessing understanding in college science classes. *Journal of College Science Teaching, 33*(3), 12–16.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163.
- Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology, 97*(1), 70–80.
- Rea, S. D., Wang, L., Muenks, K., & Yan, V. X. (2022). Students can (mostly) recognize effective learning, so why do they not do it? *Journal of Intelligence, 10*(4), 1–28.
- Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday attention: Variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology, 26*(2), 234–242.
- Roediger, H. L. III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27.
- Roediger, H. L. III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210.
- Roediger, H. L. III, & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.
- Rohrer, D. (1996). On the relative and absolute strength of a memory trace. *Memory & Cognition, 24*(2), 188–201.

- Rothkopf, E. Z. (1966). Learning from written instructive material: an exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, 3, 241–250.
- Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, 58(1), 56–61.
- 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(6), 1064–1072.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353.
- Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing concept maps: A meta-analysis. *Educational Psychology Review*, 30(2), 431–455.
- Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science*, 168(3939), 1601–1603.
- Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(3), 384–397.
- Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, 6(4), 342–353.
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, 8(2), 203–220.

- Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language*, 73, 99–115.
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, 30(9), 641–656.
- Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, 3(3), 161–164.
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America*, 110(16), 6313–6317.
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1392–1399.
- Taylor, P. D. (2017). Fake fossils by the hundred – the story of Johann Beringer’s “lying-stones”. *Deposits Magazine* (52), 38–41.
- Tempel, T., & Frings, C. (2019). Testing enhances motor practice. *Memory & Cognition*, 47(7), 1270–1283.
- Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology*, 81(2), 264–273.
- Thorndike, E. L. (1914). Communications and discussion. Repetition versus recall in memorizing vocabularies; The effect of continuous exercise and of rest upon difficult mental multiplication. *Journal of Educational Psychology*, 5(10), 596–599.

- Todd, K., Therriault, D. J., & Angerhofer, A. (2021). Improving students' summative knowledge of introductory chemistry through the forward testing effect: examining the role of retrieval practice quizzing. *Chemistry Education Research and Practice*, 22, 175–181.
- Tricio, J., Salles, P., Orsini, C., & Aravena, J. (2023). Effect of an online module designed to promote first-year dental students' evidence-based study strategies. *European Journal of Dental Education*, 27(3), 428–437.
- Tullis, J. G., & Maddox, G. B. (2020). Self-reported use of retrieval practice varies across age and domain. *Metacognition and Learning*, 15(2), 129–154.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson, *Organization of memory*. Academic Press.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53(1), 1–25.
- Tulving, E., Patterson, R. D., & Malis, G. (1967). *[Total recall in AB-AD and AB-CD paradigms]* [Unpublished raw data]. University of Toronto.
- Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning & Verbal Behavior*, 13(2), 181–193.
- Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, 64(1), 49–60.
- Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, 105(1), 33–57.
- Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition*, 43(1), 27–38.
- Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive reminders in proactive effects of memory. *Memory & Cognition*, 41(1), 1–15.

- 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.
- Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 1039–1048.
- Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face–name learning. *Psychonomic Bulletin & Review*, 18(3), 518–523.
- Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology*, 132(4), 408–428.
- Wiley, J., & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes*, 36(2), 109–129.
- Wissman, K. T., & Rawson, K. A. (2015). Grain size of recall practice for lengthy text material: Fragile and mysterious effects on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 439–455.
- Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, 18(6), 1140–1147.
- Witherby, A. E., & Carpenter, S. K. (2022). The rich-get-richer effect: Prior knowledge predicts new learning of domain-relevant information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(4), 483–498.
- Wixted, J. T. (2022). Absolute versus relative forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(12), 1775–1786.
- Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1024–1039.

- Yang, C., Chew, S.-J., Sun, B., & Shanks, D. R. (2019). The forward effects of testing transfer to different domains of learning. *Journal of Educational Psychology*, *111*(5), 809–826.
- Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435.
- Yang, C., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied*, *23*(3), 263–277.
- Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *npj Science of Learning*, *3*, 8.
- Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 485–492.
- Yang, C., Sun, B., Potts, R., Yu, R., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? *Journal of Experimental Psychology: Applied*, *26*(4), 724–738.
- Yang, C., Zhao, W., Luo, L., Sun, B., Potts, R., & Shanks, D. R. (2022). Testing potential mechanisms underlying test-potentiated new learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(8), 1127–1143.
- Yue, C. L., Soderstrom, N. C., & Bjork, E. L. (2015). Partial testing can potentiate learning of tested and untested material from multimedia lessons. *Journal of Educational Psychology*, *107*(4), 991–1005.