Testing the blocking and context change accounts of Retrieval-induced Forgetting using item recognition

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

vorgelegt von

Julia Rupprecht

aus Weiden i.d.OPf.

Regensburg 2016

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

Zweitgutachter: Prof. Dr. Klaus W. Lange

Acknowledgement

First and foremost, I want to thank my supervisor, mentor, and initiator of this thesis, Prof. Dr. Karl-Heinz Bäuml, for sharing his scientific expertise, supporting me, for his patience with me, his availability at any time for my questions and concerns, and for creating such a pleasant atmosphere to work in. I admire him for his well-balanced lifestyle handling both his job and his family with great care, his structured way of approaching problems, and his genuine interest in and profound knowledge of human memory.

I also want to thank Lisa Wallner, Eva Lehmer, Ina Dobler, Magdalena Abel, Sarah Pötschke, Oliver Kliegl, Bernhard Pastötter, Michael Wirth, Christoph Holtermann, Andreas Schlichting, and Alp Aslan for helpful professional as well as private advice, for delightful conversations at lunch, and for making the time at the institute a precious experience I don't want to miss. Further thanks go to Petra Witzmann and Schoki who took care of any administrative barriers and readily engaged in dog-related discussions.

Likewise, I want to give credit to the diligent research assistants who helped gathering the data and, of course, all the students who were willing to partake in the present study. Without their participation, the present thesis would not have been possible.

Special thanks go to my partner Lion who supported me throughout the last five years by giving me advice, lifting my spirits, and encouraging me in everything I do. I also want to thank my parents for their support and assistance - especially my dad's expertise in Excel facilitated my work tremendously. Last but not least, I want to thank my brother, my dog Sammy, and my dear friends for always being there for me.

Preface

While the functioning of the human memory is often compared to the functioning of recording devices or computers (see Pear, 1922; Posner & Warren, 1972; Simon & Feigenbaum, 1964), everyday experience as well as laboratory work indicates that we are not able to rewind and perfectly reproduce every single information that we encountered earlier. Instead, we forget. At first sight, it might seem likely that all instances of forgetting are caused by the same single mechanism, which time may represent a suitable candidate for. It has been reliably found that the more time passes by between the last encounter of a particular set of information and the moment in which it should be retrieved, the less information will be successfully reproduced (e.g., Ebbinghaus, 1885). Over time, encoded traces are supposed to become weaker such that, after a certain amount of time, the information cannot be recovered again. However, empirical evidence suggests that time itself may not be the primary driving force of forgetting but that two other mechanisms have a crucial role in forgetting: interference and context change.

Over the last century, it has been robustly established that memory for information is affected by encoding of other, related information, i.e., currently irrelevant information interferes with the retrieval of currently relevant information. In a prominent experiment by Müller and Pilzecker (1900), participants first studied a list of nonsense syllable pairs for a later test. After study of the first list, either a second list of syllable pairs was presented or not, and, after an equivalent amount of time after study, the test was administered. Participants who engaged in the study of a second list recalled the first list of syllables at a lower rate compared to participants who did not study a second list. This finding provides a demonstration that other encoded information can block retrieval of particular target information. This associative blocking is based on the idea that encoded information is accessible via cues, i.e., providing information that is associated with the target information can aid retrieval. When such a cue is given, activation is supposed to spread from the cue to the target information, and if the level of activation is sufficient to exceed a particular threshold, the target information will be retrieved. According to a blocking-based explanation of forgetting, forgetting arises when the cue is associated to multiple traces in memory. Then, in face of the cue, all traces are activated and compete for retrieval which decreases the chances of recalling a particular, sought-after information. Furthermore, this blocking process supposedly depends on the strength of the association between information and cue. More strongly represented information may be recalled preferentially and prior to weaker information, representing a stronger source of interference and blocking.

Apart from interference by other information, memory is reliably affected by changes in context. Most prominently, the role of context in memory has been demonstrated by Godden and Baddeley (1975): Participants studied a list of items either on land or underwater. Then, participants were tested on the list either in the study environment or in the alternative environment, i.e., when they had been studying underwater they were tested on land and vice versa. The results showed that a mismatch between study and test environment decreased memory performance. This indicates that changes in the context between encoding and test impede retrieval while a constant context aids retrieval. According to a context change explanation of forgetting, information is encoded together with the context it is presented in. Thus, the context itself is a valid cue to assist retrieval of relevant information. If the context at test is sufficiently similar to the context during encoding, retrieval will be successful. In contrast, if the context at test differs substantially from the context during study, the chances to retrieve the sought-after information decrease.

Both mechanisms, interference and context change, are appealing due to their simplicity, their empirical robustness, and their successful application to a variety of memory phenomena (see Chapter 1.1). More than twenty years ago, M. C. Anderson, Bjork, and Bjork (1994) found that when a subset of information is retrieved, memory for related information is impaired, and termed the observed effect *Retrieval-induced Forgetting*. To this point, it is not clear whether this phenomenon is merely a type of interference-based or context-based forgetting or whether another mechanism is involved. The blocking account of Retrieval-induced Forgetting proposes that the retrieved information is more strongly associated to the retrieval cue compared to non-retrieved information. Thus, the weaker, non-retrieved information is blocked in a later test. According to the context change account, the act of retrieval induces a context change, decreasing the congruence of context cues during encoding and test. This contextual mismatch is supposed to be responsible for Retrieval-induced Forgetting.

Originally, however, the phenomenon was attributed to a different mechanism: inhibition. According to the inhibition account, Retrieval-induced Forgetting arises because non-retrieved information may interfere with retrieval and is therefore inhibited in order to grant retrieval success. Inhibition is supposed to last and may thus show later in impaired memory for the non-retrieved information. Some researchers view the idea of inhibition with skepticism due to its relation to Freud's (1915) controversial concept of repression and due to the presence of seemingly better alternatives. Well-established, simply structured mechanisms like blocking and context change might seem favorable to account for this forgetting effect, yet the results on their accountability for Retrieval-induced Forgetting are mixed.

The present thesis is dedicated to contribute to the comprehension of the mechanisms underlying Retrieval-induced Forgetting. In six experiments, predictions by the blocking account (Experiments 1-3) and the context change account (Experiments 4-6) are tested and the findings are discussed in terms of compatibility with each of the three accounts. In the first chapter of this dissertation, potential factors underlying forgetting in episodic memory are presented, i.e., interference, context, and inhibition. Then, Retrieval-induced Forgetting and related empirical findings are introduced. The three most prominent accounts with the objective of explaining Retrieval-induced Forgetting are outlined and their consistency with previous findings is discussed. In the second chapter, the goals of the present experiments are derived from the current state of research and the account-specific predictions while in Chapters 3 and 4 methods and results of the particular experiments are reported. Finally, the results are summarized and discussed, and conclusions are drawn on the validity of the three accounts of Retrieval-induced Forgetting in light of the present findings.

Contents

Parts of the present thesis are published as:

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

Rupprecht, J., & Bäuml, K.-H. T. (2016). Retrieval-induced versus context-induced forgetting: Can restudy preceded by context change simulate retrieval-induced forgetting? (*Submitted manuscript*)

Abstract

Retrieval-induced Forgetting (RIF) refers to the finding that practicing information by retrieval impairs recall for related, but not practiced information. The underlying processes of RIF have not been conclusively identified yet, but three theories have prevailed: The inhibition account attributes RIF to long-lasting inhibition of the unpracticed items in order to facilitate retrieval of the practiced items. The blocking account attributes RIF to the disproportionately stronger associations of the practiced items impeding recall of the unpracticed items during the final test. Therefore, equivalent effects of retrieval practice and other practice methods that enhance the associations adequately are predicted. The context change account attributes RIF to a contextual mismatch between study and test for the unpracticed items reducing the likelihood to recall those items as the act of retrieval during practice is supposed to accelerate context drift. Therefore, equivalent effects of retrieval and other practice methods that are preceded by a contextual change are predicted. Here, these equivalence assumptions fundamental to the blocking and the context change account were tested by varying the final test format. The blocking account predicts that if retrieval induces forgetting in a test, then restudy formats should induce forgetting in that test as well. Analogously, the context change account predicts that restudy preceded by context change should mimic the effects of retrieval practice irrespective of the test format. In six experiments, it was investigated whether restudy formats or restudy preceded by context change affect cued recall and item recognition similar to how retrieval practice does. The results showed that all forms of

practice decreased recall of related, unpracticed items. In the recognition test, however, only retrieval but none of the other practice types induced forgetting. The finding conflicts with the equivalence assumptions fundamental to the blocking and the context change accounts, suggesting that neither blocking nor context change alone can account for the entirety of RIF findings. Rather, the results indicate a critical role of inhibition in RIF.

Chapter 1

Forgetting in episodic memory

1.1 Mechanisms underlying forgetting in episodic memory

Episodic memory refers to the collection of memories that are associated with a particular episode in time and space (Baddeley, 2001; Tulving, 1972, 1983). For example, episodic memory is engaged when one is at the mall trying to recollect all the items on the shopping list that was left on the kitchen counter. Everyday experience and empirical work concur, that episodic memory is not perfect but is subject to flaws and forgetting. Such flaws of memory include for example misattribution of information to an incorrect source (e.g., B. P. Allen & Lindsay, 1998; M. K. Johnson, Hashtroudi, & Lindsay, 1993) or the suggestibility to false memories (e.g., Loftus, 1992; Loftus & Palmer, 1974) which have been reliably shown in lab and field research. One can surely imagine that not all items on the list will be retrieved during the shopping trip, especially if the list contains more than a few items. Moreover, one may think that other items that the list did not contain were part of it, maybe thinking of items on the list from last week or one's kid assures that chocolate bars were on the list and one in fact becomes subjectively confident that chocolate bars were part of the grocery list even when they were not. It might be less problematic or annoying to buy additional items compared to realizing at home that essential items were forgotten, like the liquid detergent or the breakfast milk.

Subjective experience and early empirical evidence suggest that forgetting in episodic memory is negatively correlated with time. Most will probably agree that one has more trouble recalling the grocery list that was written a few days ago compared to a shopping list that was written just before leaving the house for the mall. The more time passes by since the last occurrence of an event or an information, the smaller is the likelihood that the information can be retrieved. This relationship has been demonstrated by Ebbinghaus (1885) in the prominent forget curve showing retention as a logarithmic function of time (see also Bahrick, 1984; Bahrick, Bahrick, & Wittlinger, 1975; Meeter, Murre, & Janssen, 2005). At first, time itself was held accountable for forgetting. With time, traces of the encoded information were believed to *decay* (Ebbinghaus, 1885), an idea also adopted by Thorndike (1914) in his *Law of Disuse*. According to the Law of Disuse, when information is not repeated or retrieved, the trace in memory gradually grows weaker with time, just like a muscle that is not exercised. However, it is difficult to find conclusive evidence in support of Trace Decay Theory showing that information is eventually lost beyond retention: Even though retrieval or recognition methods are used to test the contents of memory, retention output does not reflect memory as the entirety of stored information but only shows fragments. As will be discussed below, particular conditions of the test may determine whether stored information will be successfully retrieved or not. In contrast to Trace Decay Theory, it has been proposed that forgetting is rather a transient than a persistent phenomenon, and that all encoded memories can be recovered under beneficial circumstances. Empirical evidence indicated that, primarily, not time itself may be responsible for forgetting but the events that occur over time. Present memory research focuses on three mechanisms, i.e., interference, context change, and inhibition, which will be introduced and discussed in the following subsections.

Interference

The finding that encoding of novel information impairs retrieval of previously learned information (Müller & Pilzecker, 1900) fueled the idea that the amount of events occurring over time determine the capability to retrieve particular information, i.e., other information *interferes* with successful retrieval. Empirically this interference effect can be demonstrated by showing that when study of English-Italian vocabulary (e.g., WATER - *acqua*) is followed by study of English-French vocabulary (e.g.,WATER - *eau*), retrieval of the Italian equivalent (*acqua*) will suffer compared to when unrelated math tasks follow study, this holds even though the time frame between study of the original cue-item pair and the test is matched and thus time can be ruled out to be the driving force of forgetting. This interference effect has not only been observed when new information was presented after study of target items, termed *Retroactive Interference*, but also when it is presented before (Underwood, 1957), termed *Proactive Interference*. This effect is particularly prominent when the to-be-studied material is similar (Osgood, 1953). Moreover, the chance of retrieving a specific target decreases as more and more items are added to the same cue such as studying also the German, Russian, and Portuguese equivalents of WATER (see e.g., J. R. Anderson, 1974; J. R. Anderson & Reder, 1999; Roediger, 1973; Tulving & Pearlstone, 1966). But not only semantically related information can represent valid cues, also episodes in time and contextual features can mark a cue common to a set of items. Consequently, the more information is studied within a distinct episode the harder it should be to recall each single item. This prediction was corroborated by Watkins' (1975) *List-length Effect*, demonstrating that recall of individual items declines as a function of list length, i.e., the more items constitute a single list the smaller the chances to recall each single item indicating that items within a single list may interfere with each other.

The role of interference in forgetting has been in the focus of memory research for the majority of the 20th century (for reviews see M. C. Anderson & Neely, 1996; Crowder, 1976; Postman, 1971). These interference effects have been attributed to response competition and associative blocking. Among memory researchers, it is widely accepted that memory retrieval is driven by cues, i.e., during study, the information is associated with, for instance, other presented information (e.g., 'GROCERIES' as a cue for '*milk*') and can be accessed at a later point in time when an associated cue is presented. Accessibility of particular information, however, decreases if the cue is associated to other traces in memory as well. It is assumed that when the cue is presented, activation spreads to all associated traces which then compete for retrieval. Thus, the likelihood for retrieval of every single piece of information is a function of the number of linked items. This idea has first been captured in McGeoch's (1942) *Response Competition Theory*, seized as one factor in Melton and Irwin's (1940) two-factor account of memory, and revived in modern computational interference models, such as the ACT (*Adaptive Control of Thought*; J. R. Anderson, 1981, 1983) model and SAM (*the Search of Associative Memory*; e.g., Gillund & Shiffrin, 1984; Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1980, 1981) and REM (*Retrieving Effectively from Memory*; Shiffrin & Steyvers, 1997) models. The idea of interference is very attractive as it represents a simple explanation to a variety of seemingly different phenomena: Forgetting effects in learning of novel paired associates or in learning of multiple lists (see Pro- and Retroactive Interference), but also forgetting within a single list (see List-length Effect) can be embraced by a common underlying effect: interference (for a discussion of further memory phenomena caused by associative blocking, see Raaijmakers, 2008).

This *response competition* or *cue overload* is moreover presumed to be determined by the individual strength of particular cue-target associations such that more strongly associated targets interfere more vigorously when a weakly associated target has to be retrieved (Raaijmakers & Shiffrin, 1981; Rundus, 1973; Wixted, Ghadisha, & Vera, 1997). Conversely, weakly associated items hardly block retrieval of strongly associated items. Thus, according to Response Competition Theory, forgetting arises because the presented cue is associated to other information apart from the target information which block retrieval of the to-be-remembered item and forgetting is thus a by-product of newly added or strengthened information to a shared retrieval cue. Evidence in favor of the idea that blocking potential varies as a function of strength comes from the *List-strength Effect* (Tulving & Hastie, 1975; Ratcliff, Clark, & Shiffrin, 1990). In this design, a subset of items is granted more time for encoding, either by varying presentation times or by repetition, in order to manipulate the items' strength within the study list. It is typically found that these 'strong' items are recalled at a higher rate compared to items of a list, in which all items were presented for the same amount of time as the 'strong' items in the mixed list, whereas the 'weak' items are recalled at a minor rate compared to a list with exclusively weakly encoded items. This indicates that the strength of the other items associated to the cue determines the accessibility of a particular target.

Returning to the initial shopping list example, Response Competition Theory assumes that the cue GROCERIES is associated to many items in the current shopping list (e.g., *milk, sugar, liquid detergent, etc.*) and probably also to many items in previous shopping lists (*bread, coffee, apples, etc.*). Recall of any single item of the current list is thus object to interference by the other items from the current list and all previous lists. Therefore the milk may have been forgotten because even though one was sure that the list contained more than the few items in the shopping cart, when trying to recall *milk*, these other items may have intruded one's mind vigorously until eventually giving up. However, it may be more likely that one fails to buy unusual items like *liquid detergent* that are not part of the weekly grocery list and are thus not as strongly associated to the cue.

Context

In addition to the interference findings, studies using context manipulations represent a further indication that time by itself is not the primary cause of forgetting but that changes in context may contribute to forgetting. These studies consistently show that changes in contextual features between study and test, compared to constant contextual features, impair memory even though the time frame of conditions is matched. As already pointed out in the Preface, a renowned experiment by Godden and Baddeley (1975) demonstrates this phenomenon: Participants studied a list of target items in very different environmental surroundings, either on land or underwater. Participants then either recalled the list in the same environment, in which they had studied the list, or in the alternative environment, i.e., when the list was studied on land, participants had to recall it underwater, and vice versa. The results demonstrated that changing the environment between study and test reduced list recall dramatically. Not only external manipulations, such as changing the room (e.g., S. M. Smith, 1979; S. M. Smith, Glenberg, & Bjork, 1978) or item features (e.g., color, font, etc.; e.g., Dulsky, 1935; Isarida & Isarida, 2007) have been shown to induce forgetting, but also changes of the internal environment. For instance, when a positive or negative mood was induced prior to study and a mismatching mood was induced prior to test, retrieval performance decreased compared to when the mood was matched for study and test phases (e.g., Bower, 1981; Bower, Monteiro, & Gilligan, 1978; E. Eich, 1995; Macht, Spear, & Levis, 1977). Moreover, mind-wandering tasks between study and test phases have been shown to reduce memory performance (Delaney, Sahakyan, Kelley, & Zimmerman, 2010; Pastötter & Bäuml, 2007; Sahakyan & Kelley, 2002; Unsworth, Spillers, & Brewer, 2012). Mind-wandering refers to any mental activity that redirects the participant's attention to content unrelated to the current experimental situation and has been operationalized by asking participants to imagine their parents' house and mentally walk through it (e.g., Sahakyan & Kelley, 2002) or to imagine a recent vacation trip (e.g., Delaney et al., 2010).

This empirical evidence suggests that retrieval success is contingent on the quantity and quality of the contextual cues available. It is widely accepted that, with encoding of information, contextual information is stored as well. Contextual information refers to temporo-spatial information such as the episode in time, the appearance of the surroundings, objects, colors, sounds, or odors, to name a few, as well as internal states such as mood, cognitive states or strategies, physical states like pain or fatigue, or mind-wandering. Crucially, this contextual information fluctuates as time passes by - or due to experimental manipulations -, e.g., leaving or rearranging the room where the information was acquired or mood changing from joy to anxiety. While Response Competition Theory relies on the idea that forgetting arises due to cue overload, context theory ascribes forgetting to the absence of an appropriate contextual stimulus to cue the target (see Tulving, 1974).

Accordingly, forgetting increases over time as time is usually attended by changes in the environment and in state of mind. With these changes, the presently available contextual cues may grow less and less similar to the contextual cues that were originally associated with the target information, and therefore the target information cannot be accessed (Estes, 1955; Mensink & Raaijmakers, 1988). This idea has been captured in Tulving and Thomson's (1973) *Encoding Specificity Principle*: If contextual cues at the time of encoding and at the time of test sufficiently match, then the target information can be retrieved. If there is a mismatch between study context and test context, the information cannot be accessed.

This theory is further corroborated by the finding that reinstatement of the contextual encoding environment can improve retention after a context change (e.g., Sahakyan & Kelley, 2002; S. M. Smith, 1979; for reviews see J. E. Eich, 1980; S. M. Smith & Vela, 2001). Both, recreating the external surroundings of the original study context and mentally reconstructing the scene, helps participants to recover memories. Brinegar, Lehman, and Malmberg (2013) even showed that *preinstatement*, i.e., imagining the situation at test during encoding, can reduce the detrimental effects of contextual change. It has further been found that Retroactive Interference is significantly reduced when the to-be-studied lists are learned in distinct contexts (Bilodeau & Schlosberg, 1951; Dallett & Wilcox, 1968; Greenspoon & Ranyard, 1957) indicating that successful context attribution and differentiation can help to eliminate impairments caused by interference. It is worth noting that when information is studied within different contexts, the significance of the presence of particular contextual cues on recall decreases as the information can be accessed with multiple cues, i.e., the information is *decontextualized* (S. M. Smith, 1982, 1984). Decontextualized information is thus less prone to effects of context fluctuation.

With respect to the shopping list example, the recall of the grocery list may depend on the external and internal contextual cues. Items of the list may be forgotten because the cues present during the shopping tour may not adequately match the ones while writing the grocery list. Clearly, the surroundings in the supermarket are different from the ones at home. Moreover, the mood might have changed, as one may have been calm at home or excited to go shopping but now one might be bothered by the traffic jam on the road or because of the crowd in the mall. However, mere imagination of the contextual situation during writing the grocery list may aid recovery of the items missing in the cart.

Inhibition

Response competition seems to pose a particular problem: Only the strongest competitors will be retrieved while weaker associated information is lost. During attempts to retrieve a weak target information, highly accessible competitors are assumed to persistently intrude and thus disturb retrieval of the target information. In order to grant efficient functioning of memory however, selection of currently relevant responses should be prioritized and irrelevant or out-of-date responses should be neglected. One can easily imagine scenarios, in which a weaker response associated with a cue might be more appropriate. Consider again the shopping list scenario: While some items may be linked strongly to the cue GROCERIES because they are bought on a frequent basis, like *bread*, other items may only be weakly associated because one may buy them only once in a while, like *liquid detergent*. However, today, *liquid detergent* was on the grocery list. In order to accomplish efficient memory functioning, retrieval of the interfering competitors needs to be stopped.

In fact, empirical evidence supports the idea that retrieval of information can effectively be stopped. List-method Directed Forgetting provides an example of such motivated forgetting showing that the instruction to forget can indeed impair memory for the to-be-forgotten items (Geiselman, Bjork, & Fishman, 1983). Typically, participants study a first list of items and are then either instructed to remember the list for an upcoming test or to forget it using a pretext, e.g., because allegedly the wrong list was presented. Subsequently a second list of items is presented for study and finally a test on both lists is employed. The data usually show impaired retrieval of the first list and facilitated retrieval of the second list in the Forget condition compared to the Remember condition.

The Think/No-think Paradigm provides a further example of effective memory control. In the Think/No-think Paradigm (M. C. Anderson & Green, 2001), participants study a list of weakly associated pairs (e.g., pillow - termite) and are trained to respond with the corresponding item (termite) when the cue is presented (pillow). Next, in the Think/No-think stage, the cues are presented and participants are instructed to either recall the associate or to avoid thinking of it, i.e., to suppress the thought. Following several trials, a cued recall test is employed typically showing that items the participants were instructed to think of are recalled at a higher rate compared to the baseline while items participants were instructed to suppress are recalled at a lower rate (e.g., M. C. Anderson & Green, 2001; M. C. Anderson et al., 2004; Depue, Banich, & Curran, 2006; Hertel & Calcaterra, 2005). Thus it seems that an instruction to refrain from retrieval can impair later attempts to recall this information.

Proponents of inhibition assume that active, temporary suppression impairs memory for particular contents and overrides dominant responses to solve problems in the face of interference (see M. C. Anderson, 2003). It should be noted, however, that the concepts of the way inhibition is supposed to control memory differ in their specifics. When speaking of inhibition, it either involves a reduction in strength of the cue-target association or a reduction in strength of the memory trace itself. To accomplish this selective suppression, it is usually assumed that executive control processes are recruited and prefrontal areas of the brain are involved (e.g., M. C. Anderson, 2003; Kuhl, Dudukovic, Kahn, & Wagner, 2007). The idea of active inhibitory mechanisms is related to the Freudian (1915) concept of intentional repression of unwanted memories and thoughts and Wundt's (1902) reactive inhibition. Melton and Irwin (1940) included a suppression mechanism in their two-factor account, according to which forgetting arises due to associative blocking and 'unlearning' of memory traces. Unlearning refers to the weakening of the cue-item association when, by mistake, an item is retrieved during search for a different information. In order to stop retrieval of this irrelevant competitor, the cue-item association is weakened, and thus the item will not distract the search of the target any longer. Similar to the unlearning assumption, Geiselman et al. (1983) proposed that recall impairment for List-1 items arises because the access to the List-1 items is actively inhibited. List-2 items are supposed to benefit from the inhibition of List-1 items as the Proactive Interference by List-1 items is reduced (for alternative views, see Bjork, 1972; Sahakyan & Kelley, 2002).

In contrast to the unlearning assumption, Postman, Stark, and Frasier (1968) proposed that the item trace itself rather than the cue-item association is inhibited. According to their Response-set Suppression view, when a cue is paired with two responses, e.g., studying first WATER - *acqua* and then WATER - *eau*, the second response (*eau*) will be primed by a selector mechanism and the first response *acqua* will be suppressed. This response suppression is supposed to be automatic and transient, i.e., suppressed items are supposed to recover after a short period of time. The Think/No-think phenomenon has been attributed to inhibitory mechanisms (M. C. Anderson & Green, 2001) suggesting that during No-think-trials the trace of the particular response is actively inhibited, similar to Postman et al.'s response suppression theory. In contrast to the inhibition account of List-method Directed Forgetting, not only the access to the response is impaired, i.e., its association to the cue, but the trace itself is suppressed, suggesting that the observed impairment should not be restricted to the original cue. Therefore, participants should have difficulty remembering the No-think-items over a variety of tests (for alternative views, see Hertel & Calcaterra, 2005; Bulevich, Roediger, Balota, & Butler, 2006).

Returning to the shopping list example, when trying to retrieve an item that one does not buy each week, e.g., *liquid detergent*, other items that are frequently part of the shopping lists, e.g., *bread, apples* etc., may intrude one's mind. To stop these intrusive thoughts from distracting, inhibitory control may intervene to suppress the dominant items in order to make retrieval of the currently relevant items possible.

1.2 Retrieval-induced Forgetting

The idea that retrieval does not merely represent a passive reproduction of memory contents has found wide acceptance. The act of retrieval changes memory representations, producing both positive and negative effects on retention (Bjork, 1975). These negative effects are at the core of the phenomena of Output Interference and Retrieval-induced Forgetting. *Output Interference* refers to the finding that recall performance decreases with serial position within a single test, i.e., the probability of retrieving a previously studied item is smaller when the word is tested at a later point in the test compared to an earlier point (Roediger, 1973; Roediger & Schmidt, 1980; A. D. Smith, 1971, 1973; A. D. Smith, D'Agostino, & Reid, 1970; Tulving & Arbuckle, 1963, 1966). This finding provided first support for the claim that retrieval can induce forgetting and that retrieval itself is 'a self-limiting process' (Roediger, 1978). The Retrieval-practice Paradigm was inspired by work on Output Interference and was supposed to explore the effects of retrieval on long-term memory (M. C. Anderson et al., 1994). While in the Output Interference Paradigm the effects of retrieval on subsequent memory performance were tested within a single test session and thus only immediate effects could be measured, in the Retrieval-practice Paradigm retrieval is completed in a separate stage prior to the test phase with a variable retention interval in between. This procedure permits to identify potential long-term impairments induced by retrieval.

The Retrieval-practice Paradigm

The typical Retrieval-practice Paradigm comprises three phases (M. C. Anderson et al., 1994): In the study phase, participants learn a list of categorized items (e.g., FURNITURE - *lamp*, INSECT - *hornet*, INSECT - *termite*, etc.). In the retrieval-practice phase, a subset of items of a subset of categories is tested multiple times given the category cue and item-specific initials (e.g., INSECT - *te*___). After a retention interval, memory performance for all items is typically assessed using a cued recall test (e.g., INSECT - *h*___; FURNITURE - *l*___; INSECT - *t*___). To avoid confounding Output Interference effects, the unpracticed items from practiced categories are tested prior to practiced items, either altogether in the first list half or blocked by category; items from unpracticed categories are usually tested interjacently. The retrieval-practice phase, thus, creates three distinct item types: practiced items, designated *rp+* (e.g., *termite*); unpracticed items from practiced categories, denoted *rp-* (e.g., *hornet*); and items from utterly unpracticed categories, denoted *c* items (e.g., *lamp*). In line with the testing literature (e.g., G. A. Allen, Mahler, & Estes, 1969; Bjork, 1975; Carrier & Pashler, 1992; Gardiner, Craik, & Bleasdale, 1973; Karpicke & Roediger, 2008), it is typically found that recall performance for *rp+* items exceeds recall performance for corresponding unpracticed controls, *c*. *Rp-* items, however, are recalled at a minor rate when compared to respective control items. This latter finding has been coined *Retrieval-induced Forgetting* (RIF).

Empirical evidence

This finding has attracted considerable attention in the field of memory research, inspiring almost 200 articles over the last two decades (Murayama, Miyatsu, Buchli, & Storm, 2014). By now, RIF has been shown to generalize over a wide range of materials, e.g., category-exemplar pairs (M. C. Anderson et al., 1994; M. C. Anderson, Bjork, & Bjork 2000; M. C. Anderson & McCulloch, 1999; M. C. Anderson & Spellman, 1995; Bäuml, 2002; Bäuml

& Hartinger, 2002; R. E. Smith & Hunt, 2000), text passages (Little, Storm, & Bjork, 2011), pictures and videos (Ford, Keating, & Patel, 2004; Migueles & García-Bajos, 2007), and eyewitness testimonies (MacLeod, 2002; Shaw, Bjork, & Handal, 1995), to name a few. Practical applicability has been discussed to extend from educational settings to eyewitness testimony, and interactions with other cognitive processes like social cognition, creative cognition, or abnormal cognitive processes in psychological disorders have been identified (for a review, see Storm et al., 2015).

Moreover, RIF has been shown over a wide range of memory tests. Empirical evidence suggests that RIF is not restricted to category-cued recall tests (e.g., M. C. Anderson et al., 1994, Experiment 1; Butler, Williams, Zacks, & Maki, 2001), but it arises as well in category-plus-stem-cued recall tests (e.g., M. C. Anderson et al., 1994, Experiment 2; M. C. Anderson, Bjork, et al., 2000), in item recognition tests (Aslan & Bäuml, 2011; Dobler & Bäuml, 2013; Gómez-Ariza, Lechuga, & Pelegrina, 2005; Hicks & Starns, 2004; Román, Soriano, Gómez-Ariza, & Bajo, 2009; Spitzer & Bäuml, 2007; Starns & Hicks, 2004; Veling & van Knippenberg, 2004; but see Koutstaal, Schacter, Galluccio, & Stofer, 1999), in tests of source memory (Hicks & Starns, 2004; Spitzer & Bäuml, 2009), and using the so-called *independent-probe technique* that employs a test cue different from the original study cue, e.g., cuing *hornet* with CREATURES LIVING IN A STATE-LIKE COMMUNITY (e.g., M. C. Anderson & Bell, 2001; M. C. Anderson & Spellman, 1995; Aslan, Bäuml, & Pastötter, 2007; S. K. Johnson & Anderson, 2004; Saunders & MacLeod, 2006; Shivde & Anderson, 2001; but see Camp, Pecher, & Schmidt, 2007; Jonker, Seli, & MacLeod, 2012; Perfect et al., 2004; Williams & Zacks, 2001). Single studies have also used implicit memory tests, e.g., perceptual identification, lexical decision, word-fragment completion, and recognition reaction time, yet producing mixed results (e.g., Bajo, Gómez-Ariza, Fernandez, & Marful, 2006; Butler et al., 2001; Perfect et al., 2004; Perfect, Moulin, Conway, & Perry, 2002; Veling & van Knippenberg, 2004; Verde & Perfect, 2011).

Limitations to RIF include item similarity, i.e., instructions to find similarities between *rp-* items and *rp+* items reduces RIF while finding similarities among *rp-* items boosts RIF (M. C. Anderson, Green, & McCullough, 2000), and integration of study material (M. C. Anderson & McCullough, 1999; see also Chan, 2009, 2010; Chan, McDermott, Roediger, 2006), which is furthermore supported by the finding that shorter study trials enhance RIF thus limiting room for strategic learning (Murayama et al., 2014; see also Goodmon & Anderson, 2011). Retention interval was discussed to be a boundary condition for RIF as studies that employed the test 12 hours or more after practice produced mixed results (Chan, 2009; MacLeod & Macrae, 2001; but see García-Bajos, Migueles, & Anderson, 2009; Storm, Bjork, & Bjork, 2012; Tandoh & Naka, 2007). The unreliability of RIF after longer delays can be attributed to methodological differences such as the use of highly integrated materials, i.e., text passages (see Murayama et al., 2014), repeated testing within-subject, both after a short delay and again after a longer delay (Migueles & García-Bajos, 2007; Storm, Bjork, Bjork, & Nestojko, 2006) versus single testing (Storm et al., 2012; for a direct comparison of repeated testing and single testing, see Saunders, Fernandes, & Kosnes, 2009, Experiment 3), and the activity within the retention interval, i.e., sleep versus wakefulness (Abel & Bäuml, 2012; Racsmany, Conway, & Demeter, 2010). Furthermore, several manipulations prior to or during practice have been found to eliminate RIF, such as inducing a negative mood (Bäuml & Kuhbandner, 2007) or stress between study and practice stages (Koessler, Engler, Riether, & Kissler, 2009) or to engage in a distracting task during retrieval practice (Román et al., 2009).

In the following passages, previous results relevant for the present thesis will be outlined, i.e., studies comparing the effects of retrieval and other practice methods. These other practice methods can be subsumed as noncompetitive retrieval and restudy variations. During all of these practice methods, participants are provided with the intact to-be-practiced items. More precisely, during noncompetitive retrieval practice, participants are provided with a subset of the studied exemplars the category's word stem as a retrieval cue which the subjects are asked to recall (e.g., in___ - *termite*). During restudy, a subset of the studied category-item pairs is re-presented (e.g., insect - *termite*) and participants are asked to study the word pairs once again.

In a first line of studies, it has been robustly found that forgetting of unpracticed items arises after standard (competitive) retrieval practice but not after noncompetitive retrieval practice (e.g., M. C. Anderson, Bjork et al., 2000; Ferreira, Marful, Staudigl, Bajo, & Hanslmayr, 2014; Hanslmayr, Staudigl, Aslan, & Bäuml, 2010; Saunders et al., 2009) and that forgetting of unpracticed items arises after standard (competitive) retrieval practice but not after restudy cycles (e.g., Bäuml & Aslan, 2004; Ciranni & Shimamura, 1999; Dobler & Bäuml, 2013; Hulbert, Shivde, & Anderson, 2012; Staudigl, Hanslmayr, & Bäuml, 2010). These results suggested that only retrieval but no other practice method can induce forgetting in the Retrieval-practice Paradigm.

A second line of studies investigated the robustness of these findings using modifications of noncompetitive retrieval and restudy. Raaijmakers and Jakab (2012) adapted M. C. Anderson, Bjork et al.'s (2000) original design to make noncompetitive retrieval practice more demanding: Participants studied category-exemplar pairs (e.g., round - *ball*) before they were asked to recall the category label when provided with the exemplar as a retrieval cue during practice (e.g., ___ - *ball*). Their procedure differed from the original one in several respects: For instance, the word stems of the category labels were not presented as retrieval cues; subjects received feedback after each noncompetitive retrieval trial; categories were chosen by shared physical properties, not by semantics; and items of low frequency within their categories were drawn. Results revealed RIF-like recall impairment of the unpracticed items after noncompetitive retrieval practice (for related findings, see Jonker & MacLeod, 2012). Verde (2013) modified the standard restudy condition, augmenting restudy with elaboration tasks: First, participants studied category-exemplar pairs; at practice, participants were instructed to either judge each single re-presented category-exemplar pair whether the category that was provided for the item was the best to classify it (best category judgment) or to rate the pleasantness of the re-presented item in the presence of the item's category cue (pleasantness rating). Finally, a cued recall test of all studied items was administered. Both restudy formats induced forgetting of the unpracticed items at test, indicating that retrieval may not be necessary for RIF to arise while particular restudy formats can be sufficient to induce RIF-like forgetting. In a further modification of the restudy method, Saunders et al. (2009) administered a mental imagery task during practice, i.e., a subset of previously studied category-exemplar pairs was presented and participants were asked to engage in mental visualization of particular exemplar features such as shape, color, or size (e.g., "Please form a mental image of the color of this item: INSECT - *termite*"). Mental imagery led to RIF-like forgetting in a cued recall test, mirroring the effects of standard (competitive) retrieval practice. Last but not least, Jonker, Seli, and MacLeod (2013) augmented restudy by inserting a context change task prior to practice. Participants first studied a categorized item list, then engaged in a mind-wandering task, a well-established method to induce an internal context change (mentally walking through the parents' house), then restudied a subset of these items, and finally were tested on all items using a cued recall test. While restudy without prior context change did not induce RIF-like forgetting replicating earlier findings, RIF-like forgetting arose when a context change task was interpolated. The results of these studies suggest that not only competitive retrieval practice but also other practice formats can induce forgetting in the Retrieval-practice Paradigm.

Critically, all these findings employed a cued recall test to assess memory performance. More recently, Grundgeiger (2014) investigated how competitive and noncompetitive retrieval practice affected recall and item recognition memory employing the adapted noncompetitive retrieval practice task by Raaijmakers and Jakab (2012). Participants studied a list of category-exemplar pairs, retrieved either the exemplar (competitive retrieval practice) or the category label (noncompetitive retrieval practice), and finally were tested on all studied items using either a cued recall test or an item recognition test. Replicating Raaijmakers and Jakab's (2012) results, both types of retrieval practice induced forgetting when the cued recall test was employed. Crucially, however, only competitive retrieval practice but not noncompetitive retrieval practice impaired recognition performance for unpracticed items. This finding indicates that the generality of the above mentioned findings may be moderated by the test format.

1.3 Underlying mechanisms

As output order is usually controlled in the Retrieval-practice Paradigm, RIF cannot simply be reduced to Output Interference during the final test. Different explanations have been discussed and tested, of which three major accounts have sustained investigations and received attention over the past two decades: strength-dependent blocking, context change, and inhibition. These accounts will be outlined in the following passages. Subsequently, the consistency of each account with the previously reported findings will be discussed.

Blocking account

As introduced above, strength-based blocking represents a well-established process and has been successfully accounted for many phenomena in memory research. According to blocking accounts, forgetting arises due to response competition between items that share a cue (see J. R. Anderson, 1983; McGeoch, 1942; Raaijmakers & Shiffrin, 1981; Rundus, 1973). The probability of recalling a target item is supposed to depend on the strength of its association with the cue relative to all other associations. When the association of a piece of information to its cue is strengthened, e.g., by re-exposure, future response competition between the information related to the cue will biased toward the strong information.

Applying this account to the Retrieval-practice Paradigm, it is assumed that cue-item associations of *rp+* items (e.g., *termite*) are strengthened through practice. In the final test, information that shares the same cue competes for retrieval (e.g., *hornet* and *termite*). The strengthened items (e.g., *termite*) intrude more persistently during attempts to retrieve other items related to the cue, blocking these *rp-* items (e.g., *hornet*) and thus lowering their likelihood of being retrieved. Links of control items (e.g., *lamp*) to their cue, on the other hand, are equally strong and thus no blocking occurs within these categories. So, the chance of retrieving *rp+* items is enhanced when compared to *c* items, but the chance of retrieving *rp-* items is reduced, when compared to *c* items. Thus, such a strength-based blocking account is compatible with the standard finding of RIF. Importantly, these strength-based blocking accounts do not require additional inhibitory mechanisms, but are solely based on strength modulations and interference.

Consequently, proponents of the blocking account of RIF predict that impaired memory for unpracticed items is not restricted to retrieval practice but can arise after any kind of adequate strengthening of the cue-item associations of the practiced items, i.e., strengthening of *rp+* items is necessary and sufficient for RIF to arise whereas retrieval practice is not a necessary means. Accordingly, equivalent effects of retrieval practice and re-exposure formats that promote adequate cue-item strengthening are anticipated inducing RIF-like forgetting of related unpracticed items.

Following this line of argumentation, RIF should also be equivalent to the List-strength Effect (Tulving & Hastie, 1972; Ratcliff et al., 1990) introduced above. Like retrieval practice is supposed to, a subset of the studied items in the mixed list receives particular strengthening, in the List-strength Paradigm either by longer study trials or by repeated presentation. These disproportionately strengthened items then block recall of the remaining items in the list, impairing retention of those weaker items. Therefore, the blocking account would predict RIF to arise whenever the List-strength Effect arises.

Context change account

As outlined above, it is widely accepted that information is encoded together with contextual information. Contextual information refers to all pieces of information that are not in the focus of study but are encoded without deliberation, such as the physical environment or internal states of mind. This context information can be used as cues to guide retrieval, such that a match of the context information provided during study and provided at test benefits retention (see Encoding Specificity Principle, Tulving & Thomson, 1973). External and internal conditions naturally vary over time but these context variations can also be induced, e.g., by changing the room, by mind-wandering - or by retrieval. The idea that retrieval triggers context change is derived from several studies showing that episodic and semantic retrieval produce similar effects like well-established context change inductions (e.g., Divis & Benjamin, 2014; Jang & Huber, 2008; Pastötter, Schicker, Niederhuber, & Bäuml, 2011; Sahakyan & Hendricks, 2012; Szpunar, McDermott, & Roediger, 2008). Szpunar and colleagues (2008) found that testing a recently studied list facilitates recall for a following list and reduces prior-list intrusions. This finding substantiates the notion, that retrieval may trigger a context change and, thus, reduces Proactive Interference by generating isolated lists (see also Pastötter et al., 2011). Moreover, testing of irrelevant items reduces recall of a list studied prior to testing, indicating that the act of retrieval may accelerate context fluctuation (Jang & Huber, 2008; see also Sahakyan & Hendricks, 2012). Further evidence has been gathered using semantic memory retrieval. When a semantic generation task is inserted between study lists, not only memory performance for lists that follow semantic generation improves, in addition, memory performance for a list that precedes semantic generation declines (Divis & Benjamin, 2014). This finding conforms to effects of context changes on recall (e.g., Pastötter, Bäuml, & Hanslmayr, 2008).

In line with the general idea of context effects, the context change account of RIF assumes that context information is routinely encoded along with the to-be-studied material. Critically, it posits that when participants switch from encoding during the study phase to active search of adequate responses during retrieval practice, context fluctuation is accelerated and therefore *rp+* items (*termite*) and their category cues (INSECT) are selectively associated with a new, distinct context. With items from unpracticed categories (FURNITURE - *lamp*) only the study context is associated and therefore the study context is reinstated in the final test, promoting recall of *c* items. For items from partially practiced categories (INSECT), the cue is linked to both study and practice contexts, but, by default, the practice context is reinstated due to its recency and/or elaboration of the practiced items in this context. Thus, for *rp+* items (*termite*), reinstatement of the practice context is beneficial whereas for *rp-* items (*hornet*) it is detrimental since *rp-* items are associated exclusively with the study context but not with the practice context. According to the context change account, two tenets are fundamental for RIF to arise: (1) a context change must be induced prior to practice, and (2) the practice context must be reinstated for recall of practiced categories.

Proponents of the context change account of RIF predict that impaired memory for unpracticed items is not restricted to retrieval practice but can arise after any kind of practice when it is preceded by context change, i.e., a context change prior to restudy is sufficient to induce RIF-like forgetting whereas retrieval practice is not necessary. Accordingly, retrieval practice and restudy preceded by context change should affect memory performance equivalently.

Inhibition account

Even though accounts of inhibition differ in their specifics, they share the assumption that inhibitory mechanisms are recruited to overcome response competition by directly impairing retrieval of allegedly irrelevant information.

This impairment can be obtained by inhibiting the access to the information or by inhibiting the trace itself. As a consequence, retrieval of allegedly relevant information is facilitated.

Applying the concept of inhibition to the Retrieval-practice Paradigm, the inhibition account of RIF assumes that, during practice, interference between members of a category arises, e.g., when INSECT - *te*___ is provided, *hornet* as a fellow member of the category INSECT may intrude, hindering retrieval of *termite* (e.g., M. C. Anderson, 2003). To enable successful retrieval of the to-be-practiced items, interference has to be resolved. It is assumed that executive control processes are recruited to override possible predominant responses by inhibiting them, thus reducing the activation of the item representation itself. This inhibition is supposed to be long-lasting, such that in the final test, *rp-* items (e.g., *hornet*) are recalled at a minor rate compared to control items (e.g., *lamp*) that were not object to inhibition.

According to the inhibition account, retrieval is critical to induce the effect as only if practiced and unpracticed items compete for retrieval, inhibitory mechanisms will be activated to suppress the unpracticed items. Therefore, proponents of the inhibition account will not predict equivalent effects of retrieval practice and any other kind of practice that does not recruit inhibitory mechanisms, in other words, retrieval is necessary for RIF to arise whereas other forms of practice are not sufficient.

Consistency of the accounts of RIF with empirical evidence

Blocking account. The blocking account assumes that RIF occurs because the strengthened cue-item associations of the practiced items block retention of the unpracticed items at test. Thus, any kind of practice that enhances cue-item associations to an adequate degree should affect memory performance equivalently to how retrieval practice does. At first glance, this prediction conflicts with the first line of studies showing that plain restudy did not produce significant forgetting like competitive retrieval practice did (e.g., Bäuml & Aslan, 2004; Ciranni & Shimamura, 1999; Dobler & Bäuml, 2013; Hulbert et al., 2012; Staudigl et al., 2010). Similarly, the absence of RIF following noncompetitive retrieval practice challenges the equivalence assumption of the blocking account (M. C. Anderson, Bjork et al., 2000; Ferreira et al., 2014; Hanslmayr et al., 2010; Saunders et al., 2009).

However, proponents of the blocking account argued that this finding may not be incompatible with the equivalence assumption. According to the blocking account, the amount of strengthening determines the amount of forgetting. Thus, the particular format of re-presentation may be critical for whether restudy induces forgetting or not (e.g., Raaijmakers & Jakab, 2012; Verde, 2013). Thus, plain restudy may not impair memory for the unpracticed items, because strengthening may not be as effective as when retrieval practice is employed (for related results, see e.g., Carrier & Pashler, 1992; Roediger & Karpicke, 2008) or strengthening may occur on the level of the item representations, not on the level of the cue-item associations which may not be sufficient to cause blocking at test (Verde, 2013). Yet, if retrieval practice was compared to restudy formats that enhance the cue-item associations of the practiced items to an adequate degree like retrieval allegedly does, equivalent effects of retrieval practice and restudy may arise.

In fact, the findings by Raaijmakers and Jakab (2012), Verde (2013), and Saunders et al. (2009) support this proposal and corroborate the blocking account. In order to compel participants to process the category-item associations and to increase their strength sufficiently, Raaijmakers and Jakab (2012) designed their noncompetitive retrieval practice task to be more demanding while Verde (2013) augmented restudy with additional tasks (e.g., best category judgments, pleasantness ratings). Furthermore, restudy augmented with mental imagery might be regarded as a means to enhance the cue-item associations of the practiced items more than standard restudy does, similar to how Verde's re-exposure formats are supposed to do, and it may thus cause blocking and RIF-like forgetting of unpracticed items. The results of the three studies consistently revealed recall impairment of the unpracticed items after noncompetitive retrieval practice, restudy augmented with category judgments or pleasantness ratings, and restudy augmented with mental imagery, suggesting that the strengthening of the category-exemplar associations can be sufficient to induce RIF-like forgetting. Whether or not the blocking account is fit to explain the finding by Jonker et al. (2013), showing recall impairment of unpracticed items following restudy preceded by a context change, remains unclear. This question will be returned to in the General Discussion of the present thesis.

Grundgeiger (2014) however provided first evidence for a moderating effect of test format on the equivalence of competitive retrieval and other practice formats. While the equivalence of the effects of competitive and noncompetitive retrieval practice could be sustained when recall was tested, the effects differed in item recognition tests showing RIF after competitive but not after noncompetitive retrieval practice. This finding challenges the idea of a general equivalence of RIF and strength-based forgetting. If blocking underlies the effects of both, competitive and noncompetitive retrieval practice, then either both or none of the practice methods should impair recognition memory. The findings rather indicate that (a) while RIF following competitive retrieval and RIF following noncompetitive retrieval may be equivalent in cued recall, they may not be equivalent in item recognition tests, and thus (b) RIF following competitive retrieval practice and RIF following noncompetitive retrieval practice are not mediated by the same mechanism, (c) findings from cued recall tests cannot be generalized to item recognition tests without reservation, and (d) recognition of unpracticed items was not impaired by noncompetitive retrieval practice and item recognition tests may therefore not be susceptible to blocking effects.

Free recall and category recall tests provide no item-specific cues, i.e., when participants are asked to recall all items from a studied list or from a studied category from the study episode, it is very likely that interference may have a role in forgetting. As more and more cues are provided specifying the to-be-retrieved item, the role of interference should decrease. Thus, in category-plus-stem-cued recall tests, the influence of interference should be considerably smaller and, in item recognition test that provide the actual item as a cue, close to zero. Because of this, one may assume that the finding of RIF in item recognition tests is incompatible with the blocking account. Blocking proponents argue however that interference effects from other items may be present in item recognition, as studies on Retroactive Interference or the List-length Effect (e.g., Chandler, 1989; Gronlund & Elam, 1994; Ratcliff et al., 1990; but see Dennis & Humphreys, 2001, or Kinnell & Dennis, 2011) and global-matching models of recognition memory (e.g., Gillund & Shiffrin, 1984; Murdock, 1982) indicate. Yet, at least some interference effects have been shown to be restricted to recall. For instance, it has been shown that the List-strength Effect is present when recall memory is assessed, but it is absent in item recognition tests (e.g., Murnane & Shiffrin, 1991; Ratcliff et al., 1990; Shiffrin, Ratcliff, & Clark, 1990). As outlined above, RIF and the List-strength Effect should have equivalent effects if RIF is induced by blocking and hence RIF should also be restricted to recall.

Context change account. The context change account assumes that RIF arises because retrieval induces a context change and the contextual mismatch between study and test impairs retention of the unpracticed items. Thus, any kind of practice if it is preceded by a context change should induce effects equivalent to retrieval practice. Jonker et al. (2013) provided direct evidence in favor of the equivalence assumption by the context change account demonstrating RIF-like forgetting in a cued recall test following a context change prior to restudy. This finding, once more, indicates that retrieval is not a necessary condition for RIF to occur but that restudy preceded by a context change is sufficient.

The context change account is furthermore in line with the finding that plain restudy does not induce RIF (e.g., Bäuml & Aslan, 2004; Ciranni & Shimamura, 1999; Dobler & Bäuml, 2013; Hulbert et al., 2012; Staudigl et al., 2010) as restudy is not supposed to accelerate context fluctuation (Jonker et al., 2013). The consistency of the context change account with the restudy variations employed by Saunders et al. (2009) and Verde (2013) cannot be determined unequivocally as it is rather a matter of speculation, which will be seized in the General Discussion of the present thesis. Nevertheless, the results by Raaijmakers and Jakab (2012) may be interpreted as a support of the context change account. Based on previous findings (e.g., Divis & Benjamin, 2014), the context change account assumes that any kind of retrieval, whether retrieval is competitive or not, will accelerate context drift and will therefore induce memory impairment. Thus, the finding that both, competitive and noncompetitive retrieval practice induce RIF is in line with predictions by the context change account. However, as reported above, noncompetitive retrieval practice failed to induce RIF in multiple studies (M. C. Anderson, Bjork et al., 2000; Ferreira et al., 2014; Hanslmayr et al., 2010; Saunders et al., 2009). These failures to find RIF following noncompetitive retrieval practice were attributed to the lack of sufficient demand characteristics of the retrieval task (Raaijmakers & Jakab, 2012). Whether or not one might assume that the difficulty of the retrieval process affects the amount of context change that is induced, these mixed results are compatible or incompatible with the context change account of RIF.

As indicated by Grundgeiger's (2014) results, equivalence of competitive retrieval and other practice methods may be modulated by the test format. While the equivalence proposal is consistent with the finding that competitive and noncompetitive retrieval practice both produce impairment of the unpracticed items when a recall test is used, the diverging effects of competitive and noncompetitive retrieval practice on recognition memory are at odds with the context change account. Again, if both retrieval practice tasks engage a context shift then either both or none of the tasks should affect recognition memory depending on whether context change effects show in recognition tests.

Context reinstatement relies on an association between the test cue and an earlier context. According to the context change account, the category label, that is routinely used to cue the target, triggers reactivation of a particular context, i.e., the practice context for *rp* categories and the study context for c categories. Thus, when a novel cue is used to test the item, it cannot reactivate a context as it is not associated to the study or practice context, for this reason, RIF should be eliminated. One might consider that participants who are usually instructed to complete the novel cue with an adequate item from the study list might attempt to reactivate the study context. Undifferentiated reinstatement of the study context, irrespective of the category, should not lead to forgetting, according to the context change account. Whether context-dependent forgetting arises in recognition tests remains an unsettled matter. Despite numerous studies on context effects in recognition, no clear-cut conclusion has been reached whether or under which circumstances recognition memory is affected by context change. While some studies reported effects of context manipulations on recognition memory (Bodner & Lindsay, 2003; Bodner & Richardson-Champion, 2007; Craik & Schloerscheidt, 2011; Light & Carter-Sobell, 1970; Tousignant & Bodner, 2012; Tulving & Thomson, 1973; Watkins, Ho, & Tulving, 1976; Winograd & Rivers-Bulkeley, 1977), others did not (Basden, Basden, & Gargano, 1993; Fernandez & Glenberg, 1985; Geiselman et al., 1983; Godden & Baddeley, 1980; Sego, Golding, & Gottlob, 2006; S. M. Smith, 1988; S. M. Smith et al., 1978) suggesting that context effects depend on the specific settings of the test and material that have not been clearly determined yet. Context effects have been found to arise selectively in the recollection component of recognition memory (Macken, 2002). The context change account states that RIF might manifest in recognition tests only if the use of context cues is encouraged (Jonker et al., 2013). In line with this assumption, Verde (2004) and Verde and Perfect (2011) found that RIF arises exclusively in "remember" judgments (recollection) and self-paced recognition tests but not in "know" judgments (familiarity) and speeded recognition tests suggesting that only the recollection component of recognition memory is affected by retrieval practice indicating a crucial role for contextual cues in identifying old items (but see Spitzer & Bäuml, 2007). Due to these mixed results, it remains unclear whether context change effects are to be expected in item recognition tests when employing the Retrieval-practice Paradigm.

Inhibition account. The inhibition account assumes that RIF arises because, during retrieval practice, the to-be-practiced items and the unpracticed items compete for retrieval and control processes are recruited to resolve competition by inhibiting the unpracticed items. As only a competitive retrieval task will require the involvement of inhibitory mechanisms, only competitive retrieval practice but no other kind of practice method may induce forgetting in the Retrieval-practice Paradigm. This assumption particular to the inhibition account has been termed *retrieval specificity*.

The first line of studies showing that plain restudy and easy noncompetitive retrieval practice tasks do not induce forgetting corroborates the retrieval specificity property of the inhibition account of RIF indicating that retrieval practice is necessary for RIF to arise. On the contrary, empirical evidence that noncompetitive retrieval practice, certain forms of restudy, or restudy preceded by a context change, i.e., practice methods designed to avoid interference and potential inhibition of unpracticed items, can induce forgetting (Jonker et al., 2013; Raaijmakers & Jakab, 2012; Saunders et al., 2009; Verde, 2013) question the retrieval specificity assumption and thus the validity of the inhibition account. It should be noted, however, that Saunders et al. (2009) originally interpreted the finding, that restudy augmented with imagination impairs recall of unpracticed items like retrieval practice does, as evidence in favor of inhibition assuming that, during visualization, interfering item features have to be inhibited. Therefore, the findings by Saunders et al. may also promote an inhibition-based explanation.

Grundgeiger's (2014) results showing equivalent effects of competitive and noncompetitive retrieval practice in recall, but disparate effects in recognition memory indicate that while retrieval specificity of RIF may not hold in cued recall tests, it may hold when item recognition tests are employed. In fact, Grundgeiger suggested that recognition tests may be a reliable way of distinguishing between forgetting due to inhibition and forgetting due to strength-based interference in the Retrieval-practice Paradigm (see also Aslan & Bäuml, 2010). The idea was that recall tests may be susceptible to both inhibition and blocking effects, and thus a variety of practice methods may induce recall impairment of unpracticed items, whereas recognition tests may be selectively affected by inhibition, and thus only practice methods that engage inhibitory processes, i.e., competitive retrieval practice, may lead to impaired recognition performance. According to the inhibition account of RIF, RIF should arise irrespective of the cue used at test, i.e., RIF is *cue-independent*. Since inhibition is supposed to affect the representations of the items themselves, not the associations to the cue, the detrimental effect of practice should therefore generalize to other, novel cues and be observable over a wide range of memory tests. Proponents of the inhibition account suggested that an instance of such cue independence can be derived from results in recognition tests (M. C. Anderson & Levy, 2007; Jonker et al., 2013; Storm & Levy, 2012). Therefore, the inhibition account of RIF predicts that RIF is not restricted to recall, but should arise in recognition tests as well.

In summary, the previous evaluation of the empirical findings and their consistency with the accounts of RIF indicates that while RIF may be (partially) caused by blocking or context change mechanisms in recall tests, item recognition tests may be susceptible to inhibition effects only. Regardless of whether blocking or context change affect item-specific recognition tests or not, both accounts allow a straightforward prediction: If RIF is equivalent to strength-dependent forgetting, as stated by the blocking account, or if it is equivalent to context-dependent forgetting, as stated by the context change account, and if RIF arises in item recognition tests, as has been frequently supported, then strength-dependent forgetting and context-dependent forgetting, respectively, should be present in a modified Retrieval-practice Paradigm when memory performance is assessed via item recognition. Previous research submitted evidence that, when memory performance is assessed using recall tests, RIF seems to be equivalent to strength-dependent or context-dependent forgetting. However, the empirical evidence supporting the claim that the equivalence assumption of both the blocking and the context change accounts will not hold in item recognition tests is scarce. The present dissertation focused on the question whether the findings in recall can be generalized to item recognition tests and thus whether the equivalence proposal of the blocking or the context change account will also hold in item recognition testing.

Chapter 2

Goals of the present study

Retrieval-induced Forgetting refers to the finding that retrieving information leads to impaired retrieval of related, but unpracticed information. After two decades of research, the debate regarding the underlying processes of this phenomenon is still unsettled. The three most promising accounts attribute RIF to blocking (e.g., Raaijmakers & Jakab, 2012, 2013; Verde, 2013), to context change (Jonker et al., 2013), and to inhibition (e.g., M. C. Anderson et al., 1994; M. C. Anderson, 2003; Storm & Levy, 2012). All of these accounts are able to explain the standard finding of RIF in cued recall. Proponents of the blocking account assume that the likelihood of recalling the unpracticed items is reduced when competing with the practiced items for recall due to the strengthening of those items. The context change account considers a contextual mismatch between study and test responsible for RIF that is induced by retrieval during practice. According to the inhibition account, RIF arises because the unpracticed items are actively inhibited during practice in order to assist retrieval of the to-be-practiced items.

The main assumption of both the blocking and the context change accounts can be summarized in an equivalence hypothesis: According to the blocking account, RIF is equivalent to strength-dependent forgetting and can therefore be simulated using any other form of practice that strengthens cue-item associations. Thus, the effects of retrieval and practice formats that enhance cue-item associations should be equivalent under any circumstances. Analogously, the context change account assumes that RIF is equivalent to context-dependent forgetting and thus can be mimicked with any form of re-exposure that is preceded by a context change. Retrieval and re-exposure preceded by context change should hence return congruent results irrespective of other methodological variables. As discussed above, recent findings showed that restudy formats that enhance cue-item associations and restudy preceded by context change induce forgetting just like retrieval practice does when cued recall tests are employed supporting the blocking and the context change account, respectively (Jonker et al., 2013; Raaijmakers & Jakab, 2012; Saunders et al., 2009; Verde, 2013). Moreover, these observations disagree with a core assumption of the inhibition account, i.e., retrieval specificity, thus questioning its validity. The inhibition account considers retrieval necessary for RIF to occur since the competition of to-be-practiced and unpracticed items during retrieval practice requires inhibitory mechanisms to assist successful retrieval.

However, a recent study by Grundgeiger (2014) provided a first demonstration that the equivalence proposal of the blocking account may not hold unconditionally but may be restricted to cued recall tests. In his experiments, a restudy format used to enhance cue-item associations mimicked the effects of retrieval practice when a cued recall test was administered. On the other hand, only retrieval practice, but not restudy, induced forgetting when item recognition memory was tested. This finding shows that the effects of retrieval and restudy formats that enhance cue-item associations are not equivalent under all conditions and it suggests that RIF in item recognition tests may not result from blocking or contextual mismatch, but from a different mechanism. Yet, Grundgeiger's finding is but a single instance providing evidence against the equivalence hypotheses and does not substantially challenge the blocking and context change accounts without further replication.

In this study, the equivalence assumptions of both the blocking and the context change accounts of RIF were revisited, examining whether restudy formats that enhance cue-item associations or restudy preceded by a context change can mimic the effects of retrieval practice. Based on Grundgeiger's (2014) finding that the equivalence of the effects may depend on the format of the final test, test mode was varied using both, cued recall and item recognition. As reported above, practice formats that increase strength of the cue-item association as well as context change tasks prior to restudy have been shown to impair recall performance very similar to retrieval practice. The goals of the present study were (a) to replicate these previous findings in cued recall tests, and (b), going beyond prior work, to examine whether retrieval and restudy formats that enhance cue item-associations, and retrieval and restudy preceded by context change will return equivalent results when an item recognition test is employed.

Experiments 1-3 focus on the equivalence hypothesis of the blocking account. Accordingly, retrieval was compared to three restudy formats that are supposed to strengthen cue-item associations in both, a cued recall test and an item recognition test. If RIF is the result of blocking processes, then in both tests the same pattern of results should be observed after retrieval practice and after restudy formats enhancing cue-item associations. Thus, if the effects of practice formats that enhance cue-item associations mimic the effects of retrieval practice in a cued recall test, showing enhancement of practiced items and forgetting of unpracticed items, they should also mimic the effects of retrieval practice in item recognition. Divergent results in the item recognition test, showing forgetting following retrieval practice but no forgetting following the restudy formats, would challenge the blocking account.

Experiments 4-6 test the equivalence hypothesis of the context change account. It was investigated how retrieval practice and two different context manipulations prior to restudy affected cued recall and item recognition performances. If RIF is the result of a contextual mismatch, then retrieval and restudy preceded by context change should return the same pattern of results. Again, if the effects of restudy preceded by context manipulations mimic the effects of retrieval practice in cued recall, they should mimic the effects of retrieval practice in the item recognition test as well. A dissimilar pattern of results in item recognition, showing forgetting following retrieval practice but no forgetting following restudy preceded by context manipulations, would contest the validity of the context change account.

Chapter 3

Experiment 1-3: Testing the blocking account

The first set of experiments investigated whether the blocking account of RIF and the incorporated equivalence assumption can be sustained. The blocking account of RIF attributes the forgetting of the unpracticed items to blocking by the practiced items that are more strongly associated to the cue. Thus, RIF should be equivalent to the effects of any practice method that strengthens cue-item associations during practice to an adequate extent. This equivalence implies that whenever RIF arises, so should strength-dependent forgetting in the Retrieval-practice Paradigm.

Based on Grundgeiger's (2014) finding, that the equivalence assumption of the blocking account may depend on the format of the final test, test mode was varied using either cued recall or item recognition. The results of three experiments (Experiments 1-3) are reported that were designed to directly compare the effects of standard (competitive) retrieval practice to those of re-exposure formats that strengthen cue-item associations, i.e., Raaijmakers and Jakab's (2012) re-exposure with retrieval of the category label (noncompetitive retrieval practice; Experiments 1a & 1b), Verde's (2013) re-exposure with pleasantness ratings (Experiments 2a & 2b), and Saunders et al.'s (2009) re-exposure with mental imagery (Experiments 3a & 3b). The blocked design used by Dobler and Bäuml (2013) was adopted to allow for comparisons of conditions within subjects. In each experiment, participants studied a list of category-exemplar pairs. Subsequently, a subset of the studied pairs was practiced, employing (competitive) retrieval practice on half of the practiced items first and one of the three re-exposure formats on the other half of the practiced items second. Following a distractor task, all studied items were tested using either a cued recall test (Experiments 1a, 2a, & 3a), in which participants retrieved the exemplars from memory given the category label and the initial letter as cues, or an item recognition test (Experiments 1b, 2b, & 3b), in which participants discriminated studied items from lures, rating their confidence of an item having been previously studied (old) or not (new).

On the basis of strength-based accounts of RIF, which posit that both competitive retrieval and (some) re-exposure formats increase cue-item associations and such strengthening induces blocking at test (Raaijmakers & Jakab, 2012; Verde, 2013), competitive retrieval and the three re-exposure formats should have similar effects on memory performance of the unpracticed items. The results of experiments employing cued recall (Experiments 1a, 2a, & 3a) are reported with the goal of demonstrating that, for the present materials and experimental design, both (competitive) retrieval practice and each of the three re-exposure formats can induce RIF or rather RIF-like forgetting when using recall at test. Such a finding would replicate the recall results of previous studies by Raaijmakers and Jakab (2012), Grundgeiger (2014), Verde (2013), and Saunders et al. (2009). The blocking account would further predict that, depending on the presence or absence of blocking effects in item recognition tests, either all practice formats - competitive retrieval practice, noncompetitive retrieval practice, re-exposure supplemented with pleasantness ratings, and re-exposure supplemented with mental imagery - should impair recognition of unpracticed items, generalizing the outcomes of previous recall studies to item recognition, or all four practice formats should have no influence on recognition performance. Both patterns of results would be consistent with the equivalence assumption of the blocking account of RIF.

3.1 EXPERIMENT 1A: THE EFFECTS OF COMPETITIVE AND NONCOMPETITIVE RETRIEVAL PRACTICE ON CUED RECALL

Experiment 1a was dedicated to replicate earlier findings showing that RIF-like forgetting can arise following noncompetitive retrieval practice when cued recall is employed (Grundgeiger, 2014; Raaijmakers & Jakab, 2012). Here, the standard (competitive) retrieval practice task was contrasted with an adapted noncompetitive retrieval practice task, highly similar to the ones used by Raaijmakers and Jakab (2012) and Grundgeiger (2014). Raaijmakers and Jakab (2012, p. 25) posit that for noncompetitive retrieval practice to increase the strength of cue-item associations of practiced items and thus to induce blocking of unpracticed items, particular requirements should be met: for instance, cue-item associations should initially be rather weak, i.e., low- to medium-frequency category members should be selected; to-be-recalled category labels should not be cued with corresponding initial letters during practice; feedback should be provided to grant strengthening of associations of all practiced word pairs. Experiments 1a and 1b were designed along these requirements. On the basis of the two previous studies, the typical beneficial and detrimental effects of retrieval after both competitive and noncompetitive retrieval practice were expected when memory performance was assessed using a cued recall test.

Methods

Participants. Thirty-six participants were recruited at Regensburg University (M = 22.2 years, range = 18-31 years, 25 female). All subjects spoke German as native language and received monetary reward for their participation.

Materials. Nine semantic categories (CAR EQUIPMENT, PROFESSIONS, KITCHEN SUPPLIES, CLOTHING, MUSICAL INSTRUMENTS, FOOD, BODY PARTS, MEANS OF TRANSPORTATION, FOUR-LEGGED ANIMALS) with six items each were drawn from published word norms (Mannhaupt, 1983; Scheithe & Bäuml, 1995). Three additional categories (ALCOHOLIC DRINKS, SANITARY ARTICLES, GEMSTONES) with two items each were selected serving as buffer items in the study list. The German translations of the category labels of the nine experimental categories consisted of a single word. Within each category, the to-be-studied exemplars began with a unique initial letter. Following Raaijmakers and Jakab (2012) and Grundgeiger (2014), low- to medium-frequency practiced and unpracticed items (*median* = 35.5) were used as category exemplars. As Grundgeiger (2014) demonstrated RIF after noncompetitive retrieval practice with both perceptual and semantic categories, semantic categories were employed here as well to allow for better comparison of the results of this experiment to the results of Experiments 2a-3b.

Design. The experiment had a 2 × 3 design with the within-subject factors of PRACTICE TYPE (competitive retrieval, noncompetitive retrieval) and ITEM TYPE (practiced, unpracticed, control). The experiment consisted of three main phases: an initial study phase, a practice phase, and a final test phase.

In the practice phase, participants retrieved three exemplars of three categories (competitive retrieval condition) and retrieved the category labels of three exemplars of three other categories (noncompetitive retrieval condition). The remaining three categories represent control categories. Practice was blocked by practice type, i.e., half of the subjects completed the competitive retrieval condition first and then the noncompetitive retrieval condition, the other half vice versa (for a similar design, see Dobler & Bäuml, 2013). Thus, six types of items were created: practiced items, i.e., competitively retrieval practiced (*crp+*) items and noncompetitively retrieval practiced (*ncrp+*) items; unpracticed items of practiced categories, i.e., items that are members of the same category as the *crp+* or *ncrp+* items but are not retrieved or reexposed in the practice phase (*crp-*, *ncrp-*); and items from unpracticed categories that serve as controls for the practiced (*c+*) and unpracticed (*c-*) items. Categories were counterbalanced across participants, to be either competitively retrieval practiced, noncompetitively retrieval practiced, or not practiced at all (control). Thus, items designated as *crp+* or *ncrp+* items for a subset of the participants served as *c+* (control) items for another subset of the participants, and items designated as *crp-* or *ncrp-* items for some participants served as *c-* (control) items for other participants.

Procedure. The procedure is displayed in Figure 1. In the study phase, participants were instructed to learn category-exemplar pairs (e.g., FURNITURE

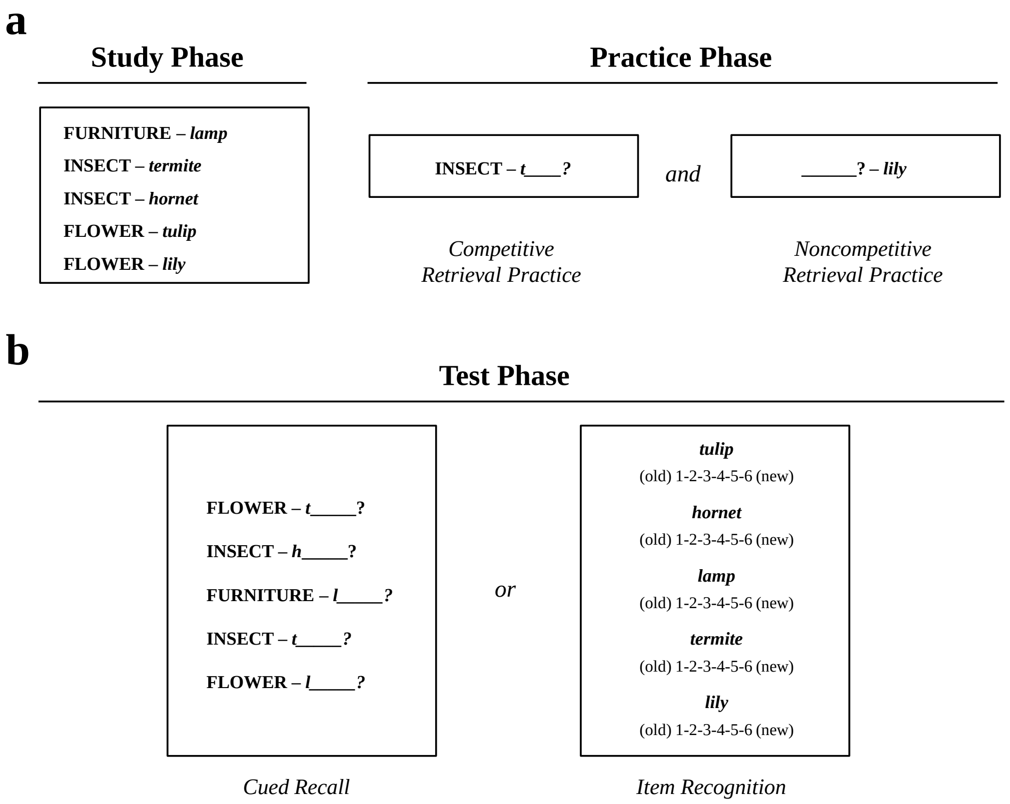

Figure 1. Procedure and conditions employed in Experiments 1a and 1b. (a) Study and practice phases of Experiments 1a and 1b: Participants studied a list of categorized items. In the practice phase, participants practiced some items by retrieving the exemplar (competitive retrieval practice) and some items by retrieving the category label (noncompetitive retrieval practice). (b) Test phase: In Experiment 1a, the category label and the exemplar's initial letter were provided and participants were asked to recall the exemplar (cued recall). In Experiment 1b, all studied exemplars and lures were provided and participants were asked to make old/new judgments (item recognition).

- *lamp*, INSECT - *termite*, INSECT - *hornet*, FLOWER - *tulip*, FLOWER - *lily*), each presented for 4 s (ISI = 500 ms) on a computer screen. The serial order of the study list was blocked randomized: six blocks were randomly arranged, with each block containing one exemplar from each category. To account for primacy and recency effects, three buffer items were presented at the beginning and the end of the study list. Half of the participants received the study list in the original order, the other half studied the items in reversed order. Next, participants were distracted for 60 s by counting backwards in steps of 3 from a three-digit number, again to control for recency effects.

In the practice phase, subjects practiced three members of six out of the nine categories; the members of three categories were practiced by competitive retrieval, the members of the other three categories by noncompetitive retrieval. In the competitive retrieval condition, category label and initial letter of a previously studied item were presented (e.g., INSECT - *t*___) for 5 s (ISI = 500 ms) and participants were asked to recall the corresponding item. Feedback was provided after each trial: the correct answer appeared on the screen for 2 s (ISI = 500 ms; e.g., INSECT - *termite*). Items were practiced twice in two consecutive cycles. The array of items within each practice cycle was blocked randomized, i.e., within blocks, items were drawn in random order, as was the order of the blocks. In the noncompetitive retrieval condition, a particular exemplar of the study list was displayed and participants tried to recall the corresponding category label assigned to the item during study (e.g., ____ - *lily*). Like in the competitive retrieval condition, the time frame for retrieval was 5 s (ISI = 500 ms). Feedback was provided after each trial presenting the correct response for 2 s (ISI = 500 ms; e.g., FLOWER - *lily*). Again, the order of items was blocked randomized and items were practiced twice in two consecutive cycles. After practice, participants completed a distractor task for 8 min (Raven's Progressive Matrices).

Finally, all to-be-studied items were tested employing a cued recall procedure. Items were cued with the category label and the initial letter (e.g., INSECT - *t*___) for 5 s (ISI = 500 ms) and the participants generated the corresponding exemplar from the study phase. Again, the test list was arranged using blocked randomization: six blocks compiled of nine items were constructed, each including one member of each category. Three blocks comprised exclusively of unpracticed items from practiced categories (*crp-*, *ncrp-*) and their control counterparts (*c-*), the remaining three blocks consisted of *crp+*, *ncrp+*, and their control counterparts (*c+*). To control confounding effects of Output Interference, blocks including *crp-*, *ncrp-*, and *c-* items were presented in the first half of the test, and the blocks including *crp+*, *ncrp+*, and *c+* items were presented in the second half. Within these boundaries, items within blocks and the blocks themselves were presented in random order. Three buffer items were tested at the beginning of the cued recall test.

Results

Practice Phase. Regarding the competitive retrieval condition, recall rates mounted up to 54.6% (SD = 0.15) of the exemplars on the first practice cycle and 75.3% ($SD = 0.09$) in total. Regarding the noncompetitive retrieval condition, recall rates mounted up to 94.4% (SD = 0.08) of the category labels on the first practice cycle and 97.2% ($SD = 0.04$) of the labels in total. The success rates in the noncompetitive condition are comparable to the ones reported by Raaijmakers and Jakab (2012) and Grundgeiger (2014).[2]

Recall Test. Figure 2 displays rates of correctly recalled practiced items, unpracticed items, and control items in the competitive and noncompetitive retrieval conditions. Regarding the beneficial effects of practice, recall rates were 86.7% (SD = 0.12) for the *crp+* items, 63.6% (SD = 0.18) for the *ncrp+* items, and 37.4% (SD = 0.18) for the *c+* items. An ANOVA with the within-subject factor of ITEM TYPE (*crp+, ncrp+, c+*) revealed a main effect of ITEM TYPE, $F(2, 70) = 113.298$, $MSE = .019$, $p < .001$, $\eta^2 = 0.764$, due to significant recall enhancement of both *crp+* and *ncrp+* items relative to the *c+* items, $t(35) = 15.707$, $p < .001$, $d = 3.264$, and $t(35) = 7.517$, $p < .001$, $d = 1.476$ respectively, and reliably higher recall levels for *crp+* compared to *ncrp+* items, $t(35) = 7.225$, $p < .001$, $d = 1.560$. These results indicate that both types of practice were successful, but that competitive retrieval practice enhanced performance to a greater degree than noncompetitive retrieval practice did. Intrusion rates for the three types of

[2] In Raaijmakers and Jakab's (2012) study, participants recalled 96.5% on the first cycle and 98.5% in total in their Experiment 1, and 97% on the first cycle and 99% in total in their Experiment 2. Grundgeiger (2014, Experiment 2a) found success rates of 90.6% on the first cycle and 96.6% in total in the noncompetitive retrieval condition.

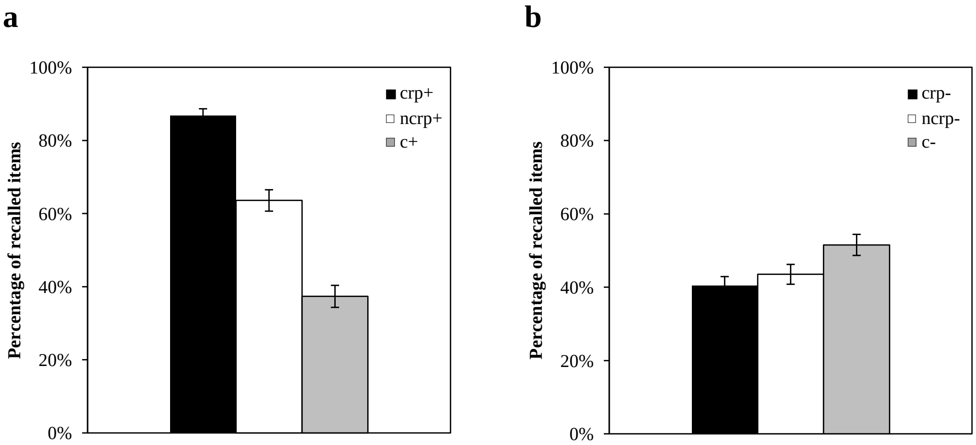

Figure 2. Results of Experiment 1a. Mean recall rates are shown as a function of item type. Error bars represent standard errors. (a) Recall percentages for competitively retrieval practiced items (crp+), noncompetitively retrieval practiced items (ncrp+), and control items (c+). (b) Recall percentages for unpracticed items of competitively retrieval practiced categories (crp-), unpracticed items of noncompetitively retrieval practiced categories (ncrp-), and control items (c-).

items were .01 ($SD = 0.03$) for the *crp+* items, .03 ($SD = 0.06$) for the *ncrp+* items, and .07 ($SD = 0.09$) for the *c+* items. Intrusion rates varied reliably with item type, $F(2, 70) = 8.980$, $MSE = .004$, $p < .001$, $\eta^2 = 0.204$. With respect to the intrusion rates, facilitation of practiced items may have been moderately underestimated.

Regarding the detrimental effects of practice, participants recalled 40.4% ($SD = 0.15$), 43.5% ($SD = 0.16$), and 51.5% ($SD = 0.17$) of *crp-* items, *ncrp-* items, and *c-* items, respectively. Using an ANOVA with the within-subject factor of ITEM TYPE (*crp-, ncrp-, c-*), significant differences between item types were found, $F(2, 70) = 6.059$, $MSE = .020$, $p = .004$, $\eta^2 = 0.149$. Planned comparisons revealed that both competitive and noncompetitive retrieval reduced recall performance of unpracticed items compared to controls, $t(35) = 3.550$, $p = .001$, $d = 0.692$, and $t(35) = 2.548$, $p = .015$, $d = 0.667$, while *crp-* and *ncrp-* items were recalled at a comparable rate, $t < 1$. Intrusion rates were .10 ($SD = 0.13$) for the *crp-* items, .10 ($SD = 0.12$) for the *ncrp-*

items, and .06 (SD = 0.08) for the *c-* items. These rates did not vary with item type, $F(2, 70) = 1.430$, $MSE = .013$, $p = .246$, $\eta^2 = 0.039$.

Discussion

Replicating prior work (Grundgeiger, 2014; Raaijmakers & Jakab, 2012), both competitive and noncompetitive retrieval practice were found to induce facilitation of the practiced items (*crp+, ncrp+*) and impairment of the unpracticed items (*crp-, ncrp-*). The finding is compatible with the equivalence assumption of the blocking account suggesting that RIF is the result of considerable strengthening of cue-item associations of the practiced items that block recall of unpracticed related items at test. Using identical material and the same procedures except for the final test, Experiment 1b was dedicated to examine whether the effects of noncompetitive retrieval practice observed in cued recall will generalize to recognition memory.

3.2 Experiment 1b: The effects of competitive and noncompetitive retrieval practice on item recognition

Experiment 1b investigated the effects of noncompetitive retrieval practice on item recognition. Like in Experiment 1a, effects of the noncompetitive retrieval practice task were contrasted with effects of the competitive retrieval practice task. In line with previous findings showing standard retrieval practice to induce forgetting when memory is assessed using recognition tests (e.g., Hicks & Starns, 2004), the expectation arose that, following competitive retrieval practice, recognition performance would be enhanced for practiced items, while recognition performance would be reduced for unpracticed items as compared to the baseline. Regarding the effects of noncompetitive retrieval practice, if recognition tests are susceptible to strength-dependent blocking effects, then the findings of Experiment 1a in cued recall should generalize to item recognition showing enhanced recognition for practiced items and impaired recognition for unpracticed items. If, however, item recognition tests are not or less prone to blocking effects, then noncompetitive retrieval practice may not induce forgetting in recognition as indicated by Grundgeiger's (2014) results.

Methods

Participants. Further 60 students of Regensburg University participated in the experiment (M = 23.5 years, range = 19-30 years, 42 female). Due to technical problems during data logging, one male participant had to be excluded in retrospect. The remaining 59 sets of data were used in the analyses. All participants spoke German as native language and received monetary reward in exchange for participation.

Materials. The material was identical to the one used in Experiment 1a. Substitution of cued recall with a recognition test demanded further category exemplars to serve as lures. For this, six further items were selected from each category with low- to medium-frequency (*median* = 35.5).

Design. Like Experiment 1a, Experiment 1b had the same 2 × 3 design with PRACTICE TYPE (competitive retrieval, noncompetitive retrieval) and ITEM TYPE (practiced, unpracticed, control) as within-subject factors.

Analogous to Experiment 1a, the practice phase divided the material into six item types: competitively practiced items (*crp+*) and unpracticed items from competitively practiced categories (*crp-*); noncompetitively practiced items (*ncrp+*) and unpracticed items from noncompetitively practiced categories (*ncrp-*); and the respective control items (*c+*, *c-*) that belong to utterly unpracticed categories.

For the final recognition test, three further item types were created: lures belonging to competitively retrieval practiced categories (*crp lures*); lures belonging to noncompetitively retrieval practiced categories (*ncrp lures*); and lures belonging to control categories (*c lures*). The items from both practice type conditions were tested within the same single recognition test together with the control items.

Procedure. The procedure was identical to Experiment 1a, except for the final test (see Figure 1). After the distractor task, participants attended to the final recognition test. All to-be-studied items and lures were displayed individually, with a schematic rating scale in the lower third of the screen. Participants were asked to rate their confidence whether an item was part of the studied list (old) or not (new) on a 6-point rating scale (1 = *definitely old*, 6 = *definitely new*). Participants responded via the keyboard, data were logged automatically. Subjects were encouraged to use the whole scope of the rating scale. No response time limit was imposed, i.e., not until the participant typed in an answer, the next item was presented (for a similar procedure, see Dobler & Bäuml, 2013, or Spitzer & Bäuml, 2007; for arguments in favor of this rating procedure compared to a procedure demanding binary old/new decisions, see Macmillan & Creelman, 2004; Parks & Yonelinas, 2008). The recognition list was arranged using blocked randomization respecting certain boundaries: studied items and lures were presented at most three times in succession; to circumvent potential Output Interference effects by practiced items, unpracticed items (*crp-*, *ncrp-*), corresponding control items (*c-*), and corresponding lures constituted the first half of the list; all other items were tested in the second half. Thus, twelve blocks were compiled with the first six blocks containing *crp-* items, *ncrp-* items, *c-* items, and lures, and the second six blocks containing *crp+* items, *ncrp+* items, *c+* items, and lures. In line with the restrictions, items from each category were selected randomly one by one and arranged in pseudo-random order for each block. Each list-half was mirror-inverted. Half of the participants were tested with the original order of the recognition list, the other participants were tested with the mirror-inverted version. Three buffer items preceded the recognition list.

Statistical Analysis. Starting at the most confident criterion, i.e., *definitely old* ("1"), the percentage of studied items correctly identified as "old" (i.e., hit rate) and the percentage of lures incorrectly identified as "old" (i.e., the false alarm rate) was accumulated across the rating scale. This approach results in an empirical Receiver Operating Characteristic (ROC) curve relating hit and false alarm rates across the particular response criteria (i.e., returning the tendency to respond with "old"; e.g., Macmillan & Creelman, 2004; Parks & Yonelinas, 2008). Thus, hit and false alarm rates under five different response criteria were derived from the present 6-point scale. The first point of the ROC ("1", i.e., *definitely old*) represents hit and false alarm rates when the strictest response criterion is applied, and each subsequent point ("2", "3", "4", "5") represents performance at an increasingly relaxed response criterion. Critically, as the function is cumulative, relaxing the scoring criterion increases both hit and false alarm rates monotonically.

First, the effect of the individual response criteria on the recognition data were examined. For this, corrected hits (hits - false alarms) were calculated as a function of item type and criterion. In an ANOVA, corrected hits were analyzed across the three most conservative ("old") response criteria ("1", "2", "3") with regard to whether participants scored higher on practiced items (*crp+*, *ncrp+*) compared to corresponding control items (*c+*), and whether participants scored lower on unpracticed items from practiced categories (*crp-*, *ncrp-*) compared to the respective controls (*c-*). If significant, post-hoc tests were conducted to test differences between practice conditions. The analyses of corrected hits (hits - false alarms) are based on linearity of the ROC function (e.g., Wixted, 2007b) although ROC functions are typically curvilinear and asymmetric along the diagonal. Thus, analysis of corrected hits can only represent an approximation towards analysis of subjects' recognition performance.

Considering the curvilinear and asymmetric shape of the ROC, a signal detection approach was employed to analyze the recognition data in a second step. Usually, it is assumed that the variance of the strength distribution is higher for the studied items relative to the lures (e.g., Dunn, 2004; Wixted, 2007a). Regarding this assumption, the unequal-variance signal detection model was applied, according to which participants rest their recognition judgments on a single source of information, i.e., the items' general memory strength. However, this does not necessarily suggest a single underlying memory process as, for instance, familiarity and recollection codes might be combined in an additive or nonadditive way (e.g., Kelley & Wixted, 2001; Wixted & Stretch, 2004). Participants rate an item with the particular level of confidence when the assessment of the item's memory strength is higher compared to response criterion c_i associated to that level of confidence. Memory strength of studied items is reflected in the distance between the means of the underlying strength distributions of old and new items (d_a). Application of the model to the present 5-point ROC data results in seven free parameters (memory strength of old items d_a, variance of the distribution of old items σ, and five criterion points $c_1 - c_5$) and thus three degrees of freedom for testing the model's goodness of fit. Maximum-likelihood techniques were used to estimate values for the model parameters and to statistically test parameter differences.

First, it was examined whether the unequal-variance signal detection model provided a good fit for item type and practice conditions. Next, differences in memory strength d_a across item types and practice conditions were tested, particularly with regard to potential beneficial and detrimental effects of practice, i.e., it was tested whether d_a for practiced items (*crp+*, *ncrp+*) was increased compared to corresponding control items (*c+*), and whether d_a for unpracticed items from practiced categories (*crp-*, *ncrp-*) was reduced compared to corresponding controls (*c-*). If analyses returned significant differences across item types, differences between practice conditions were analyzed applying the same likelihood-ratio methods. Finally, variance of the model's other parameters across item type was examined.

Results

Practice Phase. In the competitive retrieval condition, recall rates constituted 47.5% ($SD = 0.17$) on the first practice cycle and 69.5% ($SD = 0.11$) in total. In the noncompetitive retrieval condition, participants successfully recalled 94.2% ($SD = 0.08$) of the category labels on the first practice cycle and 97.0% ($SD = 0.04$) of the labels in total. These values resemble closely the ones reported in Experiment 1a.

Recognition Test: ANOVA of Corrected Hits. In Table 1, mean false alarm rates and corrected hit rates are displayed for the the single response criteria and item types. In ANOVAs, variance of corrected hits across item type was examined for the three most conservative ("old") response criteria. Regarding the beneficial effects of competitive retrieval practice, a 2 × 3 ANOVA with the within-participants factors of ITEM TYPE (*crp+*, *c+*) and RESPONSE CRITERION ("1", "2", "3") was conducted. A main effect of ITEM TYPE, $F(1, 58) = 95.457$, $MSE = 0.072$, $p < .001$, $\eta^2 = 0.622$, revealed higher corrected hits for the practiced than the control items. Even though this effect was qualified by an interaction with RESPONSE CRITERION, $F(2, 116) = 17.824$, $MSE = 0.006$, $p < .001$, $\eta^2 = 0.0235$, it was observable across all three response criteria, $ts(58) > 7.225$, $ps < .001$, $ds > 1.355$.[3] Regarding the beneficial effects of noncompetitive retrieval practice, a corresponding analysis comparing *ncrp+* and *c+* items showed a main effect of ITEM TYPE, $F(1, 58) = 89.110$, $MSE = 0.063$, $p < .001$, $\eta^2 = 0.606$, with higher corrected hits for the practiced than the control items. This main effect also varied with response criterion, $F(2, 116) = 7.726$, $MSE = 0.007$, $p = .001$, $\eta^2 = 0.118$, but was present across all three criteria, $ts(58) > 7.462$, $ps < .001$, $ds > 1.252$. Moreover, contrasting the practice methods, corrected hits for *crp+* items exceeded corrected hits for *ncrp+* items, $F(1, 58) = 4.665$, $MSE = 0.013$,

[3]This ANOVA, and all forthcoming related ANOVAs of Experiments 2b and 3b, and Experiments 4-6, returned also a significant main effect of RESPONSE CRITERION. Due to the cumulative nature of the ROC function (see Method above), the resulting main effect of RESPONSE CRITERION is rather trivial, and thus detailed reports are omitted.

Table 1 False alarm rates and corrected hit rates for Experiment 1b

Item type		Response criteria				
		"1"	"2"	"3"	"4"	"5"
crp+	False alarms	.033	.060	.114	.230	.479
	Corrected hits	.963	.938	.886	.770	.521
ncrp+	False alarms	.022	.057	.111	.238	.491
	Corrected hits	.920	.917	.872	.753	.509
c+	False alarms	.027	.058	.116	.246	.513
	Corrected hits	.623	.667	.662	.585	.416
crp-	False alarms	.033	.060	.114	.230	.479
	Corrected hits	.551	.614	.660	.595	.429
ncrp-	False alarms	.022	.057	.111	.238	.491
	Corrected hits	.615	.672	.675	.625	.439
c-	False alarms	.027	.058	.116	.246	.513
	Corrected hits	.627	.671	.673	.624	.416

Note. False alarm and corrected hit rates are shown as a function of item type and response criterion. crp+ = competitively retrieval practiced items; ncrp+ = noncompetitively retrieval practiced items; c+ = unpracticed items from unpracticed categories; crp- = unpracticed items from competitively retrieval practiced categories; ncrp- = unpracticed items from noncompetitively retrieval practiced categories; c- = unpracticed items from unpracticed categories. "1" reflects the strictest response criterion, i.e., definitely old, and each subsequent number ("2", "3", etc.) reflects a more and more relaxed criterion. Corrected hits = hits - false alarms.

p = .035, η^2 = 0.074, indicating that competitive retrieval practice results in greater enhancement than noncompetitive retrieval practice does. The main effect was qualified by a significant interaction, $F(2, 116) = 4.090$, MSE = 0.002, p = .019, η^2 = 0.066. The differences were only reliable with respect to the most confident response criterion ("1"), $t(58) = 3.204$, p = .002, d = 0.483, both other $ts(58) < 1.716$, ps > .091, ds < 0.258.

Regarding the detrimental effects of competitive retrieval practice, analogously, a 2 × 3 ANOVA with the factors of ITEM TYPE (*crp*-, *c*-) and RESPONSE CRITERION ("1", "2", "3") was conducted. The analysis showed no main effect of ITEM TYPE, $F(1, 58) = 2.666$, MSE = 0.079, p = .108, η^2 = 0.044, but a reliable interaction between the two factors, $F(2, 116)$ = 4.376, MSE = 0.007, p = .015, η^2 = 0.070. Post-hoc t-tests returned a significant difference between corrected hits for *crp*- and *c*- items for criterion "1", $t(58) = 2.242$, p = .029, d = 0.350, with lower values for *crp*- items, a marginally significant difference for criterion "2", $t(58) = 1.815$, p = .075, d = 0.275, and no significant difference for criterion "3", $t(58) < 1$. Regarding the detrimental effects of noncompetitive retrieval practice, the results of a 2 × 3 ANOVA contrasting *ncrp*- and *c*- items showed no main effect of ITEM TYPE, $F(1, 58) < 1$, and no interaction between the two factors, $F(1, 58) < 1$. These results indicate that competitive retrieval practice induced RIF, at least for the two most conservative response criteria, whereas noncompetitive retrieval practice did not induce any RIF-like forgetting. Consistent with this indication, a 2 × 3 ANOVA with the factors of ITEM TYPE (*crp*-, *ncrp*-) and RESPONSE CRITERION ("1", "2", "3") showed no main effect of ITEM TYPE, $F(1, 58) = 2.461$, MSE = 0.073, p = .122, η^2 = 0.041, but an interaction between the two factors, $F(2, 116) = 3.627$, MSE = 0.006, p = .030, η^2 = 0.059. Follow-up tests revealed significantly lower corrected hits for *crp*- than *ncrp*- items for criterion "1", $t(58) = 2.102$, p = .040, d = 0.325 a marginal significant difference for criterion "2", $t(58) = 1.810$, p = .076, d = 0.319, and no significant difference for criterion "3", $t(58) < 1$. These

Table 2 Unequal-variance signal detection model for Experiment 1b

	Parameter estimates		Goodness of fit		
Item type	d_a	σ	χ^2	*df*	*p*
crp+	11.70*	3.60	0.52	3	.915
ncrp+	4.42*	1.51	0.89	3	.829
c+	2.67	1.92	1.87	3	.600
crp-	2.29*	1.63	2.88	3	.411
ncrp-	2.64	1.77	0.07	3	.995
c-	2.67	1.81	0.36	3	.948

Note. crp+ = competitively retrieval practiced items; ncrp+ = noncompetitively retrieval practiced items; c+ = unpracticed items from unpracticed categories; crp- = unpracticed items from competitively retrieval practiced categories; ncrp- = unpracticed items from noncompetitively retrieval practiced categories; c- = unpracticed items from unpracticed categories. d_a = general memory strength; σ = variance of the target distribution.
* Significant deviations from control performance ($p < .05$).

results indicate that the RIF findings were retrieval specific.

Recognition Test: Analysis of Hit and False Alarm Rates Using the Unequal-Variance Signal Detection Model. In the second step, the unequal-variance signal detection model was employed to analyze the data, which takes the curvilinear and asymmetric form of the ROC into account. Figure 3a and Figure 3b depict the ROCs for the practiced items, the unpracticed items, and the respective control items in the competitive and noncompetitive retrieval conditions, as well as the fit of the unequal-variance signal detection model to the data of each single condition. Table 2 shows the statistics of goodness-of-fit and maximum-likelihood estimates of the model's parameters d_a and σ for practiced and unpracticed items and their control

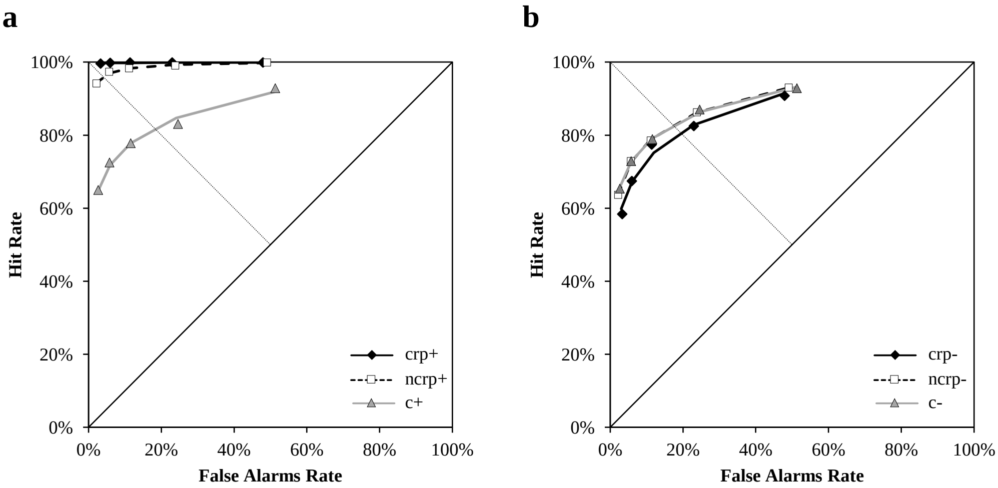

Figure 3. Item recognition Receiver Operating Characteristics (ROCs) depicting cumulative hit and false alarm rates as a function of item type. Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model. (a) ROCs for competitively retrieval practiced items (crp+), noncompetitively retrieval practiced items (ncrp+), and control items (c+). (b) ROCs for unpracticed items of competitively retrieval practiced categories (crp-), unpracticed items of noncompetitively retrieval practiced categories (ncrp-), and control items (c-).

counterparts.

The unequal-variance signal detection model provided a good fit to the recognition data of the six item types items (*crp+*, *ncrp+*, *c+*, *crp-*, *ncrp-*, *c-*), all $\chi^2 s(3) < 2.878$, $ps > .410$. Regarding the beneficial effects of practice, both competitively retrieval practiced (*crp+*) items and noncompetitively retrieval practiced (*ncrp+*) items showed enhanced memory strength as measured by d_a relative to the control (*c+*) items, $\chi^2(1) = 21.290$, $p < .001$, and $\chi^2(1) = 17.787$, $p < .001$, indicating that both types of practice were successful. The numerical difference in d_a between *crp+* and *ncrp+* items reached marginal significance, $\chi^2(1) = 3.751$, $p = .053$, with competitive retrieval practice inducing a higher memory strength than noncompetitive retrieval practice for the practiced items.[4]

[4]If recognition performance gets close to ceiling, as is the case for the practiced items in the competitive retrieval condition of Experiment 1b and the re-exposure condition of

Regarding the detrimental effects of practice, retrieval practice reduced d_a for the unpracticed (*crp-*) items in the competitive retrieval practice condition relative to the *c-* items, $\chi^2(1) = 4.386$, $p = .036$, but did not affect d_a for the unpracticed (*ncrp-*) items in the noncompetitive retrieval practice condition relative to the controls, $\chi^2(1) = 0.016$, $p = .899$. Consistently, d_a varied reliably between the two types of unpracticed items, $\chi^2(1) = 3.957$, $p = .047$, indicating that competitive retrieval practice, but not noncompetitive retrieval practice, induced forgetting of unpracticed items.

For both the practiced items and their controls, and the unpracticed items and their controls, the variance of the old items' distribution, as estimated by parameter σ, did not vary significantly across item type, $\chi^2 s(2) < 2.137$, $ps > .344$, but was larger than 1.0, $\chi^2 s(1) > 73.321$, $ps < .001$, indicating that the model's assumption of unequal variances for old and new items improved the description of the data significantly. The placement of the five confidence criteria varied across item type, for both sets of items, $\chi^2 s(10) > 6.402$, $ps <$ .049.

Discussion

Consistent with the results from previous item recognition studies (e.g., Hicks & Starns, 2004; Spitzer & Bäuml, 2007), competitive retrieval practice enhanced recognition of the practiced items (*crp+*) but reduced recognition of the unpracticed items (*crp-*). Moreover, noncompetitive retrieval practice enhanced recognition of re-exposed items (*ncrp+*). In contrast to the competitive retrieval practice condition, however, noncompetitive retrieval practice did not induce forgetting of the unpracticed items (*nrcp-*). The

Experiment 3b, parameter d_a typically gets overestimated when fitting the unequal-variance signal detection model to the data (e.g., Macmillan & Creelman, 2004; Macmillan, Rotello, & Miller, 2004). Such overestimation also occurred in the present experiments (see Tables 2 and 6). However, despite the resulting strong numerical difference between the two types of practiced items in Experiment 1b, the parameters for these items did not differ significantly between practice conditions. When fitting the model to the data in such cases, we followed prior work and substituted values of 100% performance by values of 99.9%.

findings arose from both, analysis of corrected hits and signal detection analysis, and demonstrated dissimilar effects of the two practice methods, with forgetting being present after competitive retrieval practice but not after noncompetitive retrieval practice.

The present finding replicates the results of a recent study by Grundgeiger (2014). Like Raaijmakers and Jakab (2012), Grundgeiger reported that noncompetitive retrieval practice can induce forgetting of other items when using recall testing, but he extended Raaijmakers and Jakab's results by showing that noncompetitive retrieval practice does not induce forgetting when using item recognition testing. Both the results of Grundgeiger (2014) and the present results arose from experiments that followed closely the requirements Raaijmakers and Jakab suggested for noncompetitive retrieval practice to enhance the cue-item associations of re-exposed items. Consistently, success rates during noncompetitive retrieval practice were comparable between the present experiment and the experiments by Grundgeiger and Raaijmakers and Jakab, indicating that recall of the category labels in the practice phase was similarly demanding across studies and thus should have increased the cue-item associations of the practiced items to a similar degree.

Importantly, because exactly the same materials and study and practice procedures were employed as in Experiment 1a, the results of Experiments 1a and 1b suggest that competitive retrieval practice reduces both recall and recognition of unpracticed items, whereas noncompetitive retrieval practice reduces recall but not recognition of these items. Thus, the results challenge the blocking account of RIF showing that RIF is not equivalent to strength-dependent forgetting when item recognition tests are employed and indicating that the strengthening of the cue-item associations of practiced items may not be sufficient to induce RIF-like forgetting.

Experiments 2a and 2b below examined whether the present findings generalize to another re-exposure format that, like noncompetitive retrieval practice, may strengthen practiced items' category-exemplar associations.

3.3 EXPERIMENT 2A: THE EFFECTS OF RETRIEVAL PRACTICE AND RESTUDY WITH PLEASANTNESS RATINGS ON CUED RECALL

Verde (2013) reported that not only noncompetitive retrieval practice but also re-exposure supplemented with best category judgments or pleasantness ratings can reduce recall of the not re-exposed items. The goal of Experiment 2a was to replicate this finding.

In the present Experiments 2a and 2b, re-exposure supplemented with pleasantness ratings was employed, because the impaired recall of the unpracticed items in Verde's study was numerically higher (14%) with this practice format relative to both the (competitive) retrieval practice condition (9%) and the category-judgment task (9%). Analogous to Experiment 1a, the effects of (competitive) retrieval practice were directly compared with the effects of this re-exposure condition, analyzing how the two types of practice affected later recall performance of practiced and unpracticed items. On the basis of Verde's previous study, the expectation arose that RIF would be observed after (competitive) retrieval practice and RIF-like forgetting after re-exposure supplemented with pleasantness ratings.

Methods

Participants. Thirty-six students of Regensburg University took part in the experiment (M = 22.4 years, range = 18-40 years, 26 female). All subjects spoke German as native language and received monetary reward for their participation.

Materials. The same material was employed as in the previous study by Dobler and Bäuml (2013). Nine categories (MUSICAL INSTRUMENTS, INSECTS, TREES, FRUITS, FURNITURE, SPICES, CLOTHING, TOOLS, FOUR-LEGGED ANIMALS) with six study items as well as three categories (GEMS, ALCOHOLIC

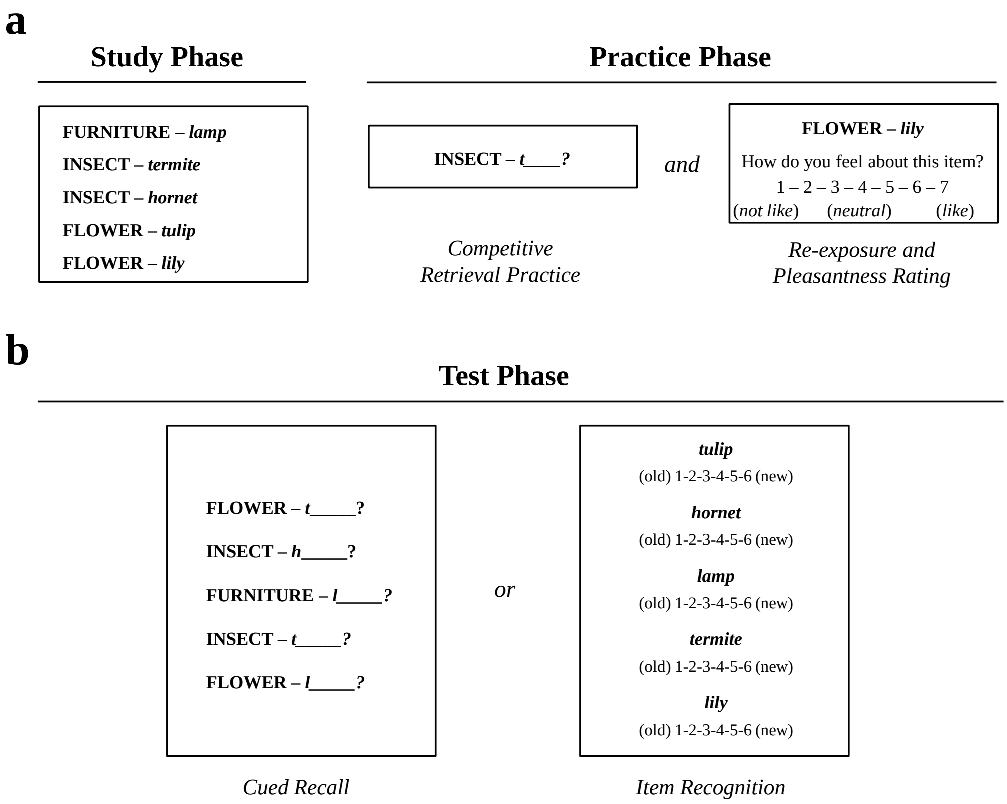

Figure 4. Procedure and conditions employed in Experiments 2a and 2b. (a) Study and practice phases of Experiments 2a and 2b: Participants studied a list of categorized items. In the practice phase, participants practiced some items by retrieving the exemplar (competitive retrieval practice) and some items by rating the item's pleasantness given the intact word pair (re-exposure and rating). (b) Test phase: In Experiment 2a, the category label and the exemplar's initial letter were provided and participants were asked to recall the exemplar (cued recall). In Experiment 2b, all studied exemplars and lures were provided and participants were asked to make old/new judgments (item recognition).

DRINKS, SANITARY ARTICLES) with two items each serving as buffer items were drawn from German published word norms (Mannhaupt, 1983) to create the study list. The German translations of the category names of the nine experimental categories consisted of a single word. The two most frequent exemplars of each category were excluded. Practiced and unpracticed items were of medium to high frequency (*median* = 10.5). Again, studied items within each category had a unique first letter.

Design and Procedure. The experiment had a 2 × 3 design with the within-subject factors of PRACTICE TYPE (competitive retrieval, re-exposure) and ITEM TYPE (practiced, unpracticed, control). The procedure was largely identical to Experiment 1a and differed only in the practice phase (see Figure 4). In the (competitive) retrieval condition, the category label and the first letter of the to-be-retrieved item were provided and participants were asked to retrieve the corresponding item (e.g., INSECT - *t*___); following Verde (2013), no feedback was provided. In the re-exposure condition, some of the original category-exemplar pairs were re-exposed and subjects were asked to judge the pleasantness of the presented exemplars on a 7-point scale (1 = *not pleasant at all*, 7 = *very pleasant*). Like in Experiments 1a and 1b, in each practice condition, items were practiced twice, in two successive practice cycles. For each practice type, three types of items were generated: practiced items, i.e., retrieval practiced (*crp+*) and re-exposed and rated (*re+*) items; unpracticed items of practiced categories (*crp-*, *re-*); and control items of unpracticed categories (*c+, c-*). Categories were counterbalanced between subjects to be either retrieval practiced, re-exposed and rated, or not practiced at all. The study list was arranged by blocked randomization as described in Experiment 1a.

Results

Practice Phase. In the competitive retrieval condition, participants correctly retrieved 67.0% ($SD = 0.20$) of the items on the first practice cycle and 68.7% ($SD = 0.20$) of the items in total.

Recall Test. Percentages of correctly recalled practiced and unpracticed items in the retrieval practice and re-exposure conditions and of the corresponding control items are depicted in Figures 5a and 5b. Regarding the beneficial effects of practice, recall rates mounted up to 68.2% ($SD = 0.21$) for the *crp+* items, 70.4% ($SD = 0.18$) for the *re+* items, and 49.7%

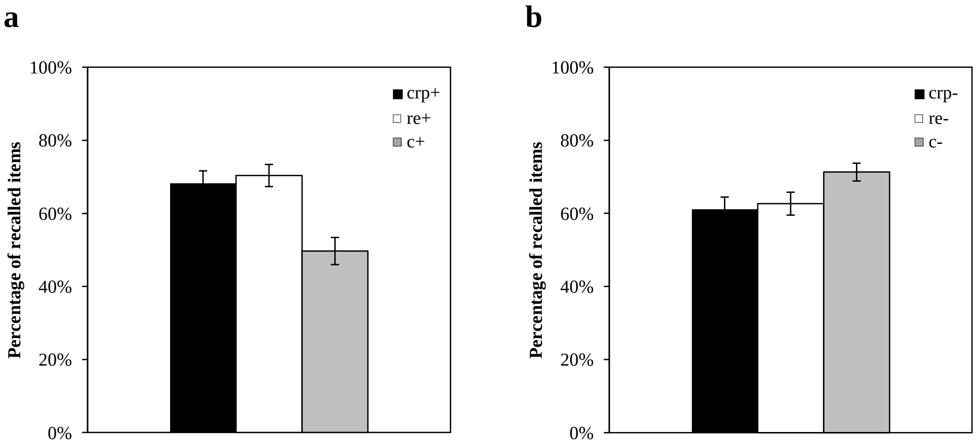

Figure 5. Results of Experiment 2a. Mean recall rates are shown as a function of item type. Error bars represent standard errors. (a) Recall percentages for competitively retrieval practiced items (crp+), re-exposed and rated items (re+), and control items (c+). (b) Recall percentages for unpracticed items of competitively retrieval practiced categories (crp-), unpracticed items of re-exposed and rated categories (re-), and control items (c-).

($SD = 0.22$) for the *c+* items. Recall levels differed significantly across item type, $F(2, 70) = 20.934$, $MSE = 0.022$, $p < .001$, $\eta^2 = 0.374$. Planned comparisons showed significant recall enhancement for both *crp+* and *re+* items when compared to *c+* items, $t(35) = 4.965$, $p < .001$, $d = 0.860$, and $t(35) = 5.462$, $p < .001$, $d = 1.030$. Recall performance of *crp+* and *re+* items did not differ reliably, $t(35) < 1$. Thus, as expected, both types of practice boosted recall of practiced items at test. Intrusion rates were .05 ($SD = 0.08$) for the *crp+* items, .06 ($SD = 0.08$) for the *re+* items, and .05 ($SD = 0.08$) for the *c+* items, and did not differ across item type, $F(2, 70) < 1$.

Regarding the detrimental effects of practice, recall rates of 61.1% ($SD = 0.20$) for the *crp-* items, 62.7% ($SD = 0.19$) for the *re-* items, and 71.3% ($SD = 0.15$) for the *c-* items were observed. Recall rates varied across item type, $F(2, 70) = 5.504$, $MSE = 0.020$, $p = .006$, $\eta^2 = 0.234$. Planned comparisons revealed significant forgetting of *crp-* relative to *c-* items, $t(35) = 3.274$, $p = .002$, $d = 0.577$, and significant forgetting of *re-* relative to *c-* items, $t(35) =$

-3.682, $p = .001$, $d = 0.502$. The numerical difference between *crp*- and *re*- items was not significant, $t(35) < 1$. Intrusion rates were .08 ($SD = 0.06$) for the *crp*- items, .10 ($SD = 0.09$) for the *re*- items, and .07 ($SD = 0.07$) for the *c*- items, and did not differ across item type, $F(2, 70) = 1.170$, $MSE = 0.007$, $p = .317$, $\eta^2 = 0.032$.

Discussion

The results demonstrate that both (competitive) retrieval practice and re-exposure with pleasantness ratings enhanced recall of the practiced items (*crp+, re+*) but reduced recall of the unpracticed items (*crp-, re-*). The finding of RIF after both practice conditions replicates the previous recall results by Verde (2013) and supports the equivalence assumption of the blocking account. In Experiment 2b, the same procedure was employed as in Experiment 2a in order to investigate whether the findings generalize to item recognition testing or whether, like in Experiments 1a and 1b, they do not.

3.4 Experiment 2b: The effects of retrieval practice and restudy with pleasantness ratings on item recognition

The goal of Experiment 2b was to extend Verde's finding, examining the effects of re-exposure supplemented with pleasantness ratings on recognition performance. Data by Grundgeiger (2014) and the present Experiment 1b indicate that RIF-like forgetting in cued recall does not necessarily generalize to item recognition tests. Therefore, Experiment 2a was repeated, however, replacing the cued recall test with an item recognition test.

Methods

Participants. Forty-eight students of Regensburg University took part in this experiment (M = 21.9 years, range = 17-28 years, 32 female).[5] All subjects spoke German as native language and received monetary reward for their participation.

Materials. The material was identical to the one employed in Experiment 2a. Additionally, six further members of each target category, that featured medium to high frequency, were chosen to be lures ($median = 8.5$).

Design and Procedure. Apart from the format of the final memory test, the same design and procedure as used in Experiment 2a was employed (see Figure 4). The procedure in the recognition test was identical to the one employed in Experiment 1b. The final test included lures that either belonged to retrieval practiced categories (*crp lures*), re-exposed and rated categories (*re lures*), or control categories (*c lures*).

Statistical Analysis. Statistical analysis of the data was analogous to Experiment 1b.

Results

Practice Phase. In the retrieval practice phase, participants successfully retrieved 67.4% (SD = 0.17) of the practiced items on the first cycle and 69.3% (SD = 0.16) of the items in total. These numbers are highly similar to

Table 3 False alarm rates and corrected hit rates for Experiment 2b

Item type		Response criteria				
		"1"	"2"	"3"	"4"	"5"
crp+	False alarms	.077	.130	.228	.374	.611
	Corrected hits	.699	.706	.649	.559	.354
re+	False alarms	.056	.114	.199	.323	.538
	Corrected hits	.868	.842	.783	.663	.460
c+	False alarms	.066	.118	.185	.313	.557
	Corrected hits	.585	.593	.581	.514	.369
crp-	False alarms	.077	.130	.228	.374	.611
	Corrected hits	.495	.572	.550	.499	.327
re-	False alarms	.056	.114	.199	.323	.538
	Corrected hits	.578	.620	.593	.545	.407
c-	False alarms	.066	.118	.185	.313	.557
	Corrected hits	.582	.646	.642	.576	.392

Note. False alarm and corrected hit rates are shown as a function of item type and response criterion. crp+ = retrieval practiced items; re+ = re-exposed and rated items; c+ = unpracticed items from unpracticed categories; crp- = unpracticed items from retrieval practiced categories; re- = unpracticed items from re-exposed and rated categories; c- = unpracticed items from unpracticed categories. “1” reflects the strictest response criterion, i.e., definitely old, and each subsequent number (“2”, “3”, etc.) reflects a more and more relaxed criterion. Corrected hits = hits - false alarms.

Table 4 Unequal-variance signal detection model for Experiment 4

	Parameter estimates		Goodness of fit		
Item type	d_a	σ	χ^2	*df*	*p*
crp+	2.65*	1.60	0.43	3	.933
re+	3.44*	1.29	0.81	3	.847
c+	2.07	1.59	1.05	3	.788
crp-	1.73*	1.27	1.41	3	.702
re-	2.04	1.39	0.92	3	.820
c-	2.08	1.32	1.24	3	.743

Note. crp+ = competitively retrieval practiced items; re+ = re-exposed and rated items; c+ = unpracticed items from unpracticed categories; crp- = unpracticed items from competitively retrieval practiced categories; re- = unpracticed items from re-exposed and rated categories; c- = unpracticed items from unpracticed categories. d_a = general memory strength; σ = variance of the target distribution.
* Significant deviations from control performance ($p < .05$).

those reported in Experiment 2a.

Recognition Test: ANOVA of Corrected Hits. Table 3 shows mean false alarm rates and corrected hit rates, separately for the five response criteria and the single item types. Regarding the beneficial effects of retrieval practice and re-exposure on the practiced items (*crp+*, *re+*) relative to their controls (*c+*), a 2 × 3 ANOVA with the factors of ITEM TYPE (*crp+*, *c+*) and

[5] In contrast to Experiment 1a and 1b, which followed Raaijmakers and Jakab (2012) and Grundgeiger (2014) and low- to medium-frequency category exemplars were used as study material, Experiments 2a-3b followed the large majority of RIF studies and medium- to high-frequency exemplars were employed. Because there is evidence that RIF is larger when using medium- to high-frequency exemplars rather than low-frequency exemplars (M. C. Anderson et al., 1994; Bäuml, 1998; Migueles & Garcia-Bajos, 2014; see also Murayama et al., 2014), the sample size was reduced for the following recognition experiments, i.e. Experiments 2b and 3b, relative to Experiment 1b.

RESPONSE CRITERION ("1", "2", "3") showed a main effect of ITEM TYPE, $F(1, 47) = 14.434$, $MSE = 0.049$, $p < .001$, $\eta^2 = 0.235$, indicating that retrieval practice was successful. Although the effect of item type varied with criterion, $F(2, 94) = 3.638$, $MSE = 0.005$, $p = .030$, $\eta^2 = 0.072$, it arose for all three response criteria, $ts(58) > 2.652$, $ps < .012$, $ds > 0.286$. Similarly, contrasting *re+* and *c+* items, a main effect of ITEM TYPE was observed, $F(1, 47) = 71.237$, $MSE = 0.061$, $p < .001$, $\eta^2 = 0.602$, that varied with response criterion, $F(2, 94) = 7.166$, $MSE = 0.006$, $p = .001$, $\eta^2 = 0.132$, but was present for all three response criteria, $ts(47) > 6.919$, $ps < .001$, $ds > 0.950$. Even though both retrieval practice and re-exposure were successful in enhancing recognition of the practiced items, corrected hits for *re+* items were higher than for *crp+* items, $F(1, 47) = 29.466$, $MSE = 0.052$, $p < .001$, $\eta^2 = 0.385$, indicating that re-exposure enhanced recognition more than competitive retrieval practice. The interaction effect was not reliable, $F(2, 49) = 1.419$, $MSE = 0.007$, $p = .247$, $\eta^2 = 0.029$.

Regarding the detrimental effects of retrieval practice on the unpracticed items (*crp-*, *re-*) relative to their controls (*c-*), a 2×3 ANOVA with the factors of ITEM TYPE (*crp-*, *c-*) and RESPONSE CRITERION ("1", "2", "3") showed a main effect of ITEM TYPE, $F(1, 47) = 10.617$, $MSE = 0.048$, $p = .002$, $\eta^2 = 0.184$, with lower corrected hits for *crp-* than *c-* items, but no interaction between the two factors, $F(2, 94) < 1$. An analogous analysis contrasting *re-* and *c-* items showed no main effect of ITEM TYPE, $F(1, 47) < 1$, and no interaction between the two factors, $F(2, 94) = 1.718$, $MSE = 0.007$, $p = .185$, $\eta^2 = 0.035$. These results indicate that competitive retrieval practice induced RIF, whereas re-exposure did not induce any RIF-like forgetting. Consistently, an ANOVA contrasting *crp-* and *re-* items showed a main effect of ITEM TYPE, $F(1, 47) = 4.039$, $MSE = 0.061$, $p = .050$, $\eta^2 = 0.079$, and no interaction between the two factors, $F(2, 94) = 1.619$, $MSE = 0.007$, $p = .204$, $\eta^2 = 0.033$, indicating that the RIF effect was retrieval specific.

Recognition Test: Analysis of Hit and False Alarm Rates using the Unequal-Variance Signal Detection Model. In the next step, the

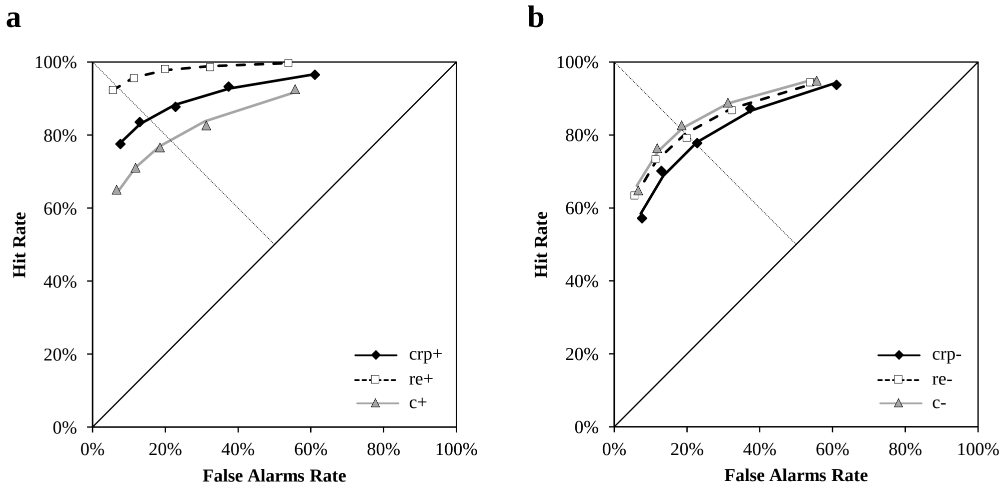

Figure 6. Item recognition Receiver Operating Characteristics (ROCs) depicting cumulative hit and false alarm rates as a function of item type. Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model. (a) ROCs for competitively retrieval practiced items (crp+), re-exposed and rated items (re+), and control items (c+). (b) ROCs for unpracticed items of competitively retrieval practiced categories (crp-), unpracticed items of re-exposed and rated categories (re-), and control items (c-).

unequal-variance signal detection model was employed to analyze hits and false alarms for the single response criteria. Figures 6a and 6b depict the ROCs for the practiced items, the unpracticed items, and the respective control items in the retrieval and re-exposure conditions, as well as the fit of the unequal-variance signal detection model to the data of each single condition. Table 4 shows the statistics of goodness-of-fit and maximum-likelihood estimates of the model's parameters d_a and σ for the practiced, unpracticed, and control items.

The unequal-variance signal detection model described the data of the six item types well, all $\chi^2 s(3) < 1.415$, $ps > .701$. Both the retrieval practiced (*crp+*) items and the reexposed (*re+*) items showed enhanced memory strength as measured by d_a relative to the control (*c+*) items, $\chi^2(1) = 5.488$, $p = .019$, and $\chi^2(1) = 16.420$, $p < .001$, indicating improved recognition of the practiced items after both retrieval and re-exposure. Retrieval and re-exposure differed marginally in their effects on d_a for the practiced items, $\chi^2(1) = 3.436$, $p = .064$, with a trend for higher d_a after re-exposure than retrieval practice. Critically, retrieval practice reduced d_a for unpracticed (*crp-*) items relative to the control (*c-*) items, $\chi^2(1) = 5.566$, $p = .018$, whereas re-exposure did not affect memory strength of the unpracticed (*re-*) items, $\chi^2(1) = 0.053$, $p = .818$. The difference in d_a between the two types of unpracticed items reached significance, $\chi^2(1) = 4.480$, $p = .034$, indicating that practice induced a detrimental effect of retrieval practice but not of re-exposure.

Further parallels to Experiment 1b arose. First, for both the practiced items and their controls, and the unpracticed items and their controls, the variance of the old items' distribution, σ, did not vary significantly across item type, $\chi^2 s(2) < 1.164$, $ps > .558$, but was larger than 1.0, $\chi^2 s(1) > 31.499$, $ps < .001$. Second, the placement of the five confidence criteria varied across item type, $\chi^2 s(10) > 26.480$, $ps < .003$.

Discussion

The results of Experiment 2b again replicate prior RIF work by finding retrieval practice to enhance recognition of the practiced items (*crp+*) but to reduce recognition of the unpracticed items (*crp-*; e.g., Hicks & Starns, 2004). Going beyond prior work, the results show that re-exposure when supplemented with a pleasantness rating task can enhance recognition of the practiced items (*re+*) but leaves recognition of the unpracticed items (*re-*) unaffected. The finding of reduced recognition of unpracticed items after retrieval practice but not after re-exposure arose from both analysis of corrected hits and signal detection analysis and indicates that, like the noncompetitive retrieval practice condition in Experiment 1b, re-exposure supplemented with pleasantness ratings does not reduce recognition of the unpracticed items.

Crucially, because exactly the same materials and study and practice procedures were employed as in Experiment 2a, the results of Experiments 2a and 2b indicate that (competitive) retrieval practice reduces both recall and recognition of unpracticed items, whereas re-exposure with pleasantness ratings reduces recall but not recognition of these items. The findings suggest that the equivalence assumption of the blocking account may hold when employing recall tests but may not hold when employing item recognition. This suggests that strengthening of cue-item associations via re-exposure and pleasantness ratings is not sufficient to induce RIF, at least when recognition memory is assessed. As a consequence, the findings challenge the blocking account of RIF.

In Experiments 3a and 3b, the effects of a third re-exposure method on recall and recognition are investigated that may also strengthen cue-item associations. It is examined whether the findings of Experiments 1 and 2 showing that retrieval practice and re-exposure induce forgetting in cued recall, but only retrieval practice induces RIF in item recognition tests, will hold when a different form of strengthening is employed.

3.5 Experiment 3a: The effects of retrieval practice and restudy with visualization on cued recall

Experiment 3a examined another re-exposure format that in previous work was shown to induce forgetting of unpracticed items, namely mental imagery. Saunders et al. (2009) found that visualization of particular features of a previously studied and re-exposed exemplar, such as the size, shape, or color of the item, can reduce recall of the unpracticed items, very similar to how retrieval practice does. Analogous to Experiments 1a and 2a, the effects of (competitive) retrieval practice were compared with the effects of such mental imagery, examining whether both types of practice reduce later recall of unpracticed items. On the basis of the previous study, and analogous to Experiments 1a and 2a, reliable RIF was expected to arise after (competitive) retrieval practice and reliable RIF-like forgetting was expected to arise after re-exposure supplemented with mental imagery.

Methods

Participants. Thirty-six students of Regensburg University participated in this experiment (M = 20.8 years, range = 18-27 years, 25 female). All participants spoke German as native language. Monetary reward was provided in exchange for participation.

Materials. Nine categories (TOOLS, BIRDS, FLOWERS, DRINKS, FRUITS, MUSICAL INSTRUMENTS, FURNITURE, SPICES, CLOTHING) with six study items each were drawn from published word norms (Battig & Montague, 1969; Mannhaupt, 1983; Scheithe & Bäuml, 1995; van Overschelde, Rawson, & Dunlosky, 2004) to compile the study list. Six of these categories matched the material used by Saunders et al. (2009).[6] Three additional categories (GEMS, AFRICAN STATES, SANITARY ARTICLES) with two items each served as buffer items. Practiced and unpracticed items were of medium to high frequency (*median* = 6.0). Studied items within each category had a unique first letter.

Design and Procedure. The experiment had a 2 × 3 design with the within-subject factors of PRACTICE TYPE (competitive retrieval, re-exposure) and ITEM TYPE (practiced, unpracticed, control). The procedure followed Experiments 1a and 2a save for the intermediate practice phase (see Figure 7). In the retrieval condition, the category label and the first letter of the to-be-retrieved item were presented (e.g., INSECT - *t*___) and subjects were asked to retrieve the corresponding exemplar within 5 s (ISI = 500 ms).

[6]Two categories from Saunders et al.'s original material (SPORTS EQUIPMENT, WEAPONS) were substituted with three categories from Experiment 2a and the remaining six categories were completed with lures from published word norms. The replaced categories did not provide enough items with unique first letters in the German language and/or reasonable frequencies to match the other categories. Furthermore, three items of the Saunders et al.'s material were replaced due to their absence in any of the available norms.

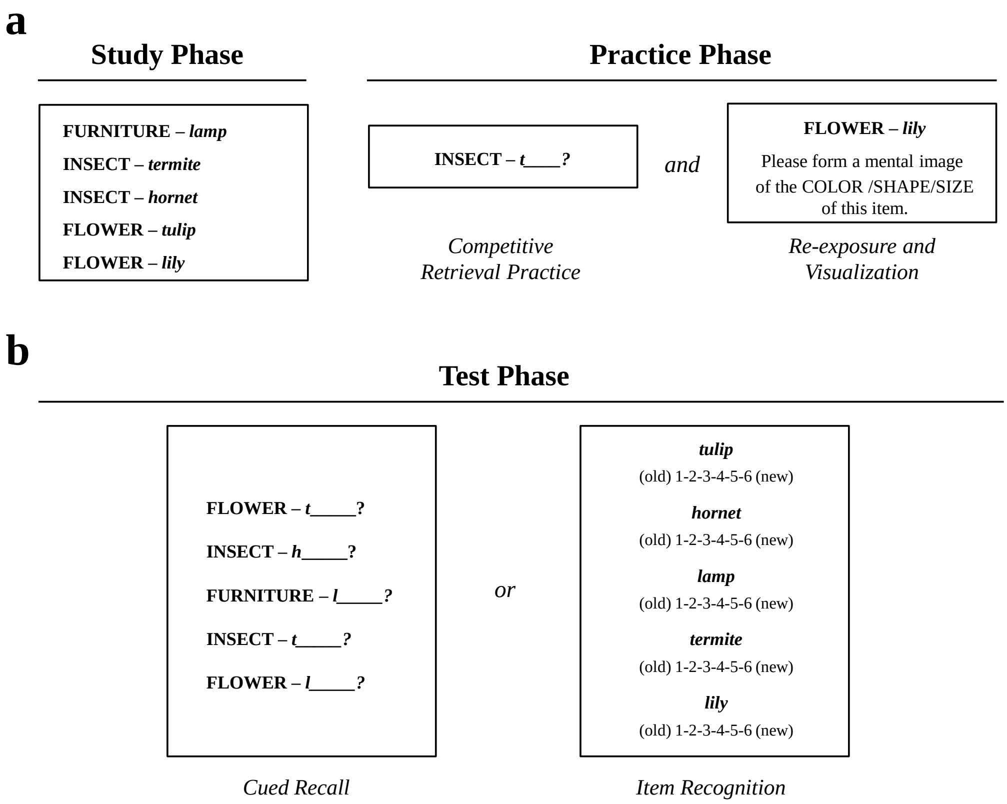

Figure 7. Procedure and conditions employed in Experiments 3a and 3b. (a) Study and practice phases of Experiments 3a and 3b: Participants studied a list of categorized items. In the practice phase, participants practiced some items by retrieving the exemplar (competitive retrieval practice) and some items by visualizing an item feature (re-exposure and visualization). (b) Test phase: In Experiment 3a, the category label and the exemplar's initial letter were provided and participants were asked to recall the exemplar (cued recall). In Experiment 3b, all studied exemplars and lures were provided and participants were asked to make old/new judgments (item recognition).

Following Saunders et al. (2009), no feedback was provided, and participants completed three consecutive cycles of retrieval practice. A distractor task of 135 s duration (summation of three-digit numbers) was included after the retrieval condition to match the time frame of the two practice conditions. Analogous to Saunders et al.'s experiment, in the re-exposure-plus-imagery condition, participants were re-exposed to the category label and the item for 10 s (ISI = 500 ms; e.g., INSECT - *termite*) and were instructed to visualize either the size, shape, or color of the item. Over the three cycles of practice, the to-be-imagined feature was held constant within subjects. The to-be-imagined feature as well as the order of practice conditions were counterbalanced across subjects.[7]

For each practice condition, three types of items were generated: practiced items, i.e., retrieval practiced (*crp+*) and reexposed and visualized (*re+*) items; unpracticed items of practiced categories (*crp-*, *re-*); and items of unpracticed categories (*c+*, *c-*).

Results

Practice Phase. Recall performance in the (competitive) retrieval practice condition mounted up to 60.5% ($SD = 0.20$) of the items on the first practice cycle, and to 63.7% ($SD = 0.20$) of the items in total.

Recall Test. Figures 8a and 8b show percentages of correctly recalled practiced and unpracticed items in the retrieval practice and re-exposure conditions, and of the corresponding control items. Regarding the beneficial effects of practice, participants recalled on average 62.7% ($SD = 0.23$) of the *crp+* items, 71.6% ($SD = 0.21$) of the *re+* items, and 44.4% ($SD = 0.19$) of the *c+* items. Recall levels differed significantly across item type, $F(2, 70) = 22.099$, $MSE = 0.031$, $p < .001$, $\eta^2 = 0.387$. When compared to the *c+*

[7]Saunders et al. (2009) employed a stronger version of mental imagery than the present experiment did. In lieu of visualizing only one particular feature in three cycles, participants in Saunders et al.'s study imagined four different features of each exemplar in four successive cycles. Moreover, the four blocks were interspersed with distractors while the three cycles in the present experiment were completed continuously. Doing so, Saunders et al. found very high levels of imagery-induced forgetting (31%), which exceeded the detrimental effect after retrieval practice in their own experiment (17%) and also exceeded the detrimental effects of noncompetitive retrieval practice and re-exposure supplemented with pleasantness ratings as they were reported in the previous studies by Raaijmakers and Jakab (2012) and Verde (2013; 6% and 14%, respectively). To improve comparability of experiments within this study, a weaker version of mental imagery was employed using only three cycles of practice, a constant feature to visualize, and no distractors between cycles. As can be seen in the Results section of Experiment 3b below, even with this reduced version, re-exposure supplemented with mental imagery led to higher recognition of practiced items than retrieval practice did.

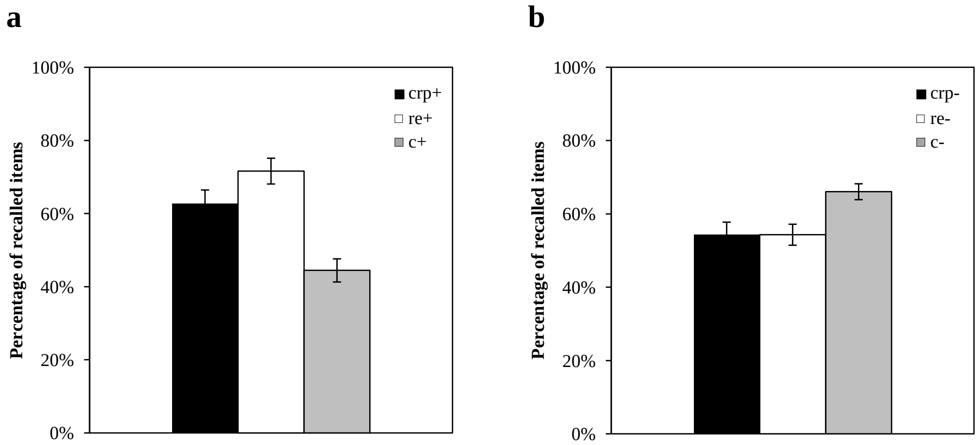

Figure 8. Recall results of Experiment 3a. Mean recall rates are shown as a function of item type. Error bars represent standard errors. (a) Recall percentages for competitively retrieval practiced items (crp+), re-exposed and visualized items (re+), and control items (c+). (b) Recall percentages for unpracticed items of competitively retrieval practiced categories (crp-), unpracticed items of re-exposed and visualized categories (re-), and control items (c-).

items, retrieval practice facilitated recall of both the *crp+* and the *re+* items, $t(35) = 4.595$, $p < .001$, $d = 0.855$, and $t(35) = 6.215$, $p < .001$, $d = 1.351$. Mental imagery boosted recall of the practiced items significantly more than retrieval practice did, $t(35) = 2.158$, $p = .038$, $d = 0.405$. Intrusion rates were .06 ($SD = 0.08$) for the *crp+* items, .04 ($SD = 0.08$) for the *re+* items, and .07 ($SD = 0.09$) for the *c+* items, and did not differ across item type, $F(2, 70) = 1.99$, $MSE = 0.005$, $p = .144$, $\eta^2 = 0.054$.

Regarding the detrimental effects of practice, recall rates for *crp-* items, *re-* items, and *c-* items reached 54.3% ($SD = 0.21$), 54.3% ($SD = 0.17$), and 66.1% ($SD = 0.13$), respectively. Recall levels varied significantly across item type, $F(2, 70) = 9.304$, $MSE = 0.018$, $p < .001$, $\eta^2 = 0.210$. Compared to *c-* items, both *crp-* and *re-* items showed a reliable reduction in recall, $t(35) = 4.245$, $p < .001$, $d = 0.698$, and $t(35) = 3.952$, $p < .001$, $d = 0.789$, but there was no difference in recall levels between *crp-* and *re-* items, $t(35) < 0.001$. Intrusion rates were .06 ($SD = 0.07$) for the *crp-* items, .04 ($SD = 0.06$) for the *re-*

items, and .07 (SD = 0.07) for the c- items, and did not differ across item type, $F(2, 70) < 1$.

Discussion

Both (competitive) retrieval practice and re-exposure augmented with mental imagery enhanced recall of the practiced items (*crp+, re+*) but reduced recall of the unpracticed items (*crp-, re-*). The observed forgetting after both practice conditions replicates the previous recall result by Saunders et al. (2009). The data thus corroborate the equivalence assumption of the blocking account when cued recall tests are administered supporting the idea that strengthening of cue-item associations is sufficient to induce RIF-like forgetting.

3.6 EXPERIMENT 3B: THE EFFECTS OF RETRIEVAL PRACTICE AND RESTUDY WITH VISUALIZATION ON ITEM RECOGNITION

Saunders et al. (2009) as well as the results of Experiment 3a showed that imagining particular attributes of re-exposed category exemplars during practice can reduce recall of the not re-exposed items at test. The goal of Experiment 3b was to investigate whether this finding generalizes from recall to item recognition.

Methods

Participants. Another 48 students of Regensburg University participated in the experiment (M = 21.0 years, range =18-29 years, 43 female). All subjects spoke German as native language. In exchange for participation, monetary reward was provided.

Materials. The same material was used as in Experiment 3a. Six lures were drawn from each category ($median = 9.5$) within the range of medium to high frequency.

Design and Procedure. Design and procedure of Experiment 3b were identical to Experiment 3a with the only exception that the final recall test was replaced by a recognition procedure (see Figure 7). The procedure in the recognition test was identical to the ones employed in Experiments 1b and 2b. The final test included lures that either belonged to retrieval practiced categories (*crp lures*), re-exposed and visualized categories (*re lures*), or control categories (*c lures*).

Statistical Analysis. The same statistical analyses as in Experiments 1b and 2b were employed.

Results

Practice Phase. Success rates in the retrieval practice phase were 66.0% ($SD = 0.18$) in the first cycle and 69.7% ($SD = 0.16$) in total. These numbers are similar to those reported in Experiment 3a.

Recognition Test: ANOVA of Corrected Hits. Table 5 shows mean false alarm rates and corrected hit rates, separately for the five response criteria and the single item types. Regarding beneficial the effects of retrieval practice and re-exposure on the practiced items (*crp+*, *re+*) relative to their controls (*c+*), a 2 × 3 ANOVA with the factors of ITEM TYPE (*crp+*, *c+*) and RESPONSE CRITERION ("1", "2", "3") showed a main effect of ITEM TYPE, $F(1, 47) = 27.519$, $MSE = 0.068$, $p < .001$, $\eta^2 = 0.369$, indicating that retrieval practice was successful. Like in Experiments 1a and 2a, the effect varied with criterion, $F(2, 94) = 4.537$, $MSE = 0.004$, $p = .013$, $\eta^2 = 0.088$, but was present for all three criteria, all $ts > 4.103$, all $ps < .001$, all $ds > 0.668$. An analogous analysis contrasting *re+* and *c+* items also showed a main effect of ITEM TYPE, $F(1, 47) = 136.946$, $MSE = 0.049$, $p < .001$, $\eta^2 = 0.744$, and a reliable interaction between the two factors, $F(2, 94) = 13.220$, $MSE = 0.005$, $p < .001$, $\eta^2 = 0.220$. Again, the practice effect arose for all three criteria, all $ts > 9.395$, all $ps < .001$, all $ds > 1.560$. Corrected hits for *re+* items were higher than for *crp+* items, $F(1, 47) = 39.239$, $MSE = 0.039$, $p < .001$, $\eta^2 = 0.455$, indicating that re-exposure enhanced recognition more than competitive retrieval practice. Moreover, the interaction was significant, $F(2, 94) = 5.520$, $MSE = 0.003$, $p = .005$, $\eta^2 = 0.105$ and t-tests showed significant differences across all response criteria, all $ts > 5.245$, all $ps < .001$, all $ds > 0.746$.

Regarding the detrimental effects of retrieval practice on the unpracticed items (*crp-*, *re-*) relative to their controls (*c-*), a 2×3 ANOVA with the factors of ITEM TYPE (*crp-*, *c-*) and RESPONSE CRITERION ("1", "2", "3") showed a main effect of ITEM TYPE, $F(1, 47) = 7.283$, $MSE = 0.076$, $p = .010$, $\eta^2 = 0.134$, with lower corrected hits for *crp-* than *c-* items, but no interaction between the two factors, $F(2, 94) < 1$. A similar analysis contrasting *re-* and *c-* items showed no main effect of ITEM TYPE, $F(1, 47) < 1$, and no interaction between the two factors, $F(2, 94) < 1$. Thus, competitive retrieval practice induced RIF, whereas re-exposure did not induce any RIF-like forgetting. Consistently, there was a main effect of ITEM TYPE when contrasting *crp-* and *re-* items, $F(1, 47) = 8.540$, $MSE = 0.056$, $p = .005$, $\eta^2 = 0.154$, but no interaction between ITEM TYPE and RESPONSE CRITERION, $F(2, 94) < 1$. These results indicate that the RIF findings were retrieval specific.

Recognition Test: Analysis of Hit and False Alarm Rates using the Unequal-Variance Signal Detection Model. Next, the unequal-variance signal detection model was employed. The ROCs in Figures 9a and 9b depict the cumulated hit and false alarm rates for each item type and practice condition and the fit of the unequal-variance signal detection model to the recognition data of each single condition. Goodness-of-fit statistics and maximum-likelihood estimates of the parameters d_a and σ for practiced, unpracticed, and control items are summarized in Table 6.

Table 5 False alarm rates and corrected hit rates for Experiment 3b

		Response criteria				
Item type		"1"	"2"	"3"	"4"	"5"
crp+	False alarms	.058	.115	.191	.329	.604
	Corrected hits	.780	.781	.733	.629	.378
re+	False alarms	.041	.077	.142	.259	.516
	Corrected hits	.952	.921	.856	.739	.482
c+	False alarms	.040	.094	.158	.294	.552
	Corrected hits	.594	.615	.601	.537	.372
crp-	False alarms	.058	.115	.191	.329	.604
	Corrected hits	.472	.556	.561	.514	.331
re-	False alarms	.041	.077	.142	.259	.516
	Corrected hits	.552	.634	.650	.605	.405
c-	False alarms	.040	.094	.158	.294	.552
	Corrected hits	.574	.633	.645	.581	.390

Note. False alarm and corrected hit rates are shown as a function of item type and response criterion. crp+ = retrieval practiced items; re+ = re-exposed and visualized items; c+ = unpracticed items from unpracticed categories; crp- = unpracticed items from retrieval practiced categories; re- = unpracticed items from re-exposed and visualized categories; c- = unpracticed items from unpracticed categories. “1” reflects the strictest response criterion, i.e., definitely old, and each subsequent number (“2”, “3”, etc.) reflects a more and more relaxed criterion. Corrected hits = hits false alarms.

Table 6 Unequal-variance signal detection model for Experiment 3b

	Parameter estimates		Goodness of fit		
Item type	d_a	σ	χ^2	*df*	*p*
crp+	3.22*	1.63	0.33	3	.958
re+	14.55*	4.95	0.97	3	.809
c+	2.30	1.76	0.97	3	.808
crp-	1.73*	1.29	0.70	3	.872
re-	2.22	1.50	2.88	3	.411
c-	2.20	1.46	0.31	3	.958

Note. crp+ = competitively retrieval practiced items; re+ = re-exposed and visualized items; c+ = unpracticed items from unpracticed categories; crp- = unpracticed items from competitively retrieval practiced categories; re- = unpracticed items from re-exposed and visualized categories; c- = unpracticed items from unpracticed categories. d_a = general memory strength; σ = variance of the target distribution.
* Significant deviations from control performance ($p < .05$).

The unequal-variance signal detection model described the data of the six item types well, all $\chi^2s(3) < 2.878$, $ps > .410$. Both for *crp+* and *re+* items memory strength as measured by d_a was larger than for *c+* items, $\chi^2s(1) > 9.500$, $ps < .002$, with the *re+* items showing higher d_a than the *crp+* items, $\chi^2(1) = 13.353$, $p < .001$, which indicates that re-exposed items gained more strength through practice than retrieval practiced items. Regarding the detrimental effects of practice, discriminability of *c-* items significantly exceeded discriminability of *crp-* items, $\chi^2(1) = 11.004$, $p < .001$, but did not exceed discriminability of *re-* items, $\chi^2(1) < 0.004$, $p = .950$. Consistently, d_a of *re-* items was larger than of *crp-* items, $\chi^2(1) = 11.706$, $p < .001$, indicating that forgetting of unpracticed items occurred after retrieval practice but not

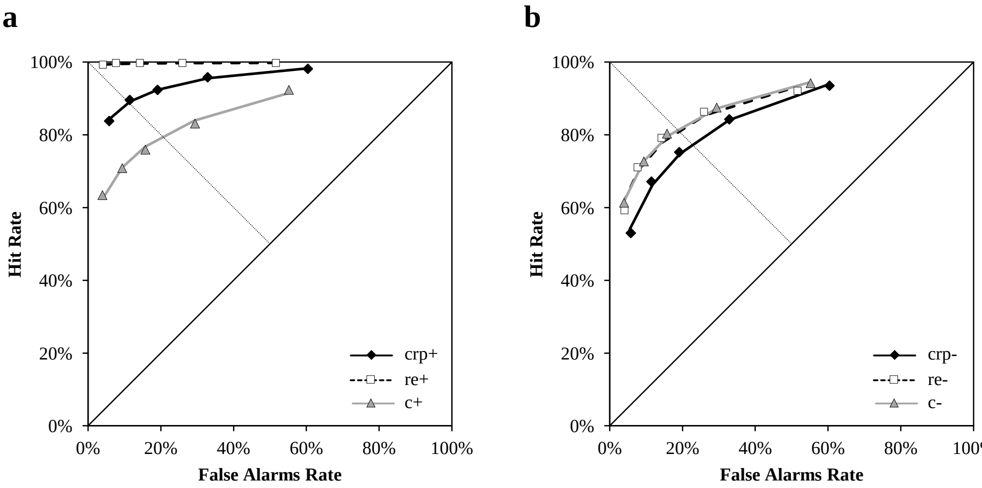

Figure 9. Item recognition Receiver Operating Characteristics (ROCs) depicting cumulative hit and false alarm rates as a function of item type. Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model. (a) ROCs for competitively retrieval practiced items (crp+), re-exposed and visualized items (re+), and control items (c+). (b) ROCs for unpracticed items of competitively retrieval practiced categories (crp-), unpracticed items of re-exposed and visualized categories (re-), and control items (c-).

after re-exposure of the practiced items. Like in Experiments 1b and 2b, for both the practiced items and their controls, and the unpracticed items and their controls, the variance of the old items' distribution, σ, did not vary significantly across item type, $\chi^2s(2) < 2.721$, $ps > .256$, but was larger than 1.0, $\chi^2s(1) > 52.488$, $ps < .001$, and the placement of the five confidence criteria varied across item type, $\chi^2s(10) > 42.514$, $ps < .001$.

Discussion

The results of Experiment 3b again replicate prior work showing RIF in item recognition tests. Moreover, the results extend previous findings. While recognition impairment was observed after retrieval practice, no such impairment was found when a subset of items was visualized. Critically, in Experiment 3b, exactly the same materials and study and practice procedures were employed as in Experiment 3a. Thus, the same materials and visualization procedure produced forgetting in recall, but failed to do so in an item recognition test. Together the findings of Experiments 3a and 3b indicate that (competitive) retrieval practice reduces both recall and recognition of unpracticed items (*crp*-), whereas re-exposure with mental imagery reduces recall but not recognition of these items (*re*-). These results provide a further demonstration of the dissimilar effects of competitive retrieval practice and strength-based practice methods in item recognition tests challenging the blocking account of RIF.

As a whole, the results of Experiments 1-3 replicate and extend prior work on Retrieval-induced Forgetting. They showed that selectively retrieving a subset of category exemplars enhanced memory performance of these practiced items while impairing memory performance of unpracticed category members. These findings held for both test types, cued recall and recognition (see also M. C. Anderson et al., 1994; Hicks & Starns, 2004). Moreover, the results replicate earlier findings supporting the equivalence assumption of the blocking account in cued recall tests. Not only after retrieval practice, but also after particular re-exposure formats that are supposed to strengthen cue-item associations RIF-like enhancement and forgetting effects arose, when memory performance was assessed employing a cued recall test (see also Raaijmakers & Jakab, 2012; Saunders et al., 2009; Verde, 2013). Thus, the assumption by the blocking account that RIF is equivalent to strength-based forgetting, i.e., re-exposure is sufficient to induce forgetting, whereas retrieval is not necessary, was sustained suggesting a role for strength-based blocking effects in RIF, at least when recall tests are employed.

However, the present results replicate and augment previous work in showing that these strength-dependent forgetting effects do not generalize to item recognition testing whereas RIF does. Like in Grundgeiger's (2014) study, noncompetitive retrieval practice induced forgetting in cued recall, but not in item recognition tests suggesting that test format may determine whether RIF and strength-dependent RIF-like forgetting are equivalent or not. The present experiments provided additional evidence that this finding can be extended to other re-exposure formats, that were shown to induce RIF-like forgetting in recall. Both re-exposure supplemented with pleasantness ratings and re-exposure supplemented with imagery were found to impair recall for unpracticed items, yet to leave recognition of unpracticed items unaffected. Thus, the equivalence assumption of the blocking account seems not to hold when assessing recognition performance at test, as retrieval but not any of the employed restudy formats reduced recognition memory after practice. This finding challenges the idea that RIF is caused solely by strength-based blocking processes that arise due to considerable strengthening of cue-target associations during practice. If RIF were equivalent to strength-based blocking processes then other forms of practice that enhance cue-item associations should induce forgetting whenever competitive retrieval induces forgetting.

Chapter 4

Experiments 4-6: Testing the context change account

The second set of experiments was designed to test the context change account of RIF (Jonker et al., 2013), in particular the implied equivalence assumption. The context change account of RIF ascribes the observed impairment of unpracticed items in the Retrieval-practice Paradigm to a mismatch of context cues during study and test. Accordingly, this contextual mismatch arises because the act of retrieval accelerates context fluctuation and thus creates distinct contexts for study and practice that are associated to the control and practiced categories, respectively. Following this line of argumentation, RIF should be equivalent to any kind of practice if it is preceded by a shift in context. This equivalence implies that whenever RIF occurs, so should context-dependent forgetting in the Retrieval-practice Paradigm.

Based on the mixed results of context change effects in item recognition tests (e.g., Bodner & Lindsay, 2003; Bodner & Richardson-Champion, 2007; Fernandez & Glenberg, 1985; Sego et al., 2006), test mode was varied using either cued recall or item recognition. In three experiments (Experiments 4-6), it was examined whether the effects of standard (competitive) retrieval practice (Experiment 4) are imitated by Jonker et al.'s (2013) re-exposure preceded by imagination (Experiment 5), and re-exposure preceded by semantic generation (Experiment 6) in a cued recall test and in an item recognition test. As context change manipulations impeded the blocked practice design used in Experiments 1-3, here, practice mode was varied across experiments. The induction of a context change by the imagination or semantic generation tasks would not only affect restudied items but also retrieval practiced items confounding the comparison of their effects on recall and recognition. Instead, cued recall and recognition data were collected within subjects. Like in Experiments 1-3, subjects studied categorized items. Next, in an intermediate phase, participants either engaged in unrelated tasks that were not supposed to shift context (counting backwards and simple calculations, Experiment 4), or engaged in tasks that have been shown to accelerate context drift (imagining particular scenarios, Experiment 5; semantic generation, Experiment 6). After a short delay, the to-be-studied items were tested using either cued recall or item recognition. A second block followed that differed only in material and final test format.

According to the context change account, which states that retrieval practice induces a context change and the resulting contextual mismatch causes RIF, re-exposure preceded by forms of mental context change should produce equivalent results as standard retrieval practice. In the cued recall test, it was expected that all three forms of practice induced forgetting of unpracticed items replicating prior findings by Jonker et al. (2013) and extending them to semantic generation as a viable method to induce mental context change. The context change account would further predict that, depending on the presence or absence of context effects in item recognition tests, either all or none of the employed practice formats - (competitive) retrieval practice (Experiment 4), restudy preceded by imagination (Experiment 5), and restudy preceded by semantic generation (Experiment 6) - should induce forgetting in item recognition. In either case, the results would support the equivalence assumption of the context change account of RIF.

4.1 EXPERIMENT 4: THE EFFECTS OF RETRIEVAL PRACTICE ON CUED RECALL AND ITEM RECOGNITION

Experiment 4 was conducted in order to replicate earlier findings, i.e., retrieval induces both enhancement of practiced items and forgetting of unpracticed items in cued recall tests as well as recognition (e.g., M. C. Anderson et al., 1994; Grundgeiger, 2014; Hicks & Starns, 2004; see also Experiments 1-3), with the materials and procedures used in Experiments 5 and 6. In this experiment, participants studied category-exemplar pairs, practiced a subset of items by retrieving the exemplar given the category and initial letter as cues, and, after a distractor, were either asked to recall the studied items using a category-plus-initial-letter recall test or to recognize them. The second block was identical except for the material and the format of the final test. Crucially, no context change task was implemented prior to the practice phase. Based on previous work, the expectation arose that retrieval induces enhancement of practiced items and impairment of unpracticed items compared to a baseline in both, cued recall and item recognition tests.

Methods

Participants. Forty-eight students of Regensburg University took part in the experiment (M = 22.83 years, $range = 18 - 29$ years, 43 female). They spoke German as native language. Monetary reward was provided in exchange for participation.

Materials. Sixteen semantic categories with six to-be-studied items and six lure items were drawn from published German word norms (Mannhaupt, 1983; Scheithe & Bäuml, 1995). Categories were allocated to one of two item sets (set 1: STATES OF THE U.S.A., MUSICAL INSTRUMENTS, FLOWERS, INSECTS, CAR EQUIPMENT, FRUITS, BIRDS, SPICES; set 2: AFRICAN STATES, KINDS OF FISH, PROFESSIONS, HOBBIES, TREES, KITCHEN EQUIPMENT, FOUR-LEGGED ANIMALS, ARTICLES OF CLOTHING). Additionally, three categories (set 1: PARTS OF GRAMMAR, SANITARY ARTICLES, TOYS; set 2: ALCOHOLIC BEVERAGES, PARTS OF THE BODY, RELATIVES) with two exemplars each were selected and used as buffer items in the study and recognition lists. The German translations of the category labels of the sixteen experimental categories consisted of a single word. The to-be-studied exemplars within each category had a unique initial letter. With respect to their frequency in the word norms, items were alternately assigned to be study items or to be lure items. The medians of the studied items were 12.5 (set 1) and 14.0 (set 2); the medians of the lure items were 14.5 (set 1) and 11.0 (set 2).

Design. The experiment had a 3 × 2 design with the within-subject factors of ITEM TYPE (practiced, unpracticed, control) and TEST TYPE (cued recall, recognition). The experiment consisted of two blocks that were identical apart from materials (set 1 or set 2) and test type (cued recall or recognition). The order of test type and material was counterbalanced across subjects. In both blocks, participants completed four main phases: an initial study phase, an intermediate phase, a practice phase, and a final test phase. In the practice phase, participants practiced three exemplars of four categories. The remaining four categories served as control categories. By this, three types of items were created: (competitively) retrieval practiced items (*crp+*); unpracticed items of retrieval practiced categories, i.e., items that were members of the same category as the *crp+* items but were not retrieved in the practice phase (*crp-*); and items from unpracticed categories that served as controls for the practiced (*c+*) and unpracticed (*c-*) items. Categories were counterbalanced across participants to be either practiced or not practiced (control). Hence, items that were practiced (*crp+* items) by half of the participants served as unpracticed control items (*c+* items) for the other half of the participants (analogously for *crp-* and *c-* items).

For the final recognition test, four further item types were generated, i.e., exemplars of retrieval practiced categories (*crp+* and *crp-* lures) and of control categories (*c+* and *c-* lures) that had not been presented in any other phase of the experiment. As described in more detail below, the differentiation between *crp+* and *crp-* lures and *c+* and *c-* lures merely stems from testing position in the recognition test: *crp-* and *c-* lures were presented in the first half of the recognition test alongside the *crp-* and *c-* items; *crp+* and *c+* lures were presented in the second half alongside the *crp+* and *c+* items. In the respective test type condition, all items were tested within the same single recognition test.

Procedure. The procedure of Experiment 4 is depicted in Figure 10a (study phase, intermediate phase, and practice phase) and 10d (test phase).

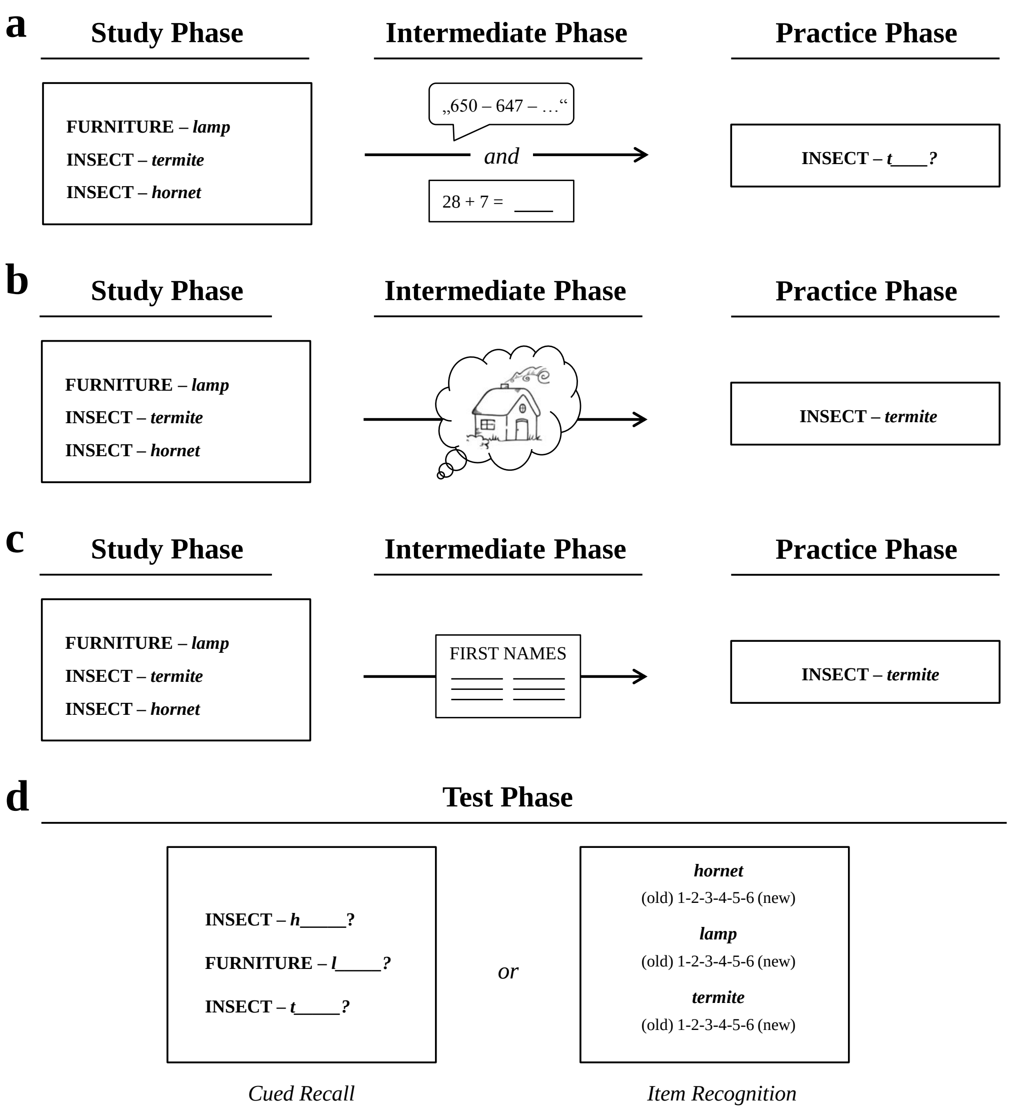

Figure 10. Procedure and conditions employed in Experiments 4-6. (a) Study and practice phases of Experiment 4: Participants studied a list of categorized items. Before practice, participants engaged in backwards counting and simple calculations. In the practice phase, participants practiced some items by retrieving the exemplar (competitive retrieval practice). (b) Study and practice phases of Experiment 5: Participants studied a list of categorized items. Before practice, participants engaged in imagination tasks. In the practice phase, participants practiced some items by restudying the pair (restudy). (c) Study and practice phases of Experiment 6: Participants studied a list of categorized items. Before practice, participants engaged in semantic generation of exemplars to novel categories. In the practice phase, participants practiced some items by restudying the pair (restudy). (d) Test phase: In one block of the experiments, the category label and the exemplar's initial letter were provided and participants were asked to recall the exemplar (cued recall). In another block of the experiments, all studied exemplars and lures were provided and participants were asked to make old/new judgments (item recognition).

Participants completed two blocks, with a five-minute break between blocks. In the study phase of each block, participants studied category-exemplar pairs (e.g., FURNITURE - *lamp*, INSECT - *termite*, INSECT - *hornet*) at a 4 s rate (ISI = 500 ms) displayed on a computer screen. The order of word pairs in the study list was blocked randomized: Six blocks were compiled, each block comprised one exemplar from each category. Order of blocks and order of word pairs within the blocks were random. Three buffer items were presented at the beginning and ending of the study list.

The study phase was followed by an intermediate phase: After studying the category-exemplar pairs, participants counted backwards in steps of three from a three-digit number for 60 s. Subsequently, simple math tasks (addition of and subtraction of two- and one-digit numbers) were provided for further 3 min. The intermediate phase in each block took 4 min to complete. These tasks were used to approximately match the time frame of the context change tasks employed in Experiments 5 and 6, and were supposed to not induce any context change (e.g., Klein, Shiffrin, & Criss, 2007). In the subsequent practice phase, half of the exemplars from half of the categories were retrieved from memory: The category label and the initial letter of an exemplar were presented for 4 s (ISI = 500 ms; e.g., INSECT - *t*___) and participants were instructed to recall the matching exemplar from the study list orally while the experimenter logged the data. Presentation was blocked randomized. The twelve exemplars were practiced twice in consecutive cycles. No feedback was provided. Before the final test, participants worked on a distractor task for another 4 min (Frankfurter Aufmerksamkeitsinventar 2, FAIR-2, Moosbrugger, Oehlschlägel, & Steinwascher, 2011).

At test, participants engaged either in a cued recall or an item recognition test. The cued recall test followed the procedure used in Experiments 1a, 2a, and 3a: The studied items were cued with the category label and the initial letter of the exemplar (e.g., INSECT - *h*___). Participants were asked to orally respond with the corresponding item within 5 s (ISI = 500 ms). The experimenter recorded the answers. Unpracticed items of retrieval practiced categories (*crp-*) and the corresponding control items (*c-*) were tested first in order to avoid confounding Output Interference effects from the practiced items. The test list was arranged compiling six blocks: three blocks with exclusively *crp-* and *c-* items and the other three blocks with the practiced items (*crp+*) and their counterparts (*c+*). For each block, one exemplar from each category was drawn randomly. Order of the three first and the three last blocks as well as of the items within each block was random. At the beginning of the test, three of the six buffer items were tested in order to familiarize participants to the procedure.

The recognition test followed the procedure employed in Experiments 1b, 2b, and 3b. All exemplars from the study list interspersed with lures were presented. Underneath each item, in the lower third of the screen, a schematic rating scale was displayed. Participants rated their confidence of an item having been previously studied (old) or not (new) on a 6-point scale (1 = *definitely old*, 6 = *definitely new*). Responses were typed in by the participants at their own pace, i.e., the next item did not appear on the screen until the subject had rated the presently displayed exemplar. Data were logged automatically by the computer. Order of the recognition list was blocked randomized with two restrictions: old and new items were presented at most three times in a row; the first half of the list contained unpracticed items of practiced categories (*crp-*), their control counterparts (*c-*), and corresponding lures to eliminate Output Interference effects. Twelve blocks were compiled: six blocks consisting of *crp-* items, *c-* items, *crp-* lures, and *c-* lures constituting the first half of the test; and six blocks containing *crp+* items, *c+* items, *crp+* lures, and *c+* lures that were presented in the second part of the recognition test. For each block, one exemplar of each category was drawn and arrayed pseudo-randomly considering the above-mentioned restrictions. The six blocks within one test half were randomly drawn. At the beginning of the recognition list, three buffer items were presented.

Statistical Analysis. Statistical analysis of the cued recall test were identical to the ones used in Experiments 1a, 2a, and 3a. Furthermore, the same statistical analysis conducted in Experiments 1b, 2b, and 3b were used to analyze item recognition data.

Results

Practice phase. During practice, participants successfully retrieved 63.4% ($SD = 0.15$) of the items on the first practice cycle and 64.7% ($SD = 0.15$) of the items in total.

Recall test. Figures 11a and 11b show percentages of correctly recalled practiced (*crp+*) and unpracticed (*crp-*) items, together with their corresponding control (*c+, c-*) items. Regarding the beneficial effect of retrieval practice, participants, on average, recalled 64.2% ($SD = 0.17$) of the *crp+* items and 50.3% ($SD = 0.17$) of the *c+* items. Recall levels differed significantly, $t(47) = 4.794$, $p < .001$, $d = 0.808$, indicating that practice was successful. Intrusion rates were .07 ($SD = 0.07$) for the *crp+* items and .06 ($SD = 0.08$) for the *c+* items, and were not significantly different, $t(47) < 1$.

Regarding the detrimental effect of retrieval practice, on average, 57.6% ($SD = 0.17$) of the *crp-* items and 64.9% ($SD = 0.16$) of the *c-* items were recalled. Recall performance was reliably reduced for the *crp-* items, $t(47) = 3.157$, $p = .003$, $d = 0.450$, indicating the presence of RIF. Intrusion rates were .07 ($SD = 0.07$) for the *crp-* items and .08 ($SD = 0.08$) for the *c-* items and did not differ reliably, $t(47) < 1$.

Recognition test: ANOVA of corrected hits. In Table 7, mean false alarm rates and corrected hit rates are displayed as a function of the five response criteria and the four item types. In the first step, ANOVAs were conducted to analyze for the three most conservative ("old") response criteria whether corrected hits varied with item type. Regarding the beneficial effects of retrieval practice on the practiced (*crp+*) items relative to their controls (*c+*), a 2 × 3 ANOVA with the within-participants factors of ITEM TYPE (*crp+*, *c+*) and RESPONSE CRITERION ("1", "2", "3") showed a main effect of ITEM TYPE, $F(1, 47) = 17.723$, $MSE = 0.056$, $p < .001$, $\eta^2 = 0.274$, with higher corrected

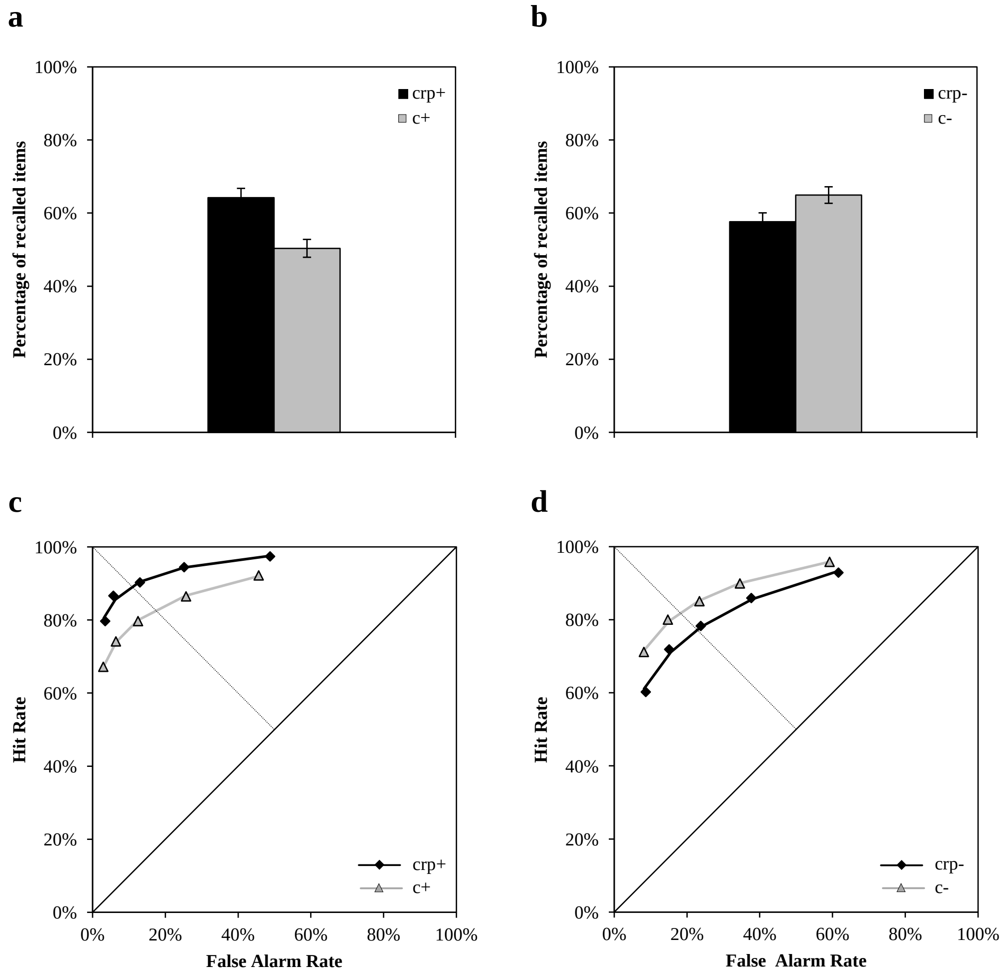

Figure 11. Results of Experiment 4. Mean recall rates and item recognition Receiver Operating Characteristics (ROCs) depicting cumulative hit and false alarm rates as a function of item type. (a) Recall percentages for competitively retrieval practiced items (crp+) and control items (c+). Error bars represent standard errors. (b) Recall percentages for unpracticed items of competitively retrieval practiced categories (crp-) and control items (c-). Error bars represent standard errors. (c) ROCs for competitively retrieval practiced items (crp+) and control items (c+). Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model. (d) ROCs for unpracticed items of competitively retrieval practiced categories (crp-) and control items (c-). Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model.

hits for the practiced than the control items, indicating that practice was successful. The effect did not vary with response criterion, $F(2, 94) = 1.726$, $MSE = 0.003$, $p = .184$, $\eta^2 = 0.035$.

Regarding the detrimental effects of retrieval practice on the unpracticed (*crp-*) items relative to their controls (*c-*), a 2 × 3 ANOVA with the factors of ITEM TYPE (*crp-*, *c-*) and RESPONSE CRITERION ("1", "2", "3") showed a main effect of ITEM TYPE, $F(1, 47) = 12.984$, $MSE = 0.045$, $p = .001$, $\eta^2 = 0.216$, with lower corrected hits for *crp-* items, but no interaction between the two factors, $F(2, 94) = 1.498$, $MSE = 0.008$, $p = .229$, $\eta^2 = 0.031$. These results indicate that retrieval practice induced RIF in item recognition.

Recognition test: analysis of hit and false alarm rates using the unequal-variance signal detection model. In the second step, the unequal-variance signal detection model was employed to analyze the data, which takes the curvilinear and asymmetric form of the ROC into account. Figure 11c and Figure 11d depict the ROCs for the practiced items, the unpracticed items, and the respective control items, as well as the fit of the unequal-variance signal detection model to the data of each single condition. Table 8 shows the statistics of goodness-of-fit and maximum-likelihood estimates of the model's parameters d_a and σ for practiced and unpracticed items and their control counterparts.

The model fit the recognition data of the four types of items well, all $\chi^2 s(3) < 1.408$, all $ps > .703$. Retrieval practiced (*crp+*) items showed enhanced d_a relative to the control (*c+*) items, $\chi^2(1) = 3.904$, $p = .048$, indicating that practice was successful. Relative to the *c-* items, retrieval practice reduced d_a for the unpracticed (*crp-*) items, $\chi^2(1) = 5.374$, $p = .020$, indicating the presence of RIF.

For both the practiced items and their controls, and the unpracticed items and their controls, the variance of the old items' distribution, as estimated by parameter σ, did not vary significantly across item type, both $\chi^2 s(1) < 0.504$, $ps > .477$, but was larger than 1.0, both $\chi^2 s(1) > 27.905$, $ps < .001$,

Table 7 False alarm rates and corrected hit rates for Experiment 4a

Item type		Response criteria				
		"1"	"2"	"3"	"4"	"5"
crp+	False alarms	.035	.057	.130	.252	.488
	Corrected hits	.762	.809	.773	.693	.486
c+	False alarms	.029	.064	.125	.257	.457
	Corrected hits	.643	.677	.672	.608	.465
crp-	False alarms	.087	.151	.238	.377	.616
	Corrected hits	.515	.568	.545	.482	.313
c-	False alarms	.082	.148	.234	.345	.592
	Corrected hits	.630	.652	.617	.554	.366

Note. False alarm and corrected hit rates are shown as a function of item type and response criterion. crp+ = retrieval practiced items; c+ = unpracticed items from unpracticed categories; crp- = unpracticed items from retrieval practiced categories; c- = unpracticed items from unpracticed categories. “1” reflects the strictest response criterion, i.e., definitely old, and each subsequent number (“2”, “3”, etc.) reflects a more and more relaxed criterion. Corrected hits = hits false alarms.

indicating that the model’s assumption of unequal variances for old and new items improved the description of the data significantly. The placement of the five confidence criteria did not vary across item type, $\chi^2 s(5) < 2.244$, $ps > .814$.[8]

[8]Half of the participants in this experiment started testing with the cued recall test and the other half with item recognition. When restricting analyses to the data of participants’ first memory test, exactly the same pattern of results arose as reported above, indicating that testing order did not influence the results. The same held true for the results of Experiments 5 and 6 below.

Table 8 Unequal-variance signal detection model for Experiment 4a

	Parameter estimates		Goodness of fit		
Item type	d_a	σ	χ^2	*df*	*p*
crp+	3.29*	1.65	1.41	3	.704
c+	2.70	1.84	0.08	3	.994
crp-	1.78*	1.38	0.84	3	.840
c-	2.20	1.40	0.15	3	.986

Note. crp+ = competitively retrieval practiced items; c+ = unpracticed items from unpracticed categories; crp- = unpracticed items from competitively retrieval practiced categories; c- = unpracticed items from unpracticed categories. d_a = general memory strength; σ = variance of the target distribution.
* Significant deviations from control performance ($p < .05$).

Discussion

The results of the present experiment provide a further demonstration of the effects of selective retrieval on both cued recall and item recognition. For both types of tests, enhancement of practiced items (*crp+*) attended by impairment of unpracticed items (*crp-*) arose (see e.g., Murayama et al., 2014). The standard finding of RIF in recall is compatible with all accounts of RIF including the context change account, as context manipulations have been found to affect recall performance (e.g., Bower, 1981; Godden & Baddeley, 1975; Sahakyan & Kelley, 2002; S. M. Smith, 1979). Whether the finding of RIF in item recognition is also compatible with the context change account, remains a matter of debate. Previous results regarding context effects on recognition have been mixed (e.g., Bodner & Lindsay, 2003; Bodner & Richardson-Champion, 2007; Fernandez & Glenberg, 1985; Sego et al., 2006) and thus further research is warranted that examines whether context effects can be held accountable for RIF in item recognition. Experiments 5 and 6 were designed to investigate this possibility.

4.2 Experiment 5: The effects of restudy preceded by imagination on cued recall and item recognition

Jonker et al. (2013) provided first evidence that a context change task prior to plain restudy can have similar effects on recall as retrieval practice does. Their experimental design (Jonker et al., 2013, Experiment 2b) included a study phase, in which participants studied a set of category-exemplar pairs, a context change phase, in which participants engaged in a frequently used context change task (imagination of the parents' home), a practice phase, in which participants restudied a subset of items while presented with the intact category-exemplar pair, a distractor phase, and a test phase, in which participants were asked to recall the corresponding exemplar when cued with the category and the exemplar's initial letter. In line with the assumptions by the context change account, restudy preceded by context change induced benefits and costs typical of retrieval practice in cued recall.

The present experiment was designed along Jonker et al.'s (2013) study with the only exception being test format of the final test. Again participants studied category-item pairs, then completed imagination tasks before restudying a subset of items. After a distractor task, all studied items were either tested using a cued recall test or a recognition test (see Experiment 4). Considering the robust findings of RIF in cued recall and in recognition tests (e.g., Hicks & Starns, 2004; see also Murayama et al., 2014, and the present Experiment 4) and the equivalence proposal entailed by the context change account, the context change account would predict that restudy preceded by a context change task should also induce forgetting in both tests. However, considering the mixed results of context change on item recognition, restudy preceded by context change might not induce forgetting in recognition in the Retrieval-practice Paradigm. Such a result, together with the finding that retrieval induces forgetting in recognition, would be incompatible with the idea that context-dependent forgetting and RIF are mediated by the same mechanisms.

Methods

Participants. Further 48 students of Regensburg University participated in the experiment ($M = 21.44$ years, $range = 18 - 29$ years, 37 female). All subjects spoke German as native language and received money in exchange for participation.

Materials. The same material as in Experiment 4 was employed.

Design. The experiment had the same 3×2 design as Experiment 4 with the within-subjects factors of ITEM TYPE (practiced, unpracticed, control) and TEST TYPE (cued recall, recognition). The only differences between the two experiments were the nature of the intermediate task and type of practice: In the intermediate phase, participants were engaged in two successive imagination trials rather than participating in counting and calculation tasks, and in the practice phase they restudied a subset of the studied items rather than retrieving these items (see Figure 10b). Restudied items are denoted *re+* items and unpracticed items of restudied categories are denoted *re-* items. Respective control items are again denoted *c+* and *c-* items. Lures belonging to restudied categories are denoted *re+ lures* and *re- lures*.

Procedure. Study phase, distractor task, and test phase did not differ from Experiment 4. In the intermediate phase, a context change task was administered. Participants were instructed to imagine a scenario as vividly as possible and to write it down within 2 min. Four imagination tasks were employed with two tasks in each experimental block (being in the parents' house; recalling a happy childhood event; winning 10 million Euro in the lottery; being able to perform magic; see Delaney et al., 2010; Sahakayn & Kelley, 2002). Participants completed a block's two imagination tasks consecutively. Following this intermediate phase, participants practiced half of the exemplars from half of the categories by extra study. The complete category-exemplar pair was re-exposed on the computer screen for 4 s (ISI = 500 ms; e.g., HORNET - *termite*). Like in Jonker et al. (2013), participants were asked to read the pairs out loud and to restudy them as thoroughly as possible. Order of presentation was blocked randomized. The twelve pairs were practiced in two consecutive cycles. The final test, cued recall or item recognition, followed after the same 4 min distractor task as was used in Experiment 4.

Statistical analysis. Statistical analysis of the data was analogous to Experiment 4.

Results

Recall test. Percentages of correctly recalled practiced (*re+*) and unpracticed (*re-*) items, together with their corresponding control (*c+, c-*) items, are displayed in Figures 12a and 12b. Regarding the beneficial effect of practice, mean recall rates mounted up to 76.0% (SD = 0.13) and 49.5% (SD = 0.16) for the *re+* and *c+* items. The numerical difference was reliable, $t(47) = 11.434$, $p < .001$, $d = 1.847$, suggesting that practice improved recall. Intrusion rates were .04 (SD = 0.06) for the *re+* items and .07 (SD = 0.10) for the *c+* items, and did not differ significantly, $t(47) = 1.785$, $p = .081$.

Regarding the detrimental effect of practice, participants recalled 57.8% (SD = 0.15) of the *re-* items and 67.7% (SD = 0.13) of the *c-* items, $t(47) = 4.189$, $p < .001$, $d = 0.696$, showing significant recall impairment for the *re-* items and thus RIF-like forgetting. Intrusion rates were .07 (SD = 0.09) for the *re-* items and .09 (SD = 0.09) for the *c-* items and did not vary significantly,

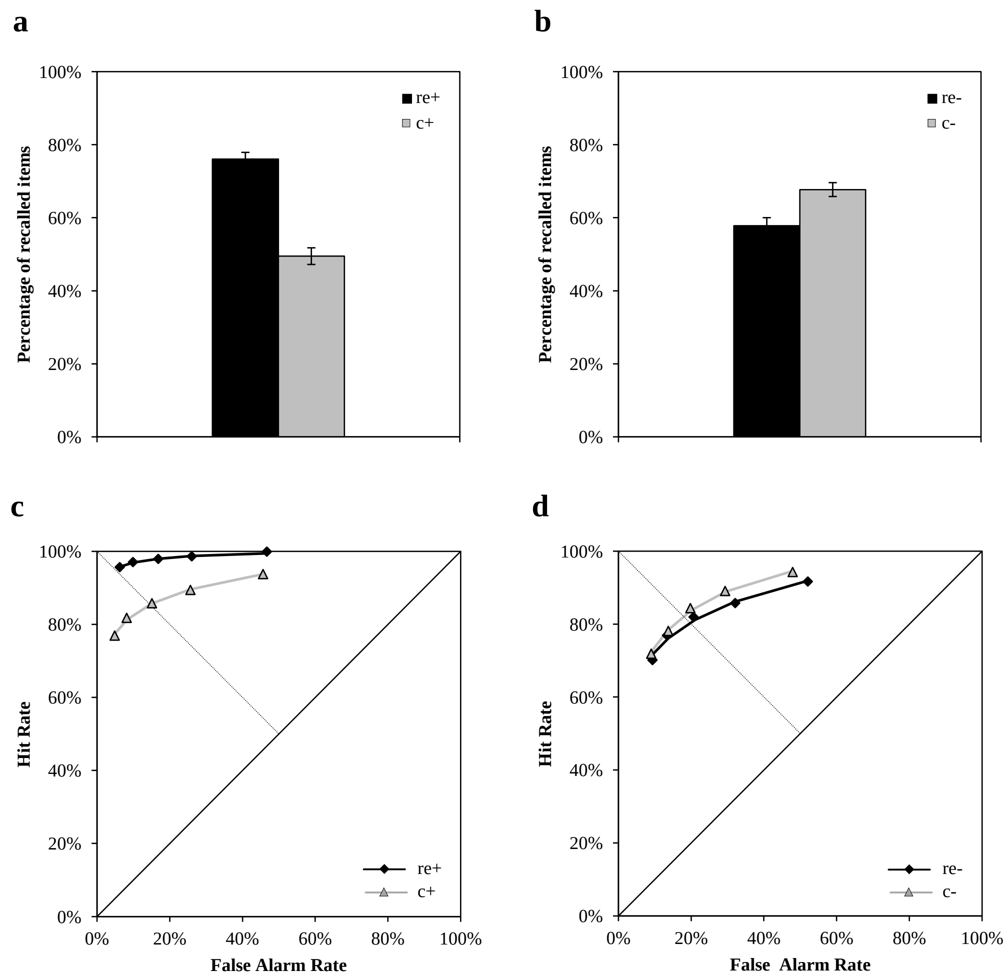

Figure 12. Results of Experiment 5. Mean recall rates and item recognition Receiver Operating Characteristics (ROCs) depicting cumulative hit and false alarm rates as a function of item type. (a) Recall percentages for re-exposed items (re+) and control items (c+). Error bars represent standard errors. (b) Recall percentages for unpracticed items of re-exposed categories (re-) and control items (c-). Error bars represent standard errors. (c) ROCs for re-exposed items (re+) and control items (c+). Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model. (d) ROCs for unpracticed items of re-exposed categories (re-) and control items (c-). Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model.

Table 9 False alarm rates and corrected hit rates for Experiment 4b

Item type		Response criteria				
		"1"	"2"	"3"	"4"	"5"
re+	False alarms	.063	.099	.168	.261	.467
	Corrected hits	.894	.872	.811	.726	.528
c+	False alarms	.049	.082	.151	.257	.457
	Corrected hits	.721	.736	.707	.637	.481
re-	False alarms	.094	.134	.207	.321	.521
	Corrected hits	.608	.635	.613	.537	.396
c-	False alarms	.090	.137	.198	.293	.479
	Corrected hits	.628	.644	.646	.597	.464

Note. False alarm and corrected hit rates are shown as a function of item type and response criterion. re+ = re-exposed items; c+ = unpracticed items from unpracticed categories; re- = unpracticed items from re-exposed categories; c- = unpracticed items from unpracticed categories. "1" reflects the strictest response criterion, i.e., definitely old, and each subsequent number ("2", "3", etc.) reflects a more and more relaxed criterion. Corrected hits = hits - false alarms.

$t(47) < 1$.

Recognition test: ANOVA of corrected hits. Table 9 depicts mean false alarm rates and mean corrected hit rates as a function of response criterion and item type. Regarding the beneficial effects of practice on corrected hits, a 2 × 3 ANOVA with the within-participants factors of ITEM TYPE (*re+*, *c+*) and RESPONSE CRITERION ("1", "2", "3") revealed a main effect of ITEM TYPE, $F(1, 47) = 36.786$, $MSE = 0.037$, $p < .001$, $\eta^2 = 0.439$, which was qualified by an interaction with the factor of response criterion, $F(2, 94) = 7.600$, $MSE = 0.004$, $p = .001$, $\eta^2 = 0.139$. However, corrected hit rates

Table 10 Unequal-variance signal detection model for Experiment 4b

	Parameter estimates		Goodness of fit		
Item type	d_a	σ	χ^2	*df*	*p*
re+	4.60*	1.78	0.11	3	.990
c+	3.15	1.98	0.16	3	.984
re-	2.28	1.67	1.09	3	.781
c-	2.10	1.28	0.31	3	.958

Note. re+ = re-exposed items; c+ = unpracticed items from unpracticed categories; re- = unpracticed items from re-exposed categories; c- = unpracticed items from unpracticed categories. d_a = general memory strength; σ = variance of the target distribution.
* Significant deviations from control performance ($p < .05$).

for *re+* items exceeded corrected hit rates for *c+* items for all three response criteria, all $ts(47) > 4.252$, all $ps < .001$, all $ds > 0.445$, indicating that practice was successful.

Regarding the detrimental effects of practice, a 2 × 3 ANOVA with the factors of ITEM TYPE (*p*-, *c*-) and RESPONSE CRITERION ("1", "2", "3") showed no main effect of ITEM TYPE, $F(1, 47) < 1$, and no interaction between the two factors, $F(2, 94) < 1$, indicating that restudy preceded by context change did not impair recognition of the unpracticed items.

Recognition test: analysis of hit and false alarm rates using the unequal-variance signal detection model. Figure 12c and Figure 12d depict the ROCs for the practiced items, the unpracticed items, and the respective control items, as well as the fit of the unequal-variance signal detection model to the data of each single condition. Table 10 shows the statistics of goodness-of-fit and maximum-likelihood estimates of the model's parameters

d_a and σ for practiced and unpracticed items and their control counterparts. The unequal-variance signal detection model described the data of the four item types well, all $\chi^2 s(3) < 1.086$, all $ps > .780$. Practiced (*re+*) items showed higher d_a than the control (*c+*) items, $\chi^2(1) = 4.415$, $p = .036$, indicating improved recognition of the practiced items after the restudy trials. Critically, however, restudy preceded by imagination did not impair d_a for unpracticed (*re-*) items relative to the control (*c-*) items, $\chi^2(1) = 0.601$, $p = .438$, again indicating that no RIF-like forgetting arose in item recognition.

Variance σ did not differ significantly between *re+* and *c+* items and between *re-* and *c-* items, $\chi^2 s(1) < 2.827$, $ps > .130$, although, like in Experiment 4, σ was significantly larger than 1.0, $\chi^2 s(1) > 25.398$, $ps < .001$, indicating that the model's assumption of unequal variance for old and new items improved the description of the data significantly. Differences in the placement of the five confidence criteria did not reach significance, $\chi^2 s(5) < 3.041$, $ps > .693$.

Discussion

The present experiment replicated prior results by Jonker et al. (2013), showing that selective restudy preceded by imagination tasks can affect recall similarly to how retrieval practice does (see Experiment 4). Like retrieval, restudy following imagination enhanced recall for practiced items (*re+*) and reduced recall performance for unpracticed items (*re-*). This finding is in line with the context change account of RIF, predicting equivalent effects of retrieval and restudy preceded by context change. Unlike retrieval, however, restudy with prior imagination did not induce impairment of unpracticed items (*re-*) when memory performance was assessed using a recognition test. The diverging effects of retrieval and restudy preceded by context change on recognition are inconsistent with the context change account of RIF suggesting different underlying mechanisms. The results support furthermore the view that context change, even though it may contribute to RIF in recall, does not affect recognition memory in the Retrieval-practice Paradigm. Before arriving at a final conclusion with respect to the role of context change in RIF, Experiment 6 was dedicated to investigating whether the results of Experiment 5 could be replicated when a different context change task is employed prior to practice.

4.3 Experiment 6: The effects of restudy preceded by semantic generation on cued recall and item recognition

Context change inductions vary from environmental changes (e.g., Godden & Baddeley, 1975, 1980; S. M. Smith et al., 1978) to changes in mood (e.g., Bower, 1981; E. Eich, 1995) and mind-wandering (e.g., Delaney et al., 2010; Sahakyan & Kelley, 2002). Recently, it has been suspected that retrieval tasks such as semantic generation can induce an internal context change. In a study by Divis and Benjamin (2014), participants studied multiple lists. Critically, between lists, they either engaged in semantic generation of category exemplars, i.e., participants were instructed to write down as many exemplars of a given category as possible (e.g., SPORTS), or they continued engaging in a distractor task (counting backward). Interpolation of semantic generation tasks improved recall of the final list and impaired recall of the first list mimicking the typical results of context change inductions. The effects of semantic generation on prior and subsequent learning are in line with the notion that semantic retrieval prompts a context change (for related results and interpretations see, Jang & Huber, 2008; Pastötter et al., 2011; Sahakyan & Hendricks, 2012).

The idea of the context change account of RIF was inspired by findings showing semantic retrieval to induce equivalent effects as other well-established context change tasks do. Furthermore, one might assume that semantic retrieval, i.e., semantic generation, and episodic retrieval, i.e., retrieval practice in the Retrieval-practice Paradigm, are more similar with regard to the underlying processes than imagination tasks, like mind-wandering, and retrieval practice might be. Therefore, the goal of Experiment 6 was to explore whether the findings obtained in Experiment 5 would be replicated replacing the imagination tasks by semantic generation tasks or whether the effects of restudy preceded by semantic generation on cued recall and item recognition of unpracticed items would mimic the effects of retrieval practice. Again, participants studied a list of category-exemplar pairs. In the intermediate phase, participants generated exemplars to different categories that were not part of the study list. Next, a subset of items from the study list were re-exposed, and participants were asked to study them once more. After a distractor task, the final test was administered, either in form of a cued recall test or an item recognition test. In line with the context change account of RIF, semantic generation should induce analogous effects as retrieval practice did (see Experiment 4), i.e., it should enhance performance for practiced items and reduce performance for unpracticed items in both types of test. If the processes underlying semantic retrieval are more similar to the ones underlying imagination tasks, then semantic generation prior to restudy should induce enhancement of practiced items and forgetting of unpracticed items when a cued recall test is employed, but should not induce forgetting of unpracticed items in the recognition test (see Experiment 5).

Methods

Participants. Another 48 students were recruited at Regensburg University to participate in the experiment ($M = 23.25$ years, $range = 20 - 29$ years, 39 female). All spoke German as native language. Participation was rewarded monetarily.

Materials. Materials were identical to Experiments 4 and 5.

Design. A 3 $\times$ 2 design was employed varying ITEM TYPE (practiced, unpracticed, control) and TEST TYPE (cued recall, recognition) within subjects. The only difference between the present experiment and Experiment 5 was that, in the intermediate phase, participants were asked to generate as many exemplars from semantic categories as possible rather than engaging in an imagination task (see Figure 10c). Analogous to Experiment 5, *re+* items represent restudied items, *re-* items represent unpracticed items of restudied categories, and *c+* and *c-* items represent corresponding control items. Again, *re+ lures* and *re- lures* label foils that are members of restudied categories.

Procedure. Study phase, practice phase, distractor task, and test phase were identical to Experiment 5. However, the intermediate phase consisted of semantic generation tasks. Participants were instructed to think of as many exemplars from a particular category (COLORS, CANDY, FIRST NAMES, MEANS OF TRANSPORT) as possible and write them down within 2 min. None of the to-be-studied items or lures belonged to one of the four semantic categories. Participants completed two out of the four semantic retrieval tasks consecutively, the remaining two tasks were presented in the second block of the experiment. Like in Experiment 5, participants then engaged in extra study of a subset of the word pairs. The complete category-exemplar pair was re-exposed on the computer screen for 4 s (ISI = 500 ms; e.g., INSECT - *termite*). Again, we asked participants to read the pairs out loud and to restudy them for a later test. The twelve pairs were practiced in two consecutive cycles in blocked randomized order. The final test, cued recall or item recognition, followed after the same 4 min distractor task as was used in Experiments 4 and 5.

Statistical analysis. The same statistical analyses as in Experiments 4 and 5 were employed.

Results

Recall test. Figures 13a and 13b show percentages of correctly recalled practiced (*re+*) and unpracticed (*re-*) items together with their corresponding control (*c+*, *c-*) items. Regarding the beneficial effect of practice, participants recalled on average 75.5% (SD = 0.17) of the *re+* items and 46.9% (SD = 0.20) of the *c+* items. Recall differed significantly between item types, $t(47) = 8.656$, $p < .001$, $d = 1.547$, indicating that practice was successful. Intrusion rates were .04 (SD = 0.06) for the *re+* items and .06 (SD = 0.07) for the *c+* items, and did not vary significantly between item types, $t(47) = 1.295$, $p = .202$.

Regarding the detrimental effect of practice, recall rates for *re-* items and *c-* items reached 54.5% (SD = 0.18) and 62.3% (SD = 0.15), respectively. The numerical difference was reliable, $t(47) = 3.986$, $p < .001$, $d = 0.462$, suggesting that selective restudy preceded by semantic generation induced RIF-like forgetting. Intrusion rates were .06 (SD = 0.08) for the *re-* items and .08 (SD = 0.07) for the *c-* items, but the difference was not significant, $t(47) = 1.400$, $p = .168$.

Recognition test: ANOVA of corrected hits. Table 11 shows mean false alarm rates and mean corrected hit rates for the five response criteria and the four item types. Regarding the beneficial effect of restudy supplemented with prior semantic generation on corrected hits, a 2 × 3 ANOVA was conducted with the within-participants factors of ITEM TYPE (*re+*, *c+*) and RESPONSE CRITERION ("1", "2", "3"). A main effect of ITEM TYPE arose, $F(1, 47) = 71.723$, $MSE = 0.044$, $p < .001$, $\eta^2 = 0.604$, suggesting that practice was successful, and an interaction of the two factors, $F(2, 94) = 9.097$, $MSE = 0.006$, $p < .001$, $\eta^2 = 0.162$. Although the size of the beneficial effect thus varied with the particular response criterion, the effect was present for each single criterion, all $ts(47) > 6.158$, all $ps < .001$, all $ds > 1.079$.

Regarding the detrimental effect of restudy supplemented with prior semantic generation, a 2 × 3 ANOVA with the factors of ITEM TYPE (*re-*,

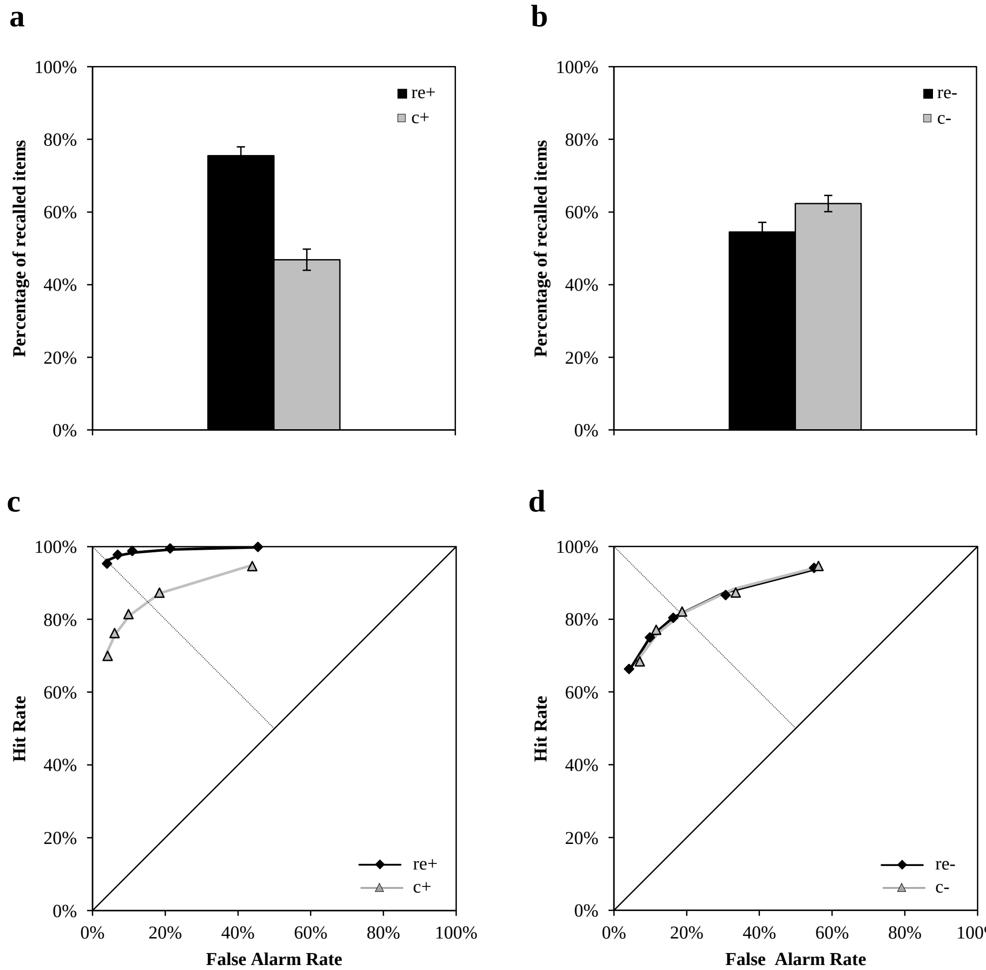

Figure 13. Results of Experiment 6. Mean recall rates and item recognition Receiver Operating Characteristics (ROCs) depicting cumulative hit and false alarm rates as a function of item type. (a) Recall percentages for re-exposed items (re+) and control items (c+). Error bars represent standard errors. (b) Recall percentages for unpracticed items of re-exposed categories (re-) and control items (c-). Error bars represent standard errors. (c) ROCs for re-exposed items (re+) and control items (c+). Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model. (d) ROCs for unpracticed items of re-exposed categories (crp-) and control items (c-). Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model.

Table 11 False alarm rates and corrected hit rates for Experiment 4c

Item type		Response criteria "1"	"2"	"3"	"4"	"5"
re+	False alarms	.040	.069	.109	.214	.455
	Corrected hits	.913	.908	.879	.781	.540
c+	False alarms	.042	.061	.099	.184	.439
	Corrected hits	.658	.701	.715	.689	.507
re-	False alarms	.042	.099	.163	.307	.550
	Corrected hits	.622	.651	.641	.559	.391
c-	False alarms	.071	.116	.187	.335	.563
	Corrected hits	.613	.655	.634	.538	.384

Note. False alarm and corrected hit rates are shown as a function of item type and response criterion. re+ = re-exposed items; c+ = unpracticed items from unpracticed categories; re- = unpracticed items from re-exposed categories; c- = unpracticed items from unpracticed categories. "1" reflects the strictest response criterion, i.e., definitely old, and each subsequent number ("2", "3", etc.) reflects a more and more relaxed criterion. Corrected hits = hits false alarms.

c-) and RESPONSE CRITERION ("1", "2", "3") showed no main effect of ITEM TYPE, $F(1, 47) < 1$, and no interaction between the two factors, $F(2, 94) < 1$, indicating that restudy preceded by semantic generation did not affect recognition of the unpracticed items.

Recognition test: analysis of hit and false alarm rates using the unequal-variance signal detection model. Separately for each item type, Figure 13c and Figure 13d display the ROCs and the fit of the unequal-variance signal detection model, while Table 12 shows the statistics of goodness-of-fit and maximum-likelihood estimates of the model's parameters d_a and σ. The

Table 12 Unequal-variance signal detection model for Experiment 4c

	Parameter estimates		Goodness of fit		
Item type	d_a	σ	χ^2	*df*	*p*
p+	4.46*	1.52	4.18	3	.243
c+	2.58	1.48	0.56	3	.906
p-	2.40	1.66	0.47	3	.926
c-	2.25	1.52	1.39	3	.707

Note. re+ = re-exposed items; c+ = unpracticed items from unpracticed categories; re- = unpracticed items from re-exposed categories; c- = unpracticed items from unpracticed categories. d_a = general memory strength; σ = variance of the target distribution.
* Significant deviations from control performance ($p < .05$).

unequal-variance signal detection model provided a good fit for the data of the four item types, all $\chi^2 s(3) < 4.178$, all $ps > .242$. Regarding the beneficial effect of practice, d_a for the practiced (*re+*) items exceeded that of the control (*c+*) items, $\chi^2(1) = 16.828$, $p < .001$, suggesting that practice was successful. Regarding the detrimental effect of practice, d_a did not differ between *re-* items and *c-* items, $\chi^2(1) = 0.541$, $p = .462$, indicating that no RIF-like forgetting arose in recognition memory.

Like in Experiments 4 and 5, σ did not vary with item type, $\chi^2 s(1) < 0.412$, $ps > .520$, and was larger than 1.0, $\chi^2 s(1) > 25.531$, $ps < .001$. Again, the placement of the five confidence criteria did not differ across item type, $\chi^2 s(5) < 7.593$, $ps > .180$.

Discussion

The effects of restudy preceded by semantic generation mirrored the effects of restudy preceded by imagination observed in Experiment 5. Like imagination, semantic generation when administered before selective restudy increased recall rates for practiced items (*re+*) and decreased recall rates for related unpracticed items (*re-*) compared to the baseline. When a recognition test was employed, on the contrary, recognition performance for practiced items (*re+*) was enhanced by restudy after semantic generation, but recognition performance for related unpracticed items (*re-*) was unaffected. These findings suggest that semantic generation and imagination tasks (see Experiment 5) trigger similar processes and may likely be mediated by an internal context change. Altogether, the results of Experiments 4-6 indicate that the effects of retrieval practice and the effects of imagination or semantic generation prior to restudy are not equivalent since selective retrieval impairs both recall and recognition whereas selective restudy with preceding context change tasks only impairs recall but leaves recognition unaffected. Consequently, it seems very unlikely that the effects of retrieval practice observed in Experiment 4 underlie the same mechanism, i.e., internal context change.

Taken together, the findings of Experiments 4-6 represent replications and extensions to prior research on Retrieval-induced Forgetting. In line with previous results (e.g., M. C. Anderson et al., 1994; Grundgeiger, 2014; Hicks & Starns, 2004; see also Murayama et al., 2014), Experiment 4 showed once more that selective retrieval practice can facilitate memory performance of practiced items and reduce memory performance of unpracticed items in both recall and recognition tests. Furthermore, Jonker et al.'s (2013) findings were replicated showing that imagination tasks prior to restudy cycles induced enhancement and forgetting in recall just like retrieval practice does.

Going beyond prior work, imagination before restudy, unlike retrieval practice, did not affect recognition memory of unpracticed items. This suggests that restudy preceded by context change may be sufficient to induce RIF-like forgetting when cued recall tests are administered, but that retrieval is necessary when recognition memory is assessed. Furthermore, the results suggest that the effects of imagination tasks generalize to other tasks that are considered a means to induce a context change, like semantic generation: Again, while retrieval practice induced forgetting in both recall and recognition, semantic generation affected recall performance of unpracticed items, but recognition performance remained unaffected. Thus, the results are incompatible with the context change account stating that RIF is equivalent to context-dependent forgetting and that it arises due to a mismatch of study and test context for the unpracticed items. If RIF were equivalent to context-dependent forgetting, then forms of context change other than retrieval when paired with re-exposure of to-be-practiced items should induce forgetting whenever retrieval practice does.

Chapter 5

General Discussion

The primary goal of the present thesis was to examine the equivalence hypotheses proposed by the blocking and the context change accounts of RIF. To address this issue, Experiments 1-3 investigated how re-exposure formats enhancing cue-item associations, namely, noncompetitive retrieval practice, restudy augmented with pleasantness ratings, and restudy augmented with visualization of item features, affected cued recall and item recognition memory of practiced and unpracticed items and whether these effects mimicked those of (competitive) retrieval practice. Experiments 4-6 explored the effects of retrieval practice and the effects of restudy preceded by context change manipulations, i.e., imagination and semantic generation, on cued recall and item recognition of practiced and unpracticed items.

The present experiments consistently showed that, when cued recall tests were employed, not only (competitive) retrieval practice but also re-exposure formats supposed to enhance cue-item associations and restudy preceded by context change tasks facilitated recall of practiced items and reduced recall of unpracticed items. When item recognition tests were employed, all practice formats significantly enhanced performance of practiced items. Crucially, only retrieval practice but none of the other practice formats impaired item recognition performance of the unpracticed items.

The results of these experiments are notable in several respects. The recall findings replicate the standard RIF effect (e.g., M. C. Anderson et al., 1994) as well as earlier studies showing that the effects of re-exposure formats that strengthen cue-item associations and of restudy preceded by context change mirror the effects of retrieval practice (Jonker et al., 2013; Raaijmakers & Jakab, 2012; Saunders et al., 2009; Verde, 2013). In addition, using semantic retrieval as a context change manipulation extends Jonker et al.'s (2013) findings suggesting equivalent effects of restudy preceded by imagination and semantic generation on cued recall and item recognition in the adapted Retrieval-practice Paradigm. With regard to the recognition results, further demonstrations of RIF in item recognition were provided across four experiments complementing the already well-established finding (see e.g., Aslan & Bäuml, 2011; Dobler & Bäuml, 2013; Gómez-Ariza et al., 2005; Román et al., 2009; Spitzer & Bäuml, 2007; Starns & Hicks, 2004; Veling & van Knippenberg, 2004). Critically, the present experiments replicated Grundgeiger's (2014) pattern of data: While cued recall was affected equivalently by competitive and noncompetitive retrieval practice, showing facilitation of practiced items and impairment of unpracticed items after both practice formats, item recognition showed diverging results. Competitive retrieval practice induced enhancement of practiced items and forgetting of unpracticed items in item recognition, noncompetitive retrieval practice induced enhancement of practiced items, but it did not induce forgetting of unpracticed items. Extending these findings, the present study provided evidence that neither re-exposure formats that strengthened cue-item associations nor restudy preceded by context change tasks induced forgetting when recognition memory was assessed. This suggests that selectively retrieval practice leads to impaired recognition performance, yet other practice formats, that reliably induced RIF-like forgetting in cued recall, do not.

In the following sections, the consistency of the present results with the accounts introduced above will be discussed.

5.1 Evaluation of the single mechanisms of Retrieval-induced Forgetting

Blocking account

According to the blocking account, RIF is the product of significant strengthening of cue-item associations during practice which causes blocking and thus forgetting at test. Therefore, any re-exposure format that allows for adequate enhancement of the associations should lead to equivalent memory impairment of unpracticed items as retrieval practice, i.e., whenever RIF arises following retrieval practice, so should RIF-like forgetting following the re-exposure formats. Experiments 1-3 were designed to directly test the equivalence assumption of the blocking account and thus to shed light on its validity. However, a comprehensive account of RIF should be able to explain the entirety of RIF findings. Therefore, the compatibility of the blocking account with the results of Experiments 4-6 as well as with relevant findings in the literature will be discussed in turn.

With respect to the recall findings of Experiments 1a, 2a, and 3a, the present results replicated those of earlier studies (Raaijmakers & Jakab, 2012; Saunders et al., 2009; Verde, 2013) showing that noncompetitive retrieval practice, restudy when augmented with pleasantness judgments, and restudy when augmented with visualization of item features enhance recall of practiced items and reduce recall of unpracticed items like competitive retrieval practice does. These findings indicate equivalent effects of competitive retrieval practice and re-exposure formats that strengthen cue-item associations, and are therefore in line with predictions derived from the blocking account. When regarding the recognition findings of Experiments 1b, 2b, and 3b, however, discrepant effects of competitive retrieval and the applied re-exposure formats were observed. Employing the same materials and procedures as in the recall experiments, competitive retrieval practice improved recognition of the practiced items and reduced recognition of the unpracticed items. All three re-exposure formats enhanced recognition memory of the practiced items reliably, reduction of recognition memory of the unpracticed items yet failed to appear. The diverging effects of retrieval practice and the employed re-exposure formats conflict with the equivalence hypothesis proposed by the blocking account of RIF. They rather indicate that (competitive) retrieval is necessary to observe RIF whereas re-exposure is not sufficient, at least when item recognition tests are employed. Thus, RIF cannot be reduced to strength-based forgetting challenging the blocking account of RIF. Moreover, strength-dependent blocking does not seem to reliably affect recognition memory in the Retrieval-practice Paradigm. Altogether, the effects of particular re-exposure formats and retrieval practice may be equivalent in recall tests, giving the impression that the underlying mechanisms are identical, the effects on recognition memory, however, differ, suggesting a dismissal of blocking as a single underlying mechanism.

Following Jonker et al. (2013), Experiments 5 and 6 investigated the effects of context change tasks prior to restudy on cued recall and item recognition as a comparison to the effects of retrieval practice (Experiment 4). It might be worth exploring the possibility that these findings may actually result from blocking. In fact, it has been shown that practice after context change may be particularly beneficial as encoding following context change is more effective compared to following no context change (Pastötter & Bäuml, 2010; Pastötter, Bäuml, & Hanslmayr, 2008; Sahakyan & Delaney, 2003).[9] Hence, in contrast to plain restudy, restudy preceded by context change may increase the practiced items' blocking potential severely, like noncompetitive retrieval practice (Raaijmakers & Jakab, 2012), restudy augmented with pleasantness ratings (Verde, 2013), or restudy augmented with visualization (Saunders et al., 2009) are supposed to do, and the observed impairment in cued recall may be the result of blocking by the disproportionately strengthened practiced items. As a consequence, the blocking account would predict the effects of restudy preceded by context change tasks to be equivalent to those of restudy formats strengthening cue-item associations, and therefore to be equivalent to the effects of retrieval practice. In line with this interpretation of the context manipulations, the results of Jonker et al. (2013, Experiment 2a) and Experiments 4-6 showed that restudy preceded by imagination and restudy preceded by semantic generation facilitated recall of practiced items and impaired recall of unpracticed items mimicking the effects of retrieval practice.

[9]It should be noted, however, that the evidence on enhanced encoding concerned exclusively study of novel material whereas in the Retrieval-practice Paradigm usually a subset of the originally studied list is practiced (for exceptions see, e.g., Bäuml, 2002; Storm et al., 2006; Verde, 2013, Experiment 6), thus inferences should be drawn with caution.

Recognition performance, however, was differently affected by retrieval practice and restudy preceded by imagination or semantic generation tasks. Even though all three practice methods induced enhancement of the practiced items, only retrieval practice but neither restudy preceded by imagination nor restudy preceded by semantic generation induced forgetting on the item recognition test. Again, blocking cannot account for these dissimilar effects of retrieval practice and enhanced encoding via context change.

Moreover, the blocking account exhibits difficulty in explaining the finding that reinstatement of the study context prior to the final test eliminates RIF. Jonker et al. (2013; Experiment 3) tested the context change account directly in the Retrieval-practice Paradigm. For this, participants studied exemplars with video clips that were distinct for each category and were supposed to serve as context cues. In the retrieval practice phase, the to-be-practiced items were paired with a novel cue. In order to reinstate either the study context or the practice context, the original or the novel cue, respectively, were presented at test of unpracticed items. In line with the context change account, RIF occurred when the novel cue was presented, i.e., the practice context was reinstated and consequently a mismatch of study and test context persisted, but RIF failed to arise when the original cue was presented, i.e., the study context was reinstated causing a match in study and test context (but see Miguez, Mash, Polack, & Miller, 2014). As pointed out above, the recall findings of Experiments 5 and 6 may be interpreted in terms of enhanced encoding following context change and blocking in the final test. A contextual mismatch however may not primarily contribute to RIF. The finding that context reinstatement eliminates RIF is therefore difficult to reconcile with the blocking hypothesis, as blocking by strengthened practiced items should persist whether the original study context is reinstated or not.

Regarding the present results and their implications, objections could be raised that will be addressed in the following passage. According to Raaijmakers and Jakab (2012), retrieval may represent a much more effective way of strengthening cue-item associations than restudy formats do. This suggestion is supported by results in the testing literature, if one concedes that the testing effect may merely be a result of strengthening (see *Bifurcation model*, Kornell, Bjork, & Garcia, 2011; for alternative views see Roediger & Butler, 2011), showing that retrieving information produces long-term benefits for the tested material when compared to restudy (e.g., Roediger & Karpicke, 2006). In turn, this might imply that, in the present study, competitive retrieval practice may have enhanced cue-item associations more than the other practice methods did and hence the blocking potential of the retrieval practiced items may have been more pronounced. Strengthening by the presently employed re-exposure methods may have sufficed to induce forgetting in cued recall, but a higher level of strengthening, like the one induced by competitive retrieval practice, may be required to induce detectable effects of blocking in recognition. However, several arguments challenge this assumption. It has been suggested that strengthening of cue-item associations is attended by improved differentiation of practiced from unpracticed items in item recognition and thus by improved recognition, compensating for the enhanced blocking potential of the practiced items (e.g., Shiffrin & Steyvers, 1997). For this reason, it remains doubtful whether RIF in recognition can represent a result of blocking in the first place. Moreover, the recognition data presented here show a dissociation of enhancement and forgetting, as enhanced recognition was found after re-exposure formats but forgetting was not. This indicates that forgetting is independent of the degree of strengthening during practice as is assumed by the inhibition account and has been documented earlier (e.g., Storm et al., 2006; see also Murayama et al., 2014). These findings challenge the assumption that RIF in the present and in previous studies is the result of higher levels of strengthening.

Context change account

The context change account states that RIF arises due to a context shift between study and practice initiated by the act of retrieval, and the resulting mismatch in context leads to the observed forgetting effect. In consequence, any kind of practice should induce RIF-like forgetting granted that practice is preceded by a context change and the study context is not reinstated prior to the final test. The empirical prediction follows that restudy preceded by context change will induce forgetting whenever retrieval practice does. Analogously to the blocking account, the context change account postulates to explain all RIF findings, so it should be consistent with the results from both sets of experiments as well as with preexisting findings.

Replicating and going beyond Jonker et al.'s (2013) study, the present findings (Experiments 4-6) show not only retrieval practice but also restudy when preceded by imagination or semantic generation to induce enhancement of practiced items and impairment of unpracticed items in cued recall. These equivalent effects of retrieval practice and context change prior to restudy suggest a common underlying mechanism, i.e., a contextual mismatch of study and test context for the unpracticed items, and are consistent with the context change account of RIF. It seems worth noting once more, that, across the three experiments, the same materials and procedures were employed and test format was varied within-subjects. With respect to the item recognition test, retrieval practice and context change manipulations prior to restudy affected recognition memory differently. Retrieval practice again facilitated recognition for practiced items and impaired recognition for unpracticed items replicating the findings of Experiments 1a, 2a, and 3a. Restudy preceded by imagination tasks and restudy preceded by semantic generation tasks, on the other hand, did not induce significant forgetting in item recognition, nevertheless enhancing recognition for practiced items. While retrieval and context change prior to restudy affect recall similarly, their divergent effects on recognition performance challenge the equivalence proposal by the context change account that predicted the effects of restudy and context change to mimic the effects of retrieval. They rather suggest that retrieval practice is necessary for RIF to arise while restudy preceded by context change is not sufficient, at least when item recognition tests are employed. Consequently, the experiments demonstrate that RIF is not merely an instance of context-dependent forgetting, and that context manipulations do not seem to have reliable effects on recognition memory in the Retrieval-practice Paradigm. In sum, the cued recall results suggest equivalent effects of retrieval practice and restudy preceded by context change, suggesting a common underlying mechanism, regarding the disparity in the recognition findings however the context change account of RIF must be dismissed.

Following Raaijmakers and Jakab (2012), Verde (2013), and Saunders et al. (2009), Experiments 1-3 compared the effects of (competitive) retrieval practice to the effects of noncompetitive retrieval practice, restudy augmented with pleasantness ratings, and restudy augmented with visualizations on cued recall and item recognition. It is worth considering the possibility that these earlier findings and the present data result from contextual mismatch between study and test. The context change account assumes that retrieval induces a context change and that the following exposure to the practiced items in the new context is responsible for the failure to remember the unpracticed items at test. Following this line of argumentation, it seems pretty clear that noncompetitive retrieval should also induce context change and thus RIF-like forgetting as it is irrelevant whether retrieval is competitive or not. Whether restudy augmented with pleasantness ratings and restudy augmented with visualizations involve retrieval and trigger context shift is a more contentious issue. However, it seems plausible that when participants judge the pleasantness of an object they retrieve information like the valence, related emotions, and former encounters with the object. Analogously, visualizing the objects requires retrieval of the visual features (for a similar argument, see Saunders et al., 2009) and should, according to the context change account, induce RIF-like forgetting. Consequently, like restudy preceded by imagination or semantic generation, noncompetitive retrieval practice, restudy augmented with pleasantness ratings, and restudy augmented with visualizations should induce a contextual mismatch similar to how retrieval practice is supposed to do and represent thus further variants of restudy attended by context change.

The context change account would therefore predict equivalent effects of the in Experiments 1-3 employed restudy formats and retrieval practice irrespective of the final test method. These predictions are consistent with Raaijmakers and Jakab's (2012), Verde's (2013), and Saunders et al.'s (2009) recall results as well as the present replications (Experiments 1a, 2a, & 3a) demonstrating enhancement of practiced items and impairment of unpracticed items after each restudy formats as well as after competitive retrieval. With respect to the recognition results, the effects of the restudy formats did not imitate the effects of competitive retrieval practice, as only retrieval induced forgetting in the recognition test. Thus, a contextual mismatch cannot account for the disparities between retrieval practice and these variants of re-exposure preceded by a potential context change.

Besides the present findings, further studies have challenged the context change account of RIF. Similar to Jonker et al.'s (2013) design, Buchli, Storm, and Bjork (2015) varied the practice tasks comparing a standard retrieval practice task, a standard restudy task, and restudy preceded by two different context manipulations. Buchli et al. used the near vs. far imagination technique (Delaney et al., 2010), i.e., participants either imagined a 'near' event like vacation in the home country or they imagined a 'far' event like vacation abroad, in order to create different degrees of context change. The rationale was that greater shifts in context should by attended by greater amounts of forgetting. However, recall impairment did not only fail to vary with the amount of context change, but forgetting was completely absent following restudy preceded by context change in a striking sample of more than 750 participants across three experiments. Similarly, Soares, Polack, and Miller (2016) manipulated context between study, practice, and test phase in the Retrieval-practice Paradigm. In two experiments, they varied processing and task demands between study and practice, i.e., participants either studied or generated the items in the study phase, and salient item-specific features, i.e., changing font or color of the presented words, in order to generate contextually matching and mismatching conditions. RIF was not influenced by manipulations of processing or task demands showing significant RIF after both study and generation. Varying item features had in fact different effects on memory performance for unpracticed items, however, the condition in which study and practice context mismatched while practice and test context matched, which reflects the particular premise of the context change account, did not lead to forgetting. Last but not least, Miguez et al. (2014) were not able to replicate Jonker et al.'s finding that context reinstatement prior to the test phase eliminates RIF when using retrieval practice instead of restudy preceded by a context change task. In sum, these findings question the validity of the context change account of RIF.

Despite the inconsistencies of the context change account with previous studies, potential objections against the interpretation of the present data might come to mind. First, the difference between effects of retrieval practice and restudy with prior context change may not be qualitative but quantitative in nature. Assuming that retrieval practice triggers a more extensive shift while study and test context overlap more after imagination or semantic generation and restudy, the present data after retrieval practice and restudy preceded by these context change tasks could still result from context shift. Granting this assumption, the effects of context change after retrieval may arise in both recall and item recognition whereas the effects of context change after imagination or semantic generation tasks show in recall tests, but are too feeble to affect recognition memory. With regard to the results of the recall tests, this assumption seems rather unlikely as the amount of forgetting in recall did not vary significantly across Experiments 4-6[10] and was numerically even

[10]Experiments 4-6 were compared with respect to beneficial and detrimental effects in recall. A 2×3 ANOVA with the within-subjects factor of ITEM TYPE (*p+*, *c+*) and the between-subjects factor of PRACTICE CONDITION (retrieval practice, restudy-plus-imagination, restudy-plus-semantic-generation) showed a significant interaction between the two factors, $F(2, 141) = 7.141$, $MSE = 0.019$, $p = .001$, $\eta^2 = .092$, indicating that the improvement effect differed statistically showing more enhancement after context change and restudy (Experiments 5 & 6) compared to retrieval practice (Experiment 4). Results of a 2×3 ANOVA with the factors of ITEM TYPE (*p-*, *c-*) and PRACTICE CONDITION showed no significant interaction between the two factors, $F(2, 141) = 1.247$, $p = .290$, $\eta^2 = .018$, indicating equal amounts of forgetting across experiments.

larger when restudy preceded by imagination (9.9%) or restudy preceded by semantic generation (7.8%) were employed than after retrieval practice (7.7%). This suggests that the alleged context change should have been at least as large after restudy as after retrieval practice. Consequently, the disparity of the effects of the single practice methods on item recognition should not have arisen from a smaller context change after restudy trials.

Second, the present Experiment 5 differed from Jonker et al.'s (2013) Experiment 2a in several respects. Unlike Jonker et al. (2013), order of items within the final test was not blocked by category but by item type. This procedure may in fact underestimate the context change effect by increasing the likelihood that the study context is reinstated. This violates Tenet 2 and potentially eliminates the context effect according to the context change account and Jonker et al.'s (2013; Experiments 2b & 3) findings. Despite this possibility, the amount of context-dependent forgetting in recall in the present experiments (9.9% and 7.8%) was comparable to the ones reported by Jonker et al. (approx. 7% and 9%) suggesting that testing blocked by item type did not bias the results substantially. Moreover, it has been common practice to block the test order by item type as well as by category, both methods producing reliable RIF. As the context change account should be able to explain the entirety of RIF findings, this methodological variance should not pose a problem. Hence, it seems legitimate to draw comparisons between Jonker et al.'s findings and the current recall and recognition data and interpret the diverging findings as valid evidence against the context change account.

Inhibition account

The inhibition account assumes that RIF arises due to interference and inhibition of the unpracticed items when participants attempt to retrieve the to-be-practiced items. Unlike (competitive) retrieval, other practice formats should not lead to inhibition of unpracticed items and RIF-like forgetting. RIF is supposed to be retrieval specific, i.e., retrieval is necessary for RIF to arise whereas restudy or any other practice format is not sufficient. This assumption disagrees with the equivalence proposals by both the blocking and the context change accounts. None of the present experiments were designed to directly test the inhibition account, yet the results may still provide an informative basis for its validity.

The recall findings of Experiments 1-6, which showed equivalent effects of retrieval practice and restudy formats enhancing cue-item associations as well as restudy preceded by context change tasks, disagree with the predictions by the inhibition account. They indicate that retrieval practice is not necessary for RIF to occur, but that other practice formats, like the ones employed here, are sufficient to induce forgetting. The recognition findings of Experiments 1-6, on the other hand, are compatible with the inhibition account of RIF. The results showed that exclusively (competitive) retrieval practice impaired recognition memory of the unpracticed items, whereas the other five practice formats did not. Thus, the inhibition account cannot explain the whole range of results presented in this thesis but it can explain the recognition findings that were inconsistent with both the blocking and the context change accounts. This indicates that RIF cannot be attributed to blocking or context change effects alone, but leaves room for a different underlying mechanism, potentially inhibition, that affects not only recall but also item recognition. It is worth noting, that proponents of the inhibition account do not claim that inhibition single-handedly causes RIF but they concede that inhibition together with other factors plays a role in RIF.

Blocking and context change proponents have pointed out theoretical and empirical objections against the inhibition account (e.g., Jonker et al, 2013; Raaijmakers & Jakab, 2013; Verde, 2013). Nonetheless, results of the large-scale meta-analysis by Murayama et al. (2014) including more than 500 samples favored the inhibition account of RIF supporting a majority of its predictions. Interestingly, however, the retrieval specificity assumption was challenged as noncompetitive retrieval practice was found to reliably induce RIF like competitive retrieval practice does, in line with the present recall results (but note that, in the meta-analyses, the few samples included in this analysis were highly heterogeneous). Apart from the negative findings on retrieval specificity in recall, the finding that reinstatement of the study context eliminates RIF (Jonker et al., 2013) is difficult to reconcile with the inhibition account, like it can hardly be explained by blocking effects. Following the rationale of the inhibition account, instructing participants to think back to the study episode should not affect the inhibited status of the unpracticed items and thus the inhibition account should predict RIF to persist. The elimination of RIF following reinstatement of the study context conflicts with this prediction and thus represents a challenge to the inhibition account.

Multiple-factor accounts

In the past, most studies have been dedicated to identifying a single mechanism underlying RIF, i.e., they addressed the question whether either inhibition or blocking or context change mediated RIF, giving the impression that these mechanisms were mutually exclusive. Importantly, Grundgeiger's (2014) and the present recognition findings disagree with the idea of attributing RIF to blocking or context change alone as neither strength-based forgetting nor context-dependent forgetting could be observed in recognition tests whereas RIF was robustly found. This indicates a crucial role for a third mechanism that either acts alone or as an additional process. Inhibition would seem the obvious candidate, yet neither inhibition can account for the entirety of findings.

Taking the deficiencies of the single accounts into consideration, it has been suggested that more than one mechanism may contribute to RIF (e.g., M. C. Anderson & Levy, 2007; Aslan & Bäuml, 2010; Bäuml, 2008; Grundgeiger, 2014; Jonker, Seli, & MacLeod, 2015; Schilling, Storm, & Anderson, 2014; Storm & Levy, 2012; see also M. C. Anderson et al., 1994, p. 1080). Most of such multiple-factor accounts assume that inhibition and blocking work in concert for RIF to arise. Inhibition is supposed to be active during the practice phase: While unpracticed category members intrude during practice potentially compromising retrieval success, inhibition processes are recruited to solve interference by reducing these items' traces. As inhibition works on the level of item representations, the effects of inhibition should be observable over a wide range of tests, including item recognition. Blocking processes are further assumed to be active during the final test. Due to significant strengthening of a subset of items during retrieval practice, these practiced items may block recall of unpracticed category members. This strength-based blocking effect should mainly be observable in tests that do not rely on item-specific cues and the contribution of blocking should decrease the more item-specific cues are provided. In item recognition, i.e., when the item itself is presented as a retrieval cue, blocking effects should be largely absent and blocking should not contribute to RIF observed in such a test. Accordingly, the type of test, particularly its susceptibility to blocking, is supposed to determine the amount to which the mechanisms mediate the observed forgetting. Thus, RIF is regarded a product of both blocking and inhibition whereas RIF-like forgetting induced by the restudy formats is the product of blocking alone. In tests that are susceptible to blocking and inhibition, like cued-recall tests, retrieval practice as well as the re-exposure formats should induce recall impairment of the unpracticed items. In tests that are not susceptible to blocking, but are susceptible to inhibition, like item recognition tests, only retrieval practice but none of the re-exposure formats should induce impairment of the unpracticed items in a final recognition test.

Such a two-factor account is consistent with the results observed in Experiments 1-3 demonstrating RIF in both cued recall and item recognition tests, but RIF-like effects only in cued recall tests. As indicated above, the present findings of Experiments 5 and 6 might also result from blocking. Conceding that restudy following a context change may lead to adequate strengthening, like practice formats that enhance cue-item associations, blocking may be accountable for the presence of RIF-like forgetting after a context change task and restudy. Therefore, the results of Experiments 4-6 which showed RIF and RIF-like forgetting in the cued recall test, and RIF but not RIF-like forgetting in the recognition test, are also compatible with this two-factor account. However, the above mentioned limitations to this interpretation should be regarded.

Further empirical evidence provides support for this account, one line of evidence concerning correlational analysis, another line concerning individual differences work. The blocking account predicts a correlation of the amounts of enhancement and forgetting: the more the practiced items are strengthened, the higher is their blocking potential, and, thus, the harder it should be to recall the unpracticed items. Consequently, the amount of strengthening should be reflected in the amount of forgetting. In line with predictions by the two-factor account specified above, such a correlation has been observed when category-cued recall tests were used to assess memory performance in the Retrieval-practice Paradigm, but not when item-specific cues were provided such as in category-plus-stem-cued recall tests or in item recognition tests (see the large-scale meta-analysis by Murayama et al., 2014). Furthermore, inhibitory capacity measured with individual motor response inhibition correlates significantly with RIF in tests that provide item-specific cues, like item recognition and category-plus-stem-cued recall, but the direction of the correlation is reversed when RIF is assessed in a category-cued recall test (Schilling et al., 2014). This indicates that tests susceptible to blocking, i.e., item-nonspecific tests, mask the inhibitory contribution to RIF whereas item-specific tests reflect RIF induced by inhibition. However, correlational results should not be interpreted as conclusive evidence, given that difference scores represent unreliable measures and that individual differences may confound the effect (for a similar argument, see Raaijmakers & Jakab, 2013).

Research on individuals whose inhibitory capabilities are believed to be compromised may provide more compelling results. In line with the two-factor account, RIF should arise in participants with impaired inhibitory capabilities when RIF is mediated mainly by strength-based mechanisms but RIF should fail to arise when it is mediated by inhibition. In fact, studies examining

RIF in children, schizophrenic patients, and ADHD patients show reliable RIF when category-cued recall tests are employed but show no RIF when item recognition tests or category-plus-stem-cued recall tests are administered (Aslan & Bäuml, 2010; Soriano, Jiménez, Román, & Bajo, 2009; Storm & White, 2010). Altogether, these findings indicate that when RIF is measured with tests in which no item-specific cues are presented, i.e., tests that are prone to blocking, then blocking effects may mask inhibition effects or rather mediate RIF as in the case of the populations lacking inhibitory control. In tests that provide item-specific cues, blocking should hardly contribute to RIF and thus RIF should mainly represent the result of inhibition.

Proposing a different two-factor account, Jonker et al. (2015) conceded the possibility that RIF may be affected by both blocking and context change. However, this two-factor account is still at odds with the current findings as RIF is found to robustly arise in item recognition tests while strength-dependent forgetting and context-dependent forgetting are eliminated when item recognition testing is employed. Furthermore, a two-factor account including inhibition and context change might be possible, attributing RIF in cued recall to both inhibition and context change and RIF in item recognition to inhibition alone. Such an account has not been specified yet, but it provides a plausible explanation for Jonker et al.'s findings and the present results of Experiments 4-6 and also for Experiments 1-3 conceding that the employed restudy formats may have engaged retrieval.

Even though it might seem plausible that two factors or even all three factors contribute to RIF in cued recall tests, the expectation may arise that forgetting mediated by multiple mechanisms may be more pronounced compared to forgetting mediated by a single mechanism. If retrieval, for instance, engaged inhibition during practice and blocking at test, then the amount of RIF should exceed the amount of forgetting induced by advanced strengthening via re-exposure formats and thus by blocking alone or by context manipulations prior to restudy and thus by context change alone. In contrast to this expectation, the present recall findings (Experiments 1a, 2a, & 3a;

Experiments 4-6) showed that RIF and RIF-like forgetting were similar in amount. A possible explanation might be provided by M. C. Anderson and Levy's (2007) correlated cost and benefit assumption. It seems plausible that inhibitory mechanisms operate not only during retrieval practice but also during retrieval in the final test: When unpracticed items are to be retrieved, the interfering practiced items should be inhibited. Consequently, inhibition may reduce the blocking contribution to RIF so that in individuals with pronounced inhibitory capabilities inhibition will contribute much more than blocking whereas in individuals with poor inhibitory capabilities blocking will mediate RIF primarily, as the practiced items are not inhibited. Furthermore, in opposition to Raaijmakers and Jakab's (2013) view, one might consider the possibility that blocking effects may be more prominent following re-exposure compared to retrieval practice. A recent study shows that retrieval practice can reduce intralist interference compared to restudy (Kliegl & Bäuml, 2016; see also Abel & Bäuml, 2014; Halamish & Bjork, 2011), suggesting that blocking effects might decrease after retrieval practice and therefore RIF-like forgetting following re-exposure may in fact be smaller than RIF. In this manner, additive effects of inhibition and reduced blocking in RIF may result in a similar amount of forgetting in recall as purely blocking-based forgetting.

Regarding previous RIF findings as a whole, a comprehensive account of RIF may include all three factors, inhibition, blocking, and context change. While the effects of blocking and context change should be restricted to tests that hardly rely on item-specific cues, such as category-cued tests and, to a lesser extent, category-plus-stem-cued recall tests, inhibition effects may be uncovered and isolated in item-specific tests such as item recognition. In addition, inhibitory effects should be retrieval specific, blocking- and context-based effects, on the contrary, should not. However, future research is warranted to determine if these three factors underlie RIF and to determine their respective contributions to the unpracticed items' impairment.

5.2 Retrieval specificity

The inhibition account of RIF assumes that retrieval practice is a necessary condition for RIF to arise as only retrieval practice triggers interference and thus inhibition of unpracticed items. In contrast, both blocking and context change accounts do not advance the view that RIF is retrieval specific, but rather assume that RIF can be induced by restudy when particular conditions are met. Thus, at its core, the retrieval specificity assumption disagrees with the equivalence assumptions of the blocking and the context change accounts. In the present thesis, the retrieval specificity proposal was rather used as a tool to operationalize the equivalence hypotheses of the blocking and the context change accounts than to be the focus of investigation. Nevertheless, the findings bear implications for the validity of the retrieval specificity assumption which will be discussed in the following passages.

The studies outlined in the introduction were originally conducted in order to test the retrieval specificity property of the inhibition account of RIF. The first line of studies provided demonstrations of retrieval specificity of RIF, showing that forgetting arises only following competitive retrieval, but not following restudy and noncompetitive retrieval practice - manipulations that were supposed to strengthen cue-item associations without the induction of interference and inhibition (e.g., M. C. Anderson, Bjork et al., 2000; Bäuml & Aslan, 2004; Ciranni & Shimamura, 1999; Dobler & Bäuml, 2013; Ferreira et al., 2014; Hanslmayr et al., 2010; Hulbert et al., 2012; Staudigl et al., 2010). Theoretical objections were raised regarding the implications of the findings on the dismissal of the blocking account (Raaijmakers & Jakab, 2012; Verde, 2013) which were empirically supported by experiments showing RIF-like forgetting following other forms of practice such as an adapted noncompetitive retrieval practice task (Raaijmakers & Jakab, 2012; see also Jonker & MacLeod, 2012) and re-exposure supplemented with category judgments or pleasantness ratings (Verde, 2013). These findings challenged the retrieval specificity assumption and the inhibition account of RIF.

As discussed in the introduction, Grundgeiger (2014) first demonstrated a possible role for the testing format in determining whether retrieval practice is necessary for RIF to arise and thus whether retrieval specificity holds. Comparing the effects of competitive retrieval practice and Raaijmaker and Jakab's (2012) variant of the noncompetitive retrieval practice task, he found both competitive retrieval practice and noncompetitive retrieval practice to impair recall performance of unpracticed items. When memory performance was assessed using item recognition testing, only competitive, but not noncompetitive retrieval practice induced forgetting. These findings indicate that while RIF is not specific to competitive retrieval practice in cued recall tests, it is in item recognition tests. The present findings support this implication. Across all six experiments, all practice formats - competitive retrieval practice, restudy formats that enhanced cue-item associations, and restudy preceded by context change tasks - reliably reduced recall performance challenging retrieval specificity. Moreover, the findings consistently showed that only competitive retrieval practice but none of the other practice formats impaired item recognition performance. Thus, the present findings together with Grundgeiger's data indicate that the retrieval specificity property of RIF is contingent on the final test format: Retrieval specificity of RIF is not supported in cued recall tests, as adequate strengthening of cue-item associations or re-exposure preceded by context change was sufficient to induce RIF-like forgetting. However, retrieval specificity holds when item recognition tests are employed to assess memory.

5.3 Further directions

The goal of the present thesis was to contribute to settling the debate on the underlying mechanisms of RIF. This task was successfully accomplished as all prevalent theories, assuming a single underlying mechanism to be responsible for RIF, could be suspended. The results presented in this thesis lead to the conclusion that neither of the most prominent processes can account for the findings suggesting two or even three mechanisms working in concert. On the basis of the present findings, it is not possible to determine any further which mechanisms act together to induce RIF, thus further research is warranted. Moreover, replications and extensions of the present experiments are necessary to increase the reliability of the present findings.

Particularly, future research regarding the role of context change in RIF is required including compelling replications of the finding by Jonker et al. (2013, Experiments 3) showing reinstatement of the context to eliminate RIF, as empirical evidence supporting its contribution seems the least well established. Furthermore, the Two-faces Paradigm may represent a viable method to test the context change account. In the Two-faces Paradigm, participants typically study a list of items and either receive a context change instruction or not before studying a second list (Bäuml & Dobler, 2015; Bäuml & Samenieh, 2010, 2012; Bäuml & Schlichting, 2014). In the final test, target items are either tested immediately or after retrieval of non-target items. The results typically show that when no context change is induced, prior non-target recall reduces target recall when compared to immediate recall of target items (see also related results on Output Interference). When context is changed between lists, prior non-target recall enhances target recall. On the basis of the context change account of RIF and the assumption that retrieval practice induces a context change, the expectation arises that the effects of retrieval practice may imitate the effects of established context manipulations in the Two-faces Paradigm: after a retrieval-practice phase, prior non-target recall may enhance target recall whereas after a restudy phase, and thus no context change, prior non-target recall may reduce target recall. Such a test may shed more light on whether retrieval practice induces a context change and thus on the contribution of context change to RIF.

Considering the role of the test format in decomposing the contributions of blocking, context change, and inhibition, as suggested by Grundgeiger

(2014) and the present results, possible alternative assessment methods of the inhibitory contribution come to mind. Item recognition tests are frequently regarded an instance of independent-probe tests (M. C. Anderson, 2003; Ortega, Gómez-Ariza, & Bajo, 2012; Román et al., 2009; Schilling et al., 2014; Spitzer & Bäuml, 2007), thus other forms of independent probes may qualify as suitable procedures to measure the inhibitory contribution to RIF. Test cues are termed independent probes if they differ from the original cue during study. As inhibition is supposed to work on the level of item representations, inhibition effects are supposed to appear regardless of the provided cue, i.e., in tests using independent probes. In contrast, blocking and context effects rely on the association between the item and the original cue. Using a novel cue should, thus, release memory impairment induced by these processes.

In the original experiment by M. C. Anderson and Spellman (1995), participants studied a categorized item list that featured items (e.g., *strawberry*) that were presented with a particular category cue (e.g., FOOD), but are also implicitly classified into a different semantic category comprised in the study list (e.g., RED). They observed that when a subset of items of the category RED was retrieval-practiced, not only the unpracticed items from practiced categories (e.g., *tomato*) showed recall impairment, but also memory performance for related control items (e.g., *strawberry*) dropped (see also M. C. Anderson & Bell, 2001; Saunders & MacLeod, 2006). Apart from this cross-category cuing, this finding has also been replicated with novel extra-list cues showing that *tomato* is impaired when cued with SAUCE when it was originally paired with FOOD, for instance (Aslan et al., 2007; S. K. Johnson & Anderson, 2004; Shivde & Anderson, 2001; but see Camp et al., 2007; Jonker et al., 2012; Perfect et al., 2004; Williams & Zacks, 2001, for failures to find RIF using independent probes). Both blocking- and context-based accounts of RIF have difficulty explaining these findings. Thus, independent-probe testing may represent a further feasible method to separate inhibitory contribution to RIF from blocking and context change contributions. Accordingly, the results from the present experiments are expected to generalize from item recognition tests to independent-probe tests.

5.4 Conclusions

The present thesis contributes to the question which mechanism underlies the observed memory impairment following selective retrieval. The results disagree with the central hypotheses of both the blocking and the context change accounts and thus strongly suggest their dismissal as the single underlying mechanism of RIF. According to the blocking account, retrieval practice and other forms of strengthening of cue-item associations should have equivalent effects on memory performance in the Retrieval-practice Paradigm. Even though this equivalence arose in recall memory, the effects of retrieval differed from the effects of other practice formats in item recognition memory. Similarly, the context change account assumes that retrieval practice and any other practice format preceded by a context change will affect memory Retrieval-practice Paradigm equivalently. Again, in the cued recall test, retrieval practice and restudy preceded by context change induced forgetting whereas, in the item recognition test, only retrieval but not restudy preceded by context change impaired memory performance. Evaluating the present findings in light of the inhibition account of RIF, it shows that, while the recall results disagree with its retrieval specificity property, the recognition findings are consistent with the inhibition account.

Given the entirety of findings, none of the most prominent accounts (blocking, context change, inhibition) can explain RIF sufficiently. Strength and context manipulations, when implemented in the Retrieval-practice Paradigm, affect recall indicating a role for these factors in RIF when recall tests are employed. These effects however do not generalize to item recognition testing whereas RIF has repeatedly been found in recognition tests. This finding suggests a critical role for inhibition in RIF, which may however be masked by other factors such as blocking or context effects when tests other than item-specific recognition are involved. Therefore, multiple-factor accounts have been discussed, of which a two-factor account attributing RIF to a combination of inhibition and blocking mechanisms seems to be the most promising and yet economical candidate. The validity of such multiple-factor accounts has yet to be established.

Moreover, the present results indicate that, although the retrieval-specificity property predicted by the inhibition account does not hold in cued recall tests, RIF seems in fact to be retrieval specific when item recognition tests are used to assess memory performance. Thus, the results on retrieval specificity assessed in a cued recall test do not simply generalize to results on retrieval specificity employing item recognition testing.

There is little doubt, that strength-based blocking and context change represent essential mechanisms in memory and viable explanations to a variety of memory phenomena. It seems neat and economical to apply these well-established accounts to new findings, whereas the idea of inhibition may strike one as a relic of the Freudian era. However, as the present thesis suggests, Retrieval-induced Forgetting cannot be simply reduced to strength-based or context-dependent forgetting but it may arise due to the combination of at least two processes. More specifically, the present findings indicate a role for inhibition in the rise of Retrieval-induced Forgetting.

Literature

Abel, M., & Bäuml, K.-H. T. (2012). Retrieval-induced forgetting, delay, and sleep. *Memory, 20*, 420-428.

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

Allen, B. P., & Lindsay, D. S. (1998). Amalgamations of memories: Intrusion of information from one event into reports of another. *Applied cognitive Psychology, 12*, 277-285.

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

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6*, 451-474.

Anderson, J. R. (1981). Interference. The relationship between response latency and response accuracy. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 326-343.

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

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General, 128*, 186-197.

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

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130*, 544-570.

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review, 7*, 522-530.

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

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

Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 1141-1159.

Anderson, M. C., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In D. S. Gorfein & C. M. MacLeod (Eds.), Inhibition in cognition (pp. 81102). Washington, DC: American Psychological Association.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 608-629.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (pp.237-313). San Diego: Academic Press.

Anderson, M. C., Ochsner, K., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., et al. (2004). Neural systems underlying the suppression of unwanted memories. *Science, 303*, 232-235.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102*, 68-100.

Aslan, A., & Bäuml, K.-H. T. (2010). Retrieval-induced forgetting in young children. *Psychonomic Bulletin & Review, 17*, 704-709.

Aslan, A., & Bäuml, K.-H. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 37*, 264-269.

Aslan, A., Bäuml, K.-H., & Pastötter, B. (2007). No inhibitory deficit in older adults episodic memory. *Psychological Science, 18*, 72-78.

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

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

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces. *Journal of Experimental Psychology: General, 104*, 54-75.

Bajo, M. T., Gomez-Ariza, C. J., Fernandez, A., & Marful, A. (2006). Retrieval-induced forgetting in perceptually driven memory tests.

Journal of Experimental Psychology: Learning, Memory, & Cognition, 32, 1185-1194.

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

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A *Journal of Experimental Psychology, 80*, 1-46.

Bäuml, K.-H. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review, 5*, 459-463.

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

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

Bäuml, K.-H., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition, 32*, 610-617.

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

Bäuml, K-H., & Hartinger, A. (2002). On the role of item similarity in retrieval-induced forgetting. *Memory, 10*, 215-224.

Bäuml, K.-H., & Kuhbandner, C. (2007). Remembering can cause forgetting - but not in negative moods. *Psychological Science, 18*, 111-115.

Bäuml, K.-H. T., & Samenieh, A. (2010). The two faces of memory retrieval. *Psychological Science, 21*, 793795.

Bäuml, K.-H. T., & Samenieh, A. (2012). Selective memory retrieval can impair and improve retrieval of other memories. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*, 488-494.

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

Bilodeau, I. M., & Schlosberg, H. (1951). Similarity in stimulating conditions as a variable in retroactive inhibition. *Journal of Experimental Psychology, 41*, 199-204.

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton and E. Martin (Eds.): *Coding Processes in Human Memory* (pp. 217-235). Washington DC: Winston.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123144). Hillsdale, NJ: Erlbaum.

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language, 48*, 563-580.

Bodner, G. E., & Richardson-Champion, D. D. L. (2007). Remembering is in the details: Effects of test-list context on memory for an event. *Memory, 15*, 718-729.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*, 129-148.

Bower, G. H., Monteiro, K. P., & Gilligan, S. G. (1978). Emotional mood as a context for learning and recall. *Journal of Verbal Learning and Verbal Behavior, 17*, 573-585.

Brinegar, K. A., Lehman, M., & Malmberg, K. J. (2013). Improving memory after environmental context change: A strategy of "preinstatement". *Psychonomic Bulletin & Review, 20*, 528-533.

Buchli, D. R., Storm, B. C., & Bjork, R. A. (2015). Explaining retrieval-induced forgetting: A change in mental context between the study and restudy practice phases is not sufficient to cause forgetting. *The Quaterly Journal of Experimental Psychology*, 1-13.

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

Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 1314-1319.

Camp, G., Pecher, D., & Schmidt, H. G. (2007). No retrieval-induced forgetting using item-specific independent cues: Evidence against a general inhibitory account. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33*, 950-958.

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

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect,and text processing. *Journal of Memory and Language, 61*, 153-170.

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

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*, 553-571.

Chandler, C. C. (1989). Specific retroactive interference in modified recognition tests: Evidence for an unknown cause of interference.

Journal of Experimental Psychology: Learning, Memory, & Cognition, 15, 256265.

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

Craik, F. I. M., & Schloerscheidt, A. M. (2011). Age-related differences in recognition memory: Effects of materials and context change. *Psychology & Aging, 26*, 671-677.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Lawrence Erlbaum.

Dallett, K., & Wilcox, S. G. (1968). Contextual stimuli and proactive inhibition. *Journal of Experimental Psychology, 78*, 475480.

Delaney, P. F., Sahakyan, L., Kelley, C. M., & Zimmerman, C. A. (2010). Remembering to forget: The amnesic effect of daydreaming. *Psychological Science, 21*, 10361042.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452-478.

Depue, B. E., Banich, M. T., & Curran, T. (2006). Suppression of emotional and nonemotional content in memory. *Psychological Science, 17*, 441-447.

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

Dobler, I. M., & Bäuml, K.-H. T. (2013). Retrieval-induced forgetting: Dynamic effects between retrieval and restudy trials when practice is mixed. *Memory & Cognition, 41*, 547-557.

Dulsky, S. G. (1935). The effect of a change of background on recall and relearning. *Journal of Experimental Psychology, 18*, 725-740.

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*, 524-542.

Ebbinghaus, H. (1885). *Über das Gedächtnis.* Leipzig: Dunker.

Eich, E. (1995). Searching for mood dependent memory. *Psychological Science, 6*, 67-75.

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

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

Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition, 13*, 333-345.

Ferreira, C. S., Marful, A., Staudigl, T., Bajo, T., & Hanslmayr, S. (2014). Medial prefrontal theta oscillations track the time course of interference during selective memory retrieval. *Journal of Cognitive Neuroscience, 26*, 777-791.

Ford, R. M., Keating, S., & Patel, R. (2004). Retrieval-induced forgetting: A developmental study. *British Journal of Developmental Psychology, 22*, 585603.

Freud, S. (1915). *Repression*, vol. XIV. London: Hogarth.

García-Bajos, E., Migueles, M., & Anderson, M. C. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory, 17*, 92-103.

Gardiner, J. M., Craik, F. I., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1*, 213-216.

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

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1-67.

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

Godden, D. R., & Baddeley, A. D. (1980). When does context influence recognition memory? *British Journal of Psychology, 71*, 99-104.

Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition, 33*, 1431-1441.

Goodmon, L. B., & Anderson, M. C. (2011). Semantic integration as a boundary condition on inhibitory processes in episodic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 416-436.

Greenspoon, J., & Ranyard, R. (1957). Stimulus conditions and retroactive inhibition. *Journal of Experimental Psychology, 53*, 55-59.

Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 1355-1369.

Grundgeiger, T. (2014). Noncompetitive retrieval practice causes retrieval-induced forgetting in cued recall but not in recognition. *Memory & Cognition, 42*, 400-408.

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

Hanslmayr, S., Staudigl, T., Aslan, A., & Bäuml, K.-H. (2010). Theta oscillations predict the detrimental effects of memory retrieval. *Cognitive, Affective, and Behavioral Neuroscience, 10*, 329-338.

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

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review*, *11*, 125-130.

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

Isarida, T., & Isarida, T. K. (2007). Environmental context effects of background color in free recall. *Memory & Cognition, 35*, 1620-1629.

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, & Cognition, 34*, 112-127.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.

Johnson, S. K., & Anderson, M. C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science, 15*, 448-453.

Jonker, T. R., & MacLeod, C. M. (2012). Retrieval-induced forgetting: Testing the competition assumption of inhibition theory. *Canadian Journal of Experimental Psychology, 66*, 204-211.

Jonker, T. R., Seli, P., & MacLeod, C. M. (2012). Less we forget: Retrieval cues and release from retrieval-induced forgetting. *Memory & Cognition, 40*, 1236-1245.

Jonker, T. R., Seli, P., & MacLeod, C. M. (2013). Putting retrieval-induced forgetting into context: An inhibition-free, context-based account. *Psychological Review, 120*, 852-872.

Jonker, T. R., Seli, P., & MacLeod, C. M. (2015). Retrieval-induced forgetting and context. *Current Directions in Psychological Science, 24*, 273-278.

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

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 701-722.

Kinnell, A., & Dennis, S. (2011). The list length effect in recognition memory: An analysis of potential confounds. *Memory & Cognition, 39*, 348-363.

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

Kliegl, O., & Bäuml, K.-H. T. (2016). Retrieval practice can insulate items against intralist interference: Evidence from the list-length effect, output interference, and retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 202-214.

Koessler, S., Engler, H., Riether, C., & Kissler, J. (2009). No retrieval-induced forgetting under stress. *Psychological Science, 20*, 1356-1363.

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*, 85-97.

Koutstaal, W., Schacter, D. L., Johnson, M. K., & Galluccio, L. (1999). Facilitation and impairment of event memory produced by photograph review. *Memory & Cognition, 27*, 478-493.

Kuhl, B., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal neural processing benefits of forgetting. *Nature Neuroscience, 10*, 908-914.

Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning & Verbal Behavior, 9*, 1-11.

Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory, 19*, 346-359.

Loftus, E. F. (1992). When a lie becomes the memory's truth: Memory distortion after exposure to misinformation. *Current Directions in Psychological Science, 13*, 145-147.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile deconstruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585-589.

Macht, M. L., Spear, N. E., & Levis, D. J. (1977). State-dependent retention in humans induced by alterations in affective state. *Bulletin of the Psychonomic Society, 10*, 415-418.

Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 153-161.

MacLeod, M. (2002). Retrieval-induced forgetting in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology, 16*, 135-149.

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

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). London, NJ: Lawrence Erlbaum Assoc Inc.

Macmillan, N.A., Rotello, C.M., & Miller, J.O. (2004). The sampling distributions of Gaussian ROC statistics. *Perception & Psychophysics, 66*, 406-421.

Mannhaupt, H.-R. (1983). Produktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien. *Sprache & Kognition, 2*, 264-278.

McGeoch, J. A. (1942). *The psychology of human learning.* New York:Longmans, Green.

Meeter, M., Murre, J. M., Janssen, S. M. (2005). Remembering the news: Modeling retention data from a study with 14,000 participants. *Memory & Cognition, 33*, 793-810.

Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology, 53*, 173-203.

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

Migueles, M., & García-Bajos, E. (2007). Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology, 21*, 11571172.

Miguez, G., Mash, L. E., Polack, C. W., & Miller, R. R. (2014). Failure to observe renewal following retrieval-induced forgetting. *Behavioural Processes, 103*, 43-51.

Moosbrugger, H., Oehlschlägel, J., & Steinwascher, M. (2011). Frankfurter Aufmerksamkeits-Inventar 2. Bern: Huber.

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

Murdock, B. B. Jr. (1982). A theory for storage and retrieval of item and associative information. *Psychological Review, 89*, 609-626.

Murnane, K., & Shiffrin, R. M. (1991). Word repetitions in sentence recognition. *Memory & Cognition, 19*, 119-130.

Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift fr Psychologie, 1*, 1-300.

Ortega, A., Gómez-Ariza, C. J., Román, P., & Bajo, M. T. (2012). Memory inhibition, aging, and the executive deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*, 178-186.

Osgood, C. E. (1953). *Method and theory in experimental psychology.* New York: Oxford University Press.

Parks, C. M., & Yonelinas, A. P. (2008). Theories of recognition memory. In H. L. Roediger, III (Ed.), *Cognitive Psychology of Memory of Learning and Memory: A comprehensive reference* (Vol. 2, pp. 389-416). Oxford: Elsevier.

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

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

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

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 37*, 287-297.

Pear, T. H. (1922). *Remembering and forgetting.* London: Metheun.

Perfect, T. J., Moulin, C. J. A., Conway, M. A., & Perry, E. (2002). Assessing the inhibitory account of retrieval-induced forgetting with implicit-memory tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 1111-1119.

Perfect, T. J., Stark, L.-J., Tree, J. J., Moulin, C. J. A., Ahmed, L., & Hutter, R. (2004). Transfer appropriate forgetting: the cue-dependent nature of retrieval-induced forgetting. *Journal of Memory & Language, 51*, 399-417.

Posner, M. I., & Warren R. E. (1972). Traces, concepts, and conscious constructions. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp.25-43). Washington, D.C: Winston.

Postman, L. (1971). Transfer, interference, and forgetting. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth and Schlosberg's Experimental Psychology* (pp.1019-1132). New York: Holt, Rinehart, and Winston.

Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning & Verbal Behavior, 7*, 672-694.

Raaijmakers, J. G. W. (2008). Mathematical models of human memory. In H. L. Roediger, III. (Ed.), *Cognitive psychology of memory. Learning and memory: A Comprehensive Reference* (Vol. 2, pp. 445-466). Oxford: Elsevier.

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

Raaijmakers, J. G. W., & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of Memory & Language, 68*, 98-122.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM. A theory of probabilistic search of associative memory. In G. Bower (Ed.), *The Psychology of learning and motivation, Vol. 14*. New York: Academic Press.

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

Racsmany, M., Conway, M. A., & Demeter, G. (2010). Consolidation of episodic memories during sleep: Long-term effects of retrieval practice. *Psychological Science, 21*, 80-85.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 163-178.

Roediger, H. L. III. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior, 12*, 644-657.

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

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 2027.

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

Roediger, H. L., & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired associate lists. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 91-105.

Román, P., Soriano, M. F., Gómez-Ariza, C. J., & Bajo, M. T. (2009). Retrieval-induced forgetting and executive control. *Psychological Science, 20*, 1053-1058.

Rundus, D. (1973). Negative effects of using list items as retrieval cues. *Journal of Verbal Learning and Verbal Behavior, 12*, 43-50.

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

Sahakyan, L., & Hendricks, H. E. (2012). Context change and retrieval difficulty in the list-before-last paradigm. *Memory & Cognition, 40*, 844-860.

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

Saunders, J., Fernandes, M., & Kosnes, L. (2009). Retrieval-induced forgetting and mental imagery. *Memory & Cognition, 37*, 819-828.

Saunders, J., & MacLeod, M. D. (2006). Can inhibition resolve retrieval competition through control of spreading activation? *Memory & Cognition, 34*, 307-322.

Scheithe, K., & Bäuml, K.-H. (1995). Deutschsprachige Normen für Vertreter von 48 Kategorien. *Sprache & Kognition, 14*, 39-43.

Schilling, C. J., Storm, B. C., & Anderson, M. C. (2014). Examining the costs and benefits of inhibition in memory retrieval. *Cognition, 133*, 358-370.

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

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2*, 249-253.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 179-195.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM - retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145-166.

Shivde, G., & Anderson, M. C. (2001). The role of inhibition in meaning selection: Insights from retrieval-induced forgetting. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 175190). Washington, DC: American Psychological Association.

Simon, H. A., & Feigenbaum, E. A. (1964). An information processing theory of some effects of similarity, familiarization, and meaningfulness in verbal learning. *Journal of Verbal Learning and Verbal Behavior, 3*, 385-396.

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

Smith, A. D. (1973). Input order and output interference in organized recall. *Journal of Experimental Psychology, 100*, 147-150.

Smith, A. D., D'Agostino, P. R., & Reid, L. S. (1970). Output interference in long-term memory. *Canadian Journal of Psychology, 24*, 85-89.

Smith, R. E., & Hunt, R. R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition, 28*, 503-508.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 460-471.

Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition, 10*, 405-412.

Smith, S. M. (1984). A comparison of two techniques for reducing context-dependent forgetting. *Memory & Cognition, 12*, 477-482.

Smith, S. M. (1988). Environmental context effects on memory. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13-34). New York: Wiley.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*, 342353.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*, 203-220.

Soares, J. S., Polack, C. W., & Miller, R. R. (2016). Retrieval-induced versus context-induced forgetting: Does retrieval-induced forgetting depend on context shifts? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 42*, 366-378.

Soriano, M. F., Jiménez, J. F., Román, P., & Bajo, M. T. (2009). Inhibitory processes in memory are impaired in schizophrenia: Evidence from retrieval-induced forgetting. *British Journal of Psychology, 100*, 661-673.

Spitzer, B., & Bäuml, K.-H. (2007). Retrieval-induced forgetting in item recognition: Evidence for a reduction in general memory strength. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33*, 863-875.

Spitzer, B., & Bäuml, K.-H. (2009). Retrieval-induced forgetting in a category recognition task. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 286-291.

Starns, J. J., & Hicks, J. L. (2004). Episodic generation can cause semantic forgetting: Retrieval-induced forgetting of false memories. *Memory & Cognition, 32*, 602-609.

Staudigl, T., Hanslmayr, S., & Bäuml, K.-H. T. (2010). Theta oscillations reflect dynamics of interference in episodic memory retrieval. *The Journal of Neuroscience, 30*, 11356-11362.

Storm, B. C., Angello, G., Buchli, D. R., Koppel, R. H., Little, J. L., & Nestojko, J. F. (2015). A review of retrieval-induced forgetting in the contexts of learning, eyewitness memory, social cognition, autobiographical memory, and creative cognition. *Psychology of Learning and Motivation, 62*, 141-194.

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

Storm, B. C., Bjork, E. L., Bjork, R. A., & Nestojko, J. F. (2006). Is retrieval success a necessary condition for retrieval-induced forgetting? *Psychonomic Bulletin & Review, 13*, 1023-1027.

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

Storm, B. C., & White, H. A. (2010). ADHD and retrieval-induced forgetting: Evidence for a deficit in the inhibitory control of memory. *Memory, 18*, 265-271.

Szpunar, K. K., McDermott, K. B., & Roediger H. L., III (2008). Testing during study insulates against the buildup of proactive interference.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 34, 1392-1399.

Tandoh, K., & Naka, M. (2007). Durability of retrieval-induced forgetting. *Japanese Journal of Psychology, 78*, 310-315.

Thorndike, E. L. (1914). *Educational Psychology: Briefer Course.* new York: Teacher's College, Columbia University.

Tousignant, C., & Bodner, G. E. (2012). Test context affects recollection and familiarity ratings: Implications for measuring recognition experiences. *Consciousness and Cognition, 21*, 994-1000.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Hrsg.), *Organization of memory* (S. 381-403). New York: Academic Press.

Tulving, E. (1974). Cue-dependent forgetting. *American Scientist, 62*, 74-82.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.

Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interferences in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 1*, 321-334.

Tulving, E., & Arbuckle, T. Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology, 72*, 145-150.

Tulving, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology, 92*, 297-304.

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

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 353-370.

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

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Dynamics of context-dependent recall: An examination of internal and external context change. *Journal of Memory and Language, 66*, 1-16.

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

Veling, H. & van Knippenberg, A. (2004). Remembering can cause inhibition: Retrieval-indiced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 315-318.

Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory & Cognition, 32*, 1265-1272.

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

Verde, M. F., & Perfect, T. J. (2011). Retrieval-induced forgetting in recognition is absent under time pressure. *Psychonomic Bulletin & Review, 18*, 1166-1171.

Watkins, M. J. (1975). Inhibition in recall with extralist "cues". *Journal of Verbal Learning & Verbal Behavior, 14*,249-303.

Watkins, M., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning & Verbal Behavior, 15*, 505-517.

Williams, C. C., & Zacks, R. T. (2001). Is retrieval-induced forgetting an inhibitory process? *American Journal of Psychology, 114*, 329-354.

Winograd, E., & Rivers-Bukeley, N. T. (1977). Effects of changing context on remembering faces. *Journal ofExperimental Psychology: Human Learning & Memory, 3*, 397-405.

Wixted, J. T. (2007a). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152-176.

Wixted, J. T. (2007b). Signal-detection theory and the neuroscience of recognition memory. In J. S. Nairne (Ed.), *The Foundations of Remembering: Essays in Honor of H. L. Roediger, III* (pp. 67-82). New York: Psychology Press.

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

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*, 616-641.

Wundt, W. (1902). *Grundzüge der physiologischen Psychologie* [Principles of physiological psychology] (5th ed.). Leipzig, Germany: Engelmann.