Journal of Memory and Language 86 (2016) 97–118

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Retrieval-induced forgetting in item recognition: Retrieval specificity revisited

Julia Rupprecht, Karl-Heinz T. Bäuml *

Department of Experimental Psychology, Regensburg University, 93053 Regensburg, Germany

ARTICLE INFO

Article history:
Received 22 December 2014
revision received 21 September 2015
Available online 10 October 2015

Keywords:
Episodic memory
Retrieval-induced forgetting
Inhibition
Blocking
Recognition

ABSTRACT

Retrieval-induced forgetting (RIF) refers to the finding that retrieval practice on a subset of studied items can induce later forgetting of related unpracticed items. Although previous studies indicated that RIF is retrieval specific – i.e., it arises after retrieval practice but not after reexposure cycles -, the results of more recent work suggest otherwise, indicating that some reexposure formats can induce RIF very similar to how retrieval practice does. Whereas this prior work employed recall at test, here we revisited retrieval specificity of RIF employing item recognition. The results of three experiments are reported, which examined the effects of retrieval practice and some of the recently suggested reexposure formats on unpracticed items' recognition. In each of these experiments, we showed RIF after retrieval practice but did not find any evidence for RIF-like forgetting after reexposure. These findings demonstrate retrieval specificity of RIF in item recognition, challenging strength-based accounts of RIF and indicating a critical role of inhibition in RIF. Together with the results from the recent recall studies, which we replicated in three further experiments, the present findings are consistent with a two-factor account of RIF, which assigns a role for both inhibition and strength-based blocking in RIF. While both inhibition and blocking may contribute to RIF in certain recall formats, only inhibition may induce RIF in item recognition.

Introduction

Selective retrieval of a subset of studied items can lead to forgetting of related, but not retrieved items. This effect has repeatedly been demonstrated in the retrieval practice paradigm (Anderson, Bjork, & Bjork, 1994). In this paradigm, subjects often learn a categorized item list (e.g., FURNITURE – *lamp*, INSECT – *hornet*, INSECT – *termite*, etc.) in an initial study phase, and, after study, repeatedly retrieve some of the items of some of the categories providing the items' category label and word stem as retrieval cues (e.g., INSECT – *te*___). On a later category-cued recall test, all originally studied items are tested. The typical finding is that recall for the practiced items (e.g., *termite*) is enhanced, but recall for the unpracticed items from the practiced categories (e.g., *hornet*) is impaired, relative to the control items from the unpracticed categories (e.g., *lamp*). The recall impairment for the unpracticed items has been termed *retrieval-induced forgetting* (RIF) and has been shown over a wide range of materials and settings (e.g., Ciranni & Shimamura, 1999; Healey, Ngo, & Hasher, 2014; Levy, McVeigh, Marful, & Anderson, 2007; Storm & Angello, 2010) and a variety of testing formats (e.g., Anderson et al., 1994; Anderson & Spellman, 1995; Hicks & Starns, 2004; Veling & van Knippenberg, 2004).

Anderson et al. (1994) suggested that RIF is induced by active inhibition operating at the retrieval practice stage. According to this view, during retrieval practice, the

* Corresponding author at: Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany.
E-mail address: karl-heinz.baeuml@ur.de (K.-H.T. Bäuml).

http://dx.doi.org/10.1016/j.jml.2015.09.003

practiced categories' not-to-be-retrieved items interfere and are actively inhibited to reduce the interference. Such inhibition is supposed to impair the memory representation of the unpracticed items, reducing access to these items on a later memory test (e.g., Anderson, 2003; Bäuml, Pastötter, & Hanslmayr, 2010; Storm & Levy, 2012). An alternative, noninhibitory account of RIF attributes the forgetting to strength-based blocking processes. Proponents of this view argue that, in the practice phase, the cue-item associations of the practiced items are strengthened, and such strengthening introduces interference of these items during recall of the unpracticed items, thus reducing unpracticed items' recall performance (e.g., Raaijmakers & Jakab, 2013; Verde, 2013). More recently, another noninhibitory account has been suggested, which assumes that the attempt to retrieve items in the practice phase induces a shift in context, thus creating distinct study and practice contexts. According to this account, subjects access the practice context at test when searching for the (practiced and unpracticed) items of the practiced categories but access the study context when searching for the control items, so that memory for the unpracticed items may be reduced relative to the control items and RIF may arise (Jonker, Seli, & MacLeod, 2013; see also Anderson & Bjork, 1994, for an early outline and rejection of a highly similar account). The primary focus of the present study was to contrast the inhibition and blocking accounts of RIF, and it was therefore designed to examine these two accounts. Although the study was not designed to examine the context account, the results will also bear implications for this account. Finally, the results will also allow evaluation of a more general two-factor account of RIF, according to which at least two mechanisms may contribute to RIF (see General discussion section).

Retrieval specificity: the first line of studies

Over the years, several RIF findings have been suggested to be indicative for a critical role of inhibition in RIF (for an overview, see Anderson, 2003). One of these is *retrieval specificity*, the finding that retrieval practice but not restudy may induce forgetting of the unpracticed items. While strength-based accounts of RIF predict that the forgetting of unpracticed items is not restricted to retrieval practice but, in principle, can arise after any kind of strengthening of the cue-item associations of the practiced items, inhibition advocates suggest that the forgetting is retrieval specific. According to this view, retrieval practice, but not restudy of the practiced items, should induce interference and inhibition of the unpracticed items during practice, and thus impair memory for the unpracticed items at test.

Two methods have originally been employed to examine retrieval specificity of RIF: restudy and noncompetitive retrieval practice. In both methods, the to-be-practiced items are reexposed intact with the goal of strengthening the items' associations to their cue without inducing interference and inhibition of related unpracticed items. When employing the restudy method, some of the originally studied category-item pairs were reexposed (e.g., INSECT – *termite*) and participants were instructed to study the word pairs once again. When employing the noncompetitive retrieval practice method, some of the originally studied items were reexposed and subjects were asked to recall the items' category label given the category's word stem as a retrieval cue (e.g., IN___ – *termite*). The results of many studies reported (i) forgetting of unpracticed items 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) and (ii) forgetting of unpracticed items after standard (competitive) retrieval practice but not after noncompetitive retrieval practice (e.g., Anderson, Bjork, & Bjork, 2000; Ferreira, Marful, Staudigl, Bajo, & Hanslmayr, 2014; Hanslmayr, Staudigl, Aslan, & Bäuml, 2010). These findings support retrieval specificity and the inhibition account of RIF and challenge strength-based explanations of RIF.

Retrieval specificity: the second line of studies

More recently, some researchers argued that the previous findings on retrieval specificity may not necessarily contradict assumptions of strength-based accounts of RIF, because reexposure format may be critical for whether reexposure induces forgetting or not (e.g., Raaijmakers & Jakab, 2012; Verde, 2013). Indeed, plain reexposure may not induce forgetting of the unpracticed items, because it may strengthen the representation of the practiced items without strengthening the items' associations to the cue, which may not be sufficient to cause blocking at test. In contrast, RIF may no longer be found to be retrieval specific if retrieval practice was compared to reexposure formats that, like retrieval practice is supposed to do, enhance the cue-item associations of the practiced items. In such case, forgetting of the unpracticed items may arise after both retrieval practice and reexposure, which would be consistent with strength-based accounts of RIF and challenge retrieval specificity and the inhibition view of RIF.

Results from two recent studies support such a view. Employing a modification of the original noncompetitive retrieval practice condition, Raaijmakers and Jakab (2012) had subjects study category-exemplar pairs (e.g., ROUND – *ball*) and asked the subjects in the practice phase to recall a pair's category label presenting the exemplar as a retrieval cue (e.g., ___ – *ball*). In contrast to Anderson et al.'s (2000) original design, the word stems of the category labels were not presented as retrieval cues and items of relatively low frequency within their categories were employed, conditions that likely make noncompetitive retrieval practice more demanding than in the original studies. Doing so, Raaijmakers and Jakab found reduced recall of the unpracticed items after noncompetitive retrieval practice, indicating that the strengthening of the category-exemplar associations can be sufficient to induce the RIF finding (for a related result, see Jonker & MacLeod, 2012).

In the second study, Verde (2013) employed a modification of the original restudy condition, testing the hypothesis that reexposure formats that strengthen category-item associations can induce forgetting similar to how retrieval practice does. Verde designed two reexposure tasks supposed to enhance the category-item associations of the practiced items. Participants learned category-item pairs and were then asked at practice to either judge for each single reexposed cue-item pair whether the category presented was the best to classify the item (best category judgment), or to rate the pleasantness of the reexposed item in the presence of the item's category cue (pleasantness rating). Consistent with the hypothesis, Verde found both reexposure formats to reduce recall of the related unpracticed items at test, indicating that retrieval may not be necessary to induce RIF and reexposure formats that enhance the cue-item associations of the practiced items can be sufficient to cause RIF-like forgetting. These results, together with those of Raaijmakers and Jakab (2012), support strength-based accounts of RIF and challenge retrieval specificity and the inhibition account of RIF.

A study by Saunders, Fernandes, and Kosnes (2009) may further corroborate the view that reexposure of the practiced items can induce RIF-like forgetting. These researchers employed a mental imagery task during practice, reexposing some of the previously studied category-item pairs and asking subjects to engage in mental visualization of exemplar features (e.g., shape, color, size). Mental imagery resulted in RIF-like recall impairment of the not reexposed items, which mimics the effects of standard (competitive) retrieval practice. Although the authors interpreted the finding as evidence for inhibition (see General discussion below), the results appear also consistent with strength-based accounts of RIF. Similar to the reexposure formats employed by Verde (2013), mental imagery may strengthen the cue-item associations of the practiced items more than plain restudy does, thus leading to blocking and RIF-like forgetting of the nonvisualized (unpracticed) items.[1] Thus, the results of Saunders et al. may also support strength-based explanations of RIF and challenge retrieval specificity.

Retrieval specificity: a new line of studies

However, before drawing firm conclusions on whether RIF is retrieval specific or not, it needs to be shown that the results from the second line of retrieval specificity studies are not restricted to recall testing but generalize to other testing formats. Indeed, RIF is not a pure recall phenomenon but can also be found in other testing formats, like, for instance, item recognition. In fact, since Hicks and Starns' (2004) original finding, numerous studies have shown that (competitive) retrieval practice of some items can reduce recognition of related unpracticed items (e.g., Aslan & Bäuml, 2011; Dobler & Bäuml, 2013; Gómez-Ariza, Lechuga, & Pelegrina, 2005; Román, Soriano, Gómez-Ariza, & Bajo, 2009; Spitzer & Bäuml, 2007; Starns & Hicks, 2004; Veling & van Knippenberg, 2004; for confirming results, see also the recent meta analysis by Murayama, Miyatsu, Buchli, & Storm, 2014). These findings are consistent with the inhibition account of RIF, according to which retrieval practice impairs unpracticed items' memory representation, and the detrimental effect of practice should therefore be observable over a wide range of memory tests, including item recognition. Importantly, however, there is reason to expect that, unlike RIF, RIF-like forgetting, as induced by noncompetitive retrieval practice or certain reexposure formats, may not generalize to item recognition.

On the one hand, interference effects from other items may be present in item recognition, as is indicated from studies on retroactive interference and the list length effect (e.g., Chandler, 1989; Gronlund & Elam, 1994; Ratcliff, Clark, & Shiffrin, 1990; but see Dennis & Humphreys, 2001, or Kinnell & Dennis, 2011), and is suggested by global-matching models of recognition memory (e.g., Gillund & Shiffrin, 1984; Murdock, 1982). According to these models, a test probe of an unpracticed item in the retrieval practice paradigm may activate older traces of the item itself as well as related item traces, like the practiced items, based on their similarity to the tested item (e.g., Criss, 2006). The activated related item traces are supposed to reduce the relative match of the test item to the test probe, which may reduce recognition of the to-be-retrieved unpracticed item, inducing RIF as well as RIF-like forgetting. On the other hand, there is clear evidence that not all types of interference effects generalize from recall to item recognition. For instance, the list strength effect – the demonstration that restudy of a subset of studied items can impair memory of the not restudied material – is usually absent in item recognition (e.g., Murnane & Shiffrin, 1991; Ratcliff et al., 1990; Shiffrin, Ratcliff, & Clark, 1990), a finding well captured by models of recognition memory, like the REM model (Shiffrin & Steyvers, 1997). This model explains the absence of list strength effects in item recognition by assuming that restudy leads to more complete memory representations of the restudied items, which are then more easily differentiated from the test probe when it comes to item recognition (see also Dennis & Humphreys', 2001, BCDMEM model of recognition memory that assumes that interference from other items is generally negligible in item recognition). On the basis of the list strength findings and the theoretical analyses of the REM model (and the BCDMEM model), the expectation may thus arise that strength-based interference effects in the retrieval-practice paradigm, as they may be induced by noncompetitive retrieval practice or certain reexposure formats, are also restricted to recall and do not generalize to item recognition. If so, retrieval specificity may not occur in certain recall formats but may occur in item recognition.

A recent study by Grundgeiger (2014) provides first evidence for the possible role of testing format in evaluating retrieval specificity. Applying Raaijmakers and Jakab's (2012) noncompetitive retrieval practice condition, Grundgeiger examined the effects of both competitive and noncompetitive retrieval practice in recall and item recognition testing. While forgetting was observed for both

[1] In both Verde (2013) and Saunders et al. (2009), word pairs were presented to subjects and subjects were asked to rate or imagine the response items in the presence of the stimulus items. In contrast, in older work by, for instance, Hunt and Einstein (1981) and Hockley and Christi (1996), either single words were exposed to subjects and subjects were asked to rate the items, or words pairs were presented and subjects were asked to form interactive images of the two words. Likely, the procedures employed by Hunt and Einstein and by Hockley and Christi induce different item processing than the procedures employed by Verde and Saunders et al. (see Verde, 2013, for corresponding evidence).

practice types when using recall testing, only competitive retrieval practice induced forgetting in item recognition testing, indicating that RIF after noncompetitive retrieval practice may be present in recall but be absent in recognition. The first goal of the present study was to replicate this result and demonstrate that competitive retrieval practice, but not noncompetitive retrieval practice, induces RIF in item recognition. The second goal was to examine whether such pattern would generalize to the other reexposure formats that supposedly strengthen the cue-item associations of the practiced items (e.g., Saunders et al., 2009; Verde, 2013). Indeed, if RIF as it is caused by competitive retrieval practice was present in item recognition, but no RIF or RIF-like forgetting arose after noncompetitive retrieval practice and the recently suggested reexposure formats, then such finding would strongly challenge strength-based accounts of RIF, indicating that the strengthening of the cue-item associations of the practiced items is not sufficient to create RIF. In particular, the finding would suggest that RIF is retrieval specific with item recognition testing.

The present study

In this study, we revisited retrieval specificity of RIF, examining whether noncompetitive retrieval practice and some of the reexposure formats that in more recent studies were shown to induce RIF-like forgetting in recall, do also induce RIF-like forgetting in item recognition. The results of three experiments (Experiments 1–3) are reported designed to directly compare the effects of standard (competitive) retrieval practice to those of Raaijmakers and Jakab's (2012) noncompetitive retrieval practice condition (Experiment 1), Verde's (2013) reexposure with pleasantness rating condition (Experiment 2), and Saunders et al.'s (2009) reexposure with mental imagery condition (Experiment 3). We adopted the blocked design used by Dobler and Bäuml (2013) 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 reexposure formats on the other half of the practiced items second, or vice versa. On the final recognition test, all studied items were tested. Participants discriminated studied items from lures, rating their confidence of an item having been previously studied (old) or not (new; e.g., Macmillan & Creelman, 2004; Parks & Yonelinas, 2008).

On the basis of strength-based accounts of RIF, which assume that both competitive retrieval and (some) reexposure formats increase cue-item associations and such strengthening induces blocking at test (Raaijmakers & Jakab, 2012; Verde, 2013), competitive retrieval and the reexposure formats should affect memory performance of the unpracticed items similarly. That is, depending on whether blocking effects may arise in item recognition or not, either all practice formats – competitive retrieval practice, noncompetitive retrieval practice, reexposure supplemented with pleasantness ratings, and reexposure supplemented with mental imagery – should reduce recognition of unpracticed items, thus generalizing the results from the prior recall studies to item recognition, or all four practice formats should leave recognition of unpracticed items unaffected. Both patterns of results would challenge retrieval specificity of RIF. In contrast, on the basis of the inhibition account of RIF, a different expectation arises. Because retrieval, but not reexposure, of practiced items should induce inhibition of the unpracticed items (e.g., Anderson et al., 1994, 2000), and the inhibitory effects, but not possible blocking effects of practice, may be present in item recognition (e.g., Hicks & Starns, 2004; Ratcliff et al., 1990), Experiments 1–3 may show RIF after standard (competitive) retrieval practice, whereas none of the three experiments may show RIF-like forgetting after any of the reexposure formats. If so, the prior results from recall studies would not generalize to item recognition and RIF would be retrieval specific with item recognition testing.

Finally, the results of three further experiments (Experiments 4–6) are reported with the goal of demonstrating that, for the materials and study and practice procedures employed in Experiments 1–3, both (competitive) retrieval practice and each of the three reexposure formats can induce RIF and RIF-like forgetting when using recall at test. Such finding would replicate the recall results from the previous studies by Raaijmakers and Jakab (2012), Grundgeiger (2014), Verde (2013), and Saunders et al. (2009). Together with the results of Experiments 1–3, it will improve our understanding of retrieval specificity in RIF and the relation between RIF in recall and RIF in item recognition. In particular, the findings will show whether the strengthening of the cue-item associations of the practiced items per se is sufficient to cause RIF, or whether active retrieval is generally necessary to induce RIF.

Experiment 1

Experiment 1 investigated whether noncompetitive retrieval practice induces RIF-like forgetting in item recognition. We compared the standard (competitive) retrieval practice task with the modified noncompetitive retrieval practice task, as it was used in the studies by Raaijmakers and Jakab (2012) and Grundgeiger (2014). According to Raaijmakers and Jakab (2012, p. 25), some requirements should be met for noncompetitive retrieval practice to increase the strengthening of cue-item associations of the practiced items and thus to induce forgetting of the unpracticed items: for instance, initial cue-item associations should be rather weak, i.e., low- to medium-frequency category members should be selected; no initial letters of the to-be-recalled category label should be presented as retrieval cues during practice; feedback should be provided to grant strengthening of associations of all practiced word pairs. We designed Experiment 1 along these requirements.

Method

Participants

Sixty students of Regensburg University took part in the experiment (M = 23.5 years, range = 19–30 years, 42 female). One male participant had to be excluded retrospectively due to technical problems during data logging. All participants spoke German as native language. They received monetary reward for their participation.

Materials

The material contained nine semantic categories (CAR EQUIPMENT, PROFESSIONS, KITCHEN SUPPLIES, CLOTHING, MUSICAL INSTRUMENTS, FOOD, BODY PARTS, MEANS OF TRANSPORTATION, FOUR-LEGGED ANIMALS), each consisting of six studied items and six lures (Mannhaupt, 1983; Scheithe & Bäuml, 1995). In addition, three categories (ALCOHOLIC DRINKS, SANITARY ARTICLES, GEMSTONES) with two items each were drawn from the word norms serving as buffer items in the study and recognition 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 first letter. Following Raaijmakers and Jakab (2012) and Grundgeiger (2014), we used low- to medium-frequency practiced and unpracticed items (*median* = 35.5), as well as low- to medium-frequency lures (*median* = 35.5) as category exemplars. Grundgeiger (2014) demonstrated RIF after noncompetitive retrieval practice with both perceptual and semantic categories. Here we employed semantic categories to allow for better comparison of the results of this experiment to the results of Experiments 2 and 3.

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, an intermediate 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 served as control categories. Practice was blocked by practice type. Half of the subjects completed the competitive retrieval condition before the noncompetitive retrieval condition, the other half vice versa (for a similar design, see Dobler & Bäuml, 2013). For each of the two practice types, three types of items were generated: 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. For the final recognition test, we created three further item types: 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.

Procedure

In the study phase, participants were instructed to learn category-exemplar pairs (e.g., FOOD – *corn*, CLOTHING – *vest*, CLOTHING – *tie*), each presented for 4 s (ISI = 500 ms) on a computer screen. To create the study list, the order of word pairs was blocked randomized: six blocks were randomly arranged, with each block consisting of one exemplar from each category. At the beginning and ending of the study list, three buffer items were presented. Half of the participants received the study list in the original order, the other half studied the items in reversed order. To control for recency effects, participants were distracted after the study phase by counting backwards in steps of 3 for 60 s. 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, subjects were exposed to the category label and initial letter of a previously studied item (e.g., CLOTHING – *v*___) for 5 s (ISI = 500 ms) and asked to recall the corresponding item. Feedback was provided after each trial and subjects were presented with the correct answer for 2 s (ISI = 500 ms; e.g., CLOTHING – *vest*). The order of items within each practice cycle was blocked randomized and items were practiced twice in two consecutive cycles. In the noncompetitive retrieval condition, participants were provided with a particular exemplar of the study list and were asked to recall the corresponding category label they had studied before (e.g., ______ – *vest*). As in the competitive retrieval condition, subjects had 5 s (ISI = 500 ms) to retrieve the category label. Feedback was provided after each trial presenting the correct response for 2 s (ISI = 500 ms; e.g., CLOTHING – *vest*). Again, order of items was blocked randomized and items were practiced twice in two consecutive cycles.

After a distractor task of 8 min (Raven's Progressive Matrices), the final recognition test was administered, in which all exemplars from the study list mingled with lures were presented, each with a schematic rating scale depicted in the lower third of the screen. 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*). Participants typed in their answers on the computer keyboard which were logged automatically. Subjects were instructed to use the whole scope of the rating scale. Presentation was subject-paced, i.e., as soon as the participant entered an answer, the next item appeared on the screen (for arguments in favor of this rating procedure compared to a procedure that requires subjects to make binary old/new decisions only, see Macmillan & Creelman, 2004, or Parks & Yonelinas, 2008). We compiled the recognition list using blocked randomization with two restrictions: old materials as well as new materials appeared at most three times in succession; unpracticed items (*crp*−, *ncrp*−) as well as their control counterparts (*c*−) were presented in the first half of the list in order to avoid confounding output interference effects from the practiced items. Therefore, twelve blocks were arranged: six blocks consisting of *crp*− items, *ncrp*− items, *c*− items, and lures, and six blocks containing *crp+* items, *ncrp+* items, *c+* items, and lures. For each block, we drew one item at a time from each category and arranged the items in random order within the boundaries of the previously mentioned restrictions. Instead of mirror-inverting the recognition list as a whole, the two halves of the recognition lists were mirror-inverted to grant preceding presentation of unpracticed items and their counterparts. Again, half of the participants received the original order of the recognition list, whereas the other participants received the mirror-inverted version. The recognition list always began with three buffer items.

Statistical analysis

Proportion of correctly recognized target items (i.e., hit rate) and proportion of incorrectly recognized lure items (i.e., the false alarm rate) were cumulated across the rating scale starting at the most confident criterion, i.e., *definitely old* ("1"). This procedure leads to an empirical Receiver Operating Characteristic (ROC) curve that relates hits and false alarm rates across variations in response criteria (i.e., the propensity to make a positive recognition response; e.g., Macmillan & Creelman, 2004; Parks & Yonelinas, 2008). With the present 6-point scale, hit and false alarm rates under five different response criteria arose. The first point of the ROC ("1") shows hit and false alarm rates when adopting the strictest response criterion, and each subsequent point ("2", "3", "4", "5") reflects performance at a more and more relaxed response criterion. Importantly, the function is cumulative and so both hit and false alarm rates are constrained to increase or remain constant as the scoring criterion is relaxed.

In the first step, recognition data were analyzed separately for the single response criteria. Corrected hits (hits – false alarms) were calculated for each combination of item type and criterion, and, using ANOVA, it was examined for the three most conservative ("old") response criteria ("1", "2", "3") whether corrected hits for practiced items (*crp+*, *ncrp+*) exceeded corrected hits for the corresponding control items (*c+*), and whether corrected hits for unpracticed items from practiced categories (*crp*−, *ncrp*−) were lower than for the respective controls (*c*−). If such statistical analysis provided significant results, differences between practice conditions were tested as well. Analysis of corrected hits (hits – false alarms) implicitly assumes that the ROC function is linear (e.g., Wixted, 2007b). However, because ROC functions are typically curvilinear and asymmetric, as is also the case in the present study (see below), analysis of corrected hits can serve as a first approximation only towards analysis of subjects' recognition performance.

In the second step, recognition data were therefore analyzed using a signal detection approach, which takes the curvilinear and asymmetric form of the ROC into account. We presumed unequal variance for the distribution of old and new items to account for the typically asymmetrical shape of the ROC and, thus, we described the data by applying the unequal-variance signal detection model (e.g., Dunn, 2004; Wixted, 2007a). This model bases recognition judgments upon a single source of information, i.e., the items' general memory strength, which does not necessarily imply a single underlying memory process but, for instance, may reflect the additive or nonadditive combination of familiarity and recollection codes (e.g., Kelley & Wixted, 2001; Wixted & Stretch, 2004). Whenever an item exceeds the response criterion c_i, which is related to a particular level of confidence i, but does not exceed criterion c_{i-1}, participants rate the item accordingly with i. The memory strength of old items relative to new items can be derived from the distance between the means of the underlying strength distributions of those old and the new items (d_a). Applying the model to our 5-point ROC data, it 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 consequently three degrees of freedom when testing the model's goodness of fit. For estimating the model parameters, we adopted maximum-likelihood methods that could be used for statistical testing as well.

We firstly analyzed whether the unequal-variance signal detection model described the data sufficiently for each item type and practice condition. Then, in order to reveal potential beneficial and detrimental effects of practice, it was tested whether memory strength d_a varied significantly across item types and practice conditions. In particular, we examined whether d_a for practiced items (*crp+*, *ncrp+*) exceeded d_a for the corresponding control items (*c+*), and whether d_a for unpracticed items from practiced categories (*crp*−, *ncrp*−) was lower than for the respective controls (*c*−). If such statistical analysis provided significant results, differences between practice conditions were tested as well, again using likelihood-ratio testing. We also examined whether the model's other parameters varied across item type.

Results

Practice phase

In the competitive retrieval condition, participants recalled 47.5% of the items on the first practice cycle and 69.5% of the items in total. In the noncompetitive retrieval condition, 94.2% of the category labels were retrieved correctly on the first practice cycle and 97.0% of the labels in total. Success rates in the noncompetitive condition thus were highly similar to the ones reported in the previous work by Raaijmakers and Jakab (2012) and Grundgeiger (2014).[2]

[2] Raaijmakers and Jakab (2012) reported success rates of 96.5% on the first cycle and 98.5% in total in their Experiment 1, and of 97% on the first cycle and 99% in total in their Experiment 2. Grundgeiger (2014) reported success rates of 90.6% on the first cycle and 96.6% in total for the noncompetitive retrieval condition of his Experiment 2a.

Table 1
Hit rates, false alarm rates, and corrected hit rates for Experiment 1.

Item type		Response criteria				
		"1"	"2"	"3"	"4"	"5"
crp+	Hits	.996	.998	1.000	1.000	1.000
	False alarms	.033	.060	.114	.230	.479
	Corrected hits	.963	.938	.886	.770	.521
ncrp+	Hits	.942	.974	.983	.991	1.000
	False alarms	.022	.057	.111	.238	.491
	Corrected hits	.920	.917	.872	.753	.509
c+	Hits	.650	.725	.778	.831	.929
	False alarms	.027	.058	.116	.246	.513
	Corrected hits	.623	.667	.662	.585	.416
crp−	Hits	.584	.674	.774	.825	.908
	False alarms	.033	.060	.114	.230	.479
	Corrected hits	.551	.614	.660	.595	.429
ncrp−	Hits	.637	.729	.786	.863	.930
	False alarms	.022	.057	.111	.238	.491
	Corrected hits	.615	.672	.675	.625	.439
c−	Hits	.654	.729	.789	.870	.929
	False alarms	.027	.058	.116	.246	.513
	Corrected hits	.627	.671	.673	.624	.416

Note: (Corrected) hit and false alarm 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.

Recognition test: ANOVA of corrected hits

Table 1 shows mean hit rates, false alarm rates, and corrected hit rates, separately for the five response criteria and the single item types. We conducted ANOVAs to analyze for the three most conservative ("old") response criteria whether corrected hits varied with item type. Regarding the effects of retrieval practice on the practiced items (*crp+*, *ncrp+*) 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, $F(1,58) = 95.457$, $MSE = 0.072$, $p < .001$, with higher corrected hits for the practiced than the control items. Although this effect varied with response criterion, $F(2,116) = 17.824$, $MSE = 0.006$, $p < .001$, it was present for all three criteria, $ts(58) > 7.225$, $ps < .001$.[3] Contrasting *ncrp+* and *c+* items, an analogous analysis led to a similar picture, showing a main effect of ITEM TYPE $F(1,58) = 89.110$, $MSE = 0.063$, $p < .001$, with higher corrected hits for the practiced than the control items. Again, this effect varied with response criterion, $F(2,116) = 7.726$, $MSE = 0.007$, $p = .001$, but arose for all three criteria, $ts(58) > 7.462$, $ps < .001$. Corrected hits for *crp+* items were higher than for *ncrp+* items, $F(1,58) = 4.665$, $MSE = 0.013$, $p = .035$, indicating that competitive retrieval practice enhanced recognition more than noncompetitive retrieval practice.

Regarding the effects of retrieval practice on the unpracticed items (*crp−*, *ncrp−*) relative to their controls (*c−*), a 2×3 ANOVA with the factors of ITEM TYPE (*crp−*, *c−*) and RESPONSE CRITERION ("1", "2", "3") showed no main effect of ITEM TYPE, $F(1,58) = 2.666$, $MSE = 0.079$, $p = .108$, but an interaction between the two factors, $F(2,116) = 4.376$, $MSE = 0.007$, $p = .015$. Follow-up tests revealed significantly lower corrected hits for *crp−* than *c−* items for criterion "1", $t(58) = 2.242$, $p = .029$, a marginal significant difference for criterion "2", $t(58) = 1.815$, $p = .075$, and no significant difference for criterion "3", $t(58) < 1$. Regarding 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, $F(1,58) = 2.461$, $MSE = 0.073$, $p = .122$, but an interaction between the two factors, $F(2,116) = 3.627$, $MSE = 0.006$, $p = .030$. Follow-up tests revealed significantly lower corrected hits for *crp−* than *ncrp−* items for criterion "1", $t(58) = 2.102$, $p = .040$, a marginal significant difference for criterion "2", $t(58) = 1.810$, $p = .076$, and no significant difference for criterion "3", $t(58) < 1$. 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

In the second step, we employed the unequal-variance signal detection model to analyze the data, which takes the curvilinear and asymmetric form of the ROC into account. Figs. 1a and 2a 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 counterparts.

The unequal-variance signal detection model provided a good fit to the recognition data of the two types of practiced items (*crp+*, *ncrp+*) and their corresponding controls (*c+*), all $\chi^2 s(3) < 1.869$, $ps > .599$. 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

[3] In this ANOVA, and all forthcoming related ANOVAs of Experiments 1–3, there was also a main effect of RESPONSE CRITERION. However, because this effect is kind of trivial since ROC functions are cumulative (see Method above), we do not report detailed results on it.

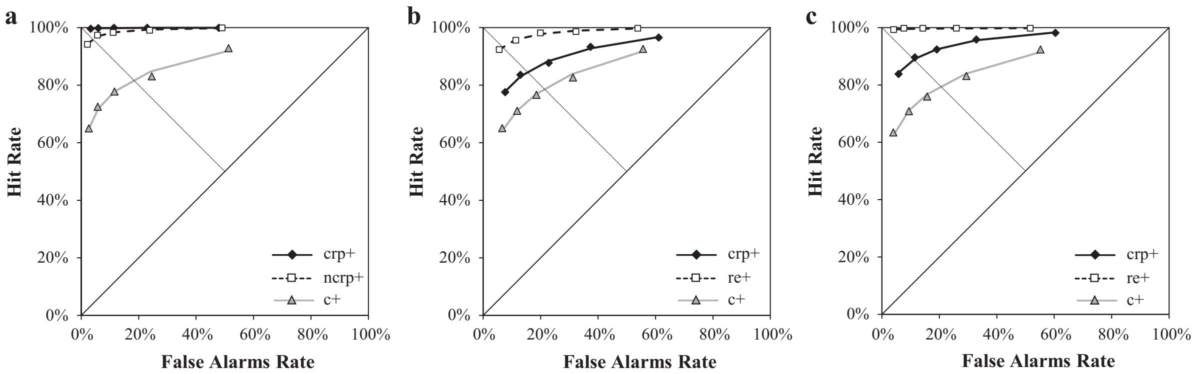

Fig. 1. Item recognition Receiver Operating Characteristics (ROCs) depicting cumulative hit and false alarm rates for the practiced item types and control items. Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model. (a) ROCs for competitively retrieval practiced items (*crp+*), noncompetitively retrieval practiced items (*ncrp+*), and control items (*c+*) of Experiment 1. (b) ROCs for competitively retrieval practiced items (*crp+*), reexposed and rated items (*re+*), and control items (*c+*) of Experiment 2. (c) ROCs for competitively retrieval practiced items (*crp+*), reexposed and visualized items (*re+*), and control items (*c+*) of Experiment 3.

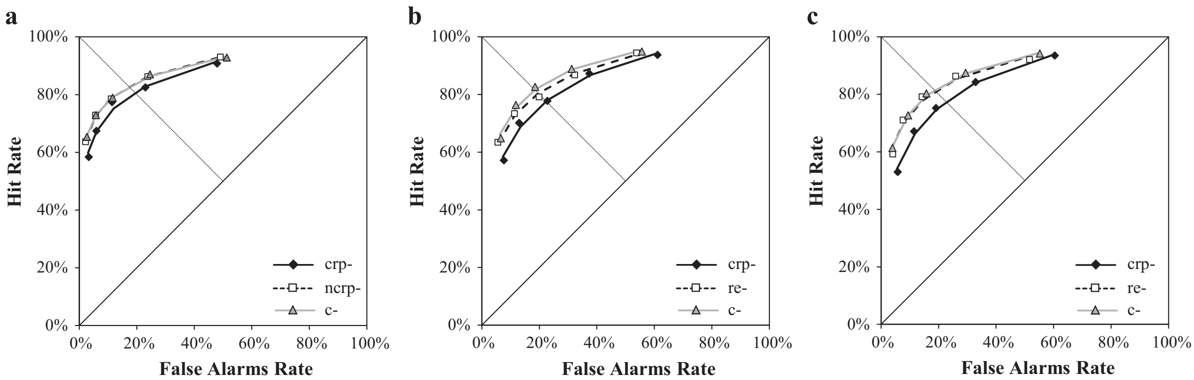

Fig. 2. Item recognition Receiver Operating Characteristics (ROCs) depicting cumulative hit and false alarm rates for the unpracticed item types of practiced categories and control items. Solid lines indicate theoretical ROCs predicted by the unequal-variance signal detection model. (a) ROCs for unpracticed items of competitively retrieval practiced categories (*crp*−), unpracticed items of noncompetitively retrieval practiced categories (*ncrp*−), and control items (*c*−) of Experiment 1. (b) ROCs for unpracticed items of competitively retrieval practiced categories (*crp*−), unpracticed items of reexposed and rated categories (*re*−), and control items (*c*−) of Experiment 2. (c) ROCs for unpracticed items of competitively retrieval practiced categories (*crp*−), unpracticed items of reexposed and visualized categories (*re*−), and control items (*c*−) of Experiment 3.

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

Item type	Parameter estimates		Goodness of fit		
	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 ($p < .05$).

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 1 and the reexposure condition of Experiment 3, 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 1, 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%.

The unequal-variance signal detection model also provided a good fit to the recognition data of the two types of unpracticed items (*crp*−, *ncrp*−), and their corresponding controls (*c*−), all $\chi^2 s(3) < 2.878,\ ps > .410$. Relative to the *c*− items, retrieval practice reduced d_a for the unpracticed (*crp*−) items in the competitive retrieval practice condition, $\chi^2(1) = 4.386,\ p = .036$, but did not affect d_a for the unpracticed (*ncrp*−) items in the noncompetitive retrieval practice condition, $\chi^2(1) = 0.016,\ p = .899$. 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, $\chi^2 s(10) > 6.402,\ ps < .049$. The order in which participants completed competitive and noncompetitive retrieval practice cycles in the practice phase of this experiment did not influence the results. Corresponding comparisons for each single item type revealed no significant effects in parameters, all $ps > .187$.

Discussion

Consistent with the results from previous item recognition studies (e.g., Hicks & Starns, 2004; Spitzer & Bäuml, 2007), we found competitive retrieval practice to enhance recognition of the practiced items but to reduce recognition of the unpracticed items. We also found retrieval practice to enhance recognition of reexposed items in the noncompetitive retrieval condition. In contrast to the competitive retrieval practice condition, however, noncompetitive retrieval practice did not induce forgetting of the unpracticed items. The findings arose from both analysis of corrected hits and signal detection analysis and demonstrate retrieval specificity of RIF, with forgetting being present after competitive retrieval practice but not after noncompetitive retrieval practice.

The present finding is consistent with the results of the 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 reexposed 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.

The results of Experiment 1 suggest that retrieval specificity of RIF is present in item recognition, at least when comparing the effects of competitive retrieval practice with the effects of noncompetitive retrieval practice. These results challenge strength-based accounts of RIF, indicating that the strengthening of the cue-item associations of practiced items may not be sufficient to induce RIF. Experiment 2 below examined whether this result generalizes to another reexposure format that, like noncompetitive retrieval practice, has been claimed to strengthen practiced items' cue-item associations.

Experiment 2

The goal of Experiment 2 was to examine whether the results for noncompetitive retrieval practice reported in Experiment 1 generalize to the restudy formats that Verde (2013) used to simulate the effects of (competitive) retrieval practice. As mentioned above, Verde (2013) employed two reexposure tasks that reduced recall of the unpracticed items: the reexposure with pleasantness rating task and the best category judgment task. We selected the reexposure-plus-pleasantness-rating task for the present Experiment 2, 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 1, we directly compared the effects of (competitive) retrieval practice with the effects of this reexposure condition, analyzing how the two types of practice affect later recognition of practiced and unpracticed items.

Method

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 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 and six lures each as well as three categories (GEMS, ALCOHOLIC DRINKS, SANITARY

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

ARTICLES) with two items each serving as buffer items were drawn from German published word norms (Mannhaupt, 1983) to create the study and recognition lists. 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 ($median = 10.5$) and so were the lures ($median = 8.5$). 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, reexposure) and ITEM TYPE (practiced, unpracticed, control). The procedure was largely identical to Experiment 1 and differed only in the intermediate practice phase. In the 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 reexposure condition, we reexposed some of the original category-exemplar pairs and asked the subjects to judge the pleasantness of the presented exemplars on a 7-point scale (1 = *not pleasant at all*, 7 = *very pleasant*). For each practice type, three types of items were generated: practiced items, i.e., retrieval practiced (*crp+*) and reexposed and rated (*re+*) items; unpracticed items of practiced categories (*crp*−, *re*−); and control items of unpracticed categories (*c+*, *c*−). Furthermore, the final test included lures that either belonged to retrieval practiced categories (*crp lures*), reexposed and rated categories (*re lures*), or control categories (*c lures*).

Statistical analysis

Statistical analysis of the data was analogous to Experiment 1.

Results

Practice phase

In the retrieval practice phase, participants successfully retrieved 67.4% of the practiced items on the first cycle and 69.3% of the items in total.

Recognition test: ANOVA of corrected hits

Table 3 shows mean hit rates, false alarm rates, and corrected hit rates, separately for the five response criteria and the single item types. Regarding the effects of retrieval practice and reexposure 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, $F(1,47) = 14.434$, $MSE = 0.049$, $p < .001$, indicating that retrieval practice was successful. Although the effect of item type varied with criterion, $F(2,94) = 3.638$, $MSE = 0.005$, $p = .030$, it arose for all three response criteria, $ts(58) > 2.652$, $ps < .012$. Similarly, contrasting *re+* and *c+* items, we found a main effect of ITEM TYPE, $F(1,47) = 71.237$, $MSE = 0.061$, $p < .001$, that varied with response criterion, $F(2,94) = 7.166$, $MSE = 0.006$, $p = .001$, but was present for all three response criteria, $ts(47) > 6.919$, $ps < .001$. Whereas both retrieval practice and reexposure were thus successful in enhancing recognition of the practiced items, corrected hits for *re+* items were higher than for *crp+* items, $F(1,47) = 29.466$, $MSE = 0.052$, $p < .001$, indicating that reexposure enhanced recognition more than competitive retrieval practice.

Table 3
Hit rates, false alarm rates, and corrected hit rates for Experiment 2.

Item type		Response criteria				
		“1”	“2”	“3”	“4”	“5”
crp+	Hits	.776	.836	.877	.933	.965
	False alarms	.077	.130	.228	.374	.611
	Corrected hits	.699	.706	.649	.559	.354
re+	Hits	.924	.956	.982	.986	.998
	False alarms	.056	.114	.199	.323	.538
	Corrected hits	.868	.842	.783	.663	.460
c+	Hits	.651	.711	.766	.827	.926
	False alarms	.066	.118	.185	.313	.557
	Corrected hits	.585	.593	.581	.514	.369
crp−	Hits	.572	.702	.778	.873	.938
	False alarms	.077	.130	.228	.374	.611
	Corrected hits	.495	.572	.550	.499	.327
re−	Hits	.634	.734	.792	.868	.945
	False alarms	.056	.114	.199	.323	.538
	Corrected hits	.578	.620	.593	.545	.407
c−	Hits	.648	.764	.827	.889	.949
	False alarms	.066	.118	.185	.313	.557
	Corrected hits	.582	.646	.642	.576	.392

Note: (Corrected) hit and false alarm rates are shown as a function of item type and response criterion. *crp+* = retrieval practiced items; *re+* = reexposed and rated items; *c+* = unpracticed items from unpracticed categories; *crp*− = unpracticed items from retrieval practiced categories; *re*− = unpracticed items from reexposed 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.

Regarding the 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, $F(1,47) = 10.617$, $MSE = 0.048$, $p = .002$, 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, $F(2,94) = 1.718$, $MSE = 0.007$, $p = .185$. These results indicate that competitive retrieval practice induced RIF, whereas reexposure did not induce any RIF-like forgetting. Consistently, an ANOVA contrasting *crp*− and *re*− items showed a main effect of ITEM TYPE, $F(1,47) = 4.039$, $MSE = 0.061$, $p = .050$, and no interaction between the two factors, $F(2,94) = 1.619$, $MSE = 0.007$, $p = .204$, 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, we employed the unequal-variance signal detection model to analyze hits and false alarms for the

Table 4
Unequal-variance signal detection model for Experiment 2.

Item type	Parameter estimates		Goodness of fit		
	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+ = retrieval practiced items; *re+* = reexposed and rated items; *c+* = unpracticed items from unpracticed categories; *crp−* = unpracticed items from retrieval practiced categories; *re−* = unpracticed items from reexposed 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 ($p < .05$).

single response criteria. Figs. 1b and 2b depict the ROCs for the practiced items, the unpracticed items, and the respective control items in the retrieval and reexposure 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 $\chi^2s(3) < 1.415$, $ps > .701$. 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 reexposure. Retrieval and reexposure differed marginally in their effects on d_a for the practiced items, $\chi^2(1) = 3.436$, $p = .064$, with a trend for better memory after reexposure 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 reexposure 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 reexposure. Three further parallels to Experiment 1 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, $\chi^2s(2) < 1.164$, $ps > .558$, but was larger than 1.0, $\chi^2s(1) > 31.499$, $ps < .001$. Second, the placement of the five confidence criteria varied across item type, $\chi^2s(10) > 26.480$, $ps < .003$. Third, the order in which the practice conditions were provided to the participants did not influence the results, all $ps > .403$.

Discussion

The results of Experiment 2 again replicate prior RIF work by finding retrieval practice to enhance recognition of the practiced items but to reduce recognition of the unpracticed items. Going beyond the prior work, the results show that reexposure when supplemented with a pleasantness rating task can enhance recognition of the practiced items but leaves recognition of the unpracticed items unaffected. The finding of reduced recognition of unpracticed items after retrieval practice but not after reexposure arose from both analysis of corrected hits and signal detection analysis and indicates that, like the noncompetitive retrieval practice condition in Experiment 1, reexposure supplemented with pleasantness ratings does not reduce recognition of the unpracticed items. These results support retrieval specificity of RIF in item recognition. Doing so, they challenge strength-based accounts of RIF and indicate a critical role for inhibition in RIF.

Experiment 3

Experiment 3 examined another reexposure 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 reexposed exemplar, such as the size, shape, color, or taste of the item, can reduce recall of the unpracticed items, very similar to how retrieval practice does. The goal of Experiment 3 was to investigate whether this finding generalizes from recall to item recognition. Analogous to Experiments 1 and 2, we directly compared the effects of (competitive) retrieval practice with the effects of such mental imagery, examining whether both types of practice reduce later recognition of unpracticed items.

Method

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

Nine categories (TOOLS, BIRDS, FLOWERS, DRINKS, FRUITS, MUSICAL INSTRUMENTS, FURNITURE, SPICES, CLOTHING) with six study items and six lures each were drawn from published word norms (Battig & Montague, 1969; Mannhaupt, 1983; Scheithe & Bäuml, 1995; Van Overschelde, Rawson, & Dunlosky, 2004) to compile the study and recognition lists. 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 as well as lures were of

[6] We substituted two categories from Saunders et al.'s original material (SPORTS EQUIPMENT, WEAPONS) with three categories from Experiment 2 and completed the remaining six categories with lures from published word norms. The replaced categories did not provide enough items (six targets, six lures) 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. material were replaced due to their absence in any of the available norms.

medium to high frequency (*median* = 6.0 and *median* = 9.5, respectively). 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, reexposure) and ITEM TYPE (practiced, unpracticed, control). The procedure was largely identical to Experiments 1 and 2 and differed only in the intermediate practice phase. In the retrieval condition, we presented the category label and the first letter of the to-be-retrieved item and asked the subjects to retrieve the corresponding exemplar within 5 s (ISI = 500 ms). Following Saunders et al. (2009), no feedback was provided, and participants completed three consecutive cycles of retrieval practice. We included a distractor task of 135 s duration (summation of three-digit numbers) after the retrieval condition to match the time frame of the two practice conditions. Analogous to the Saunders et al. experiment, in the reexposure-plus-imagery condition, participants were reexposed to the category label and the item for 10 s (ISI = 500 ms) 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−*). Furthermore, the final test included lures that either belonged to retrieval practiced categories (*crp lures*), reexposed and visualized categories (*re lures*), or control categories (*c lures*).

Statistical analysis

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

Results

Practice phase

Success rates in the retrieval practice phase were 66.0% in the first cycle and 69.7% in total.

Recognition test: ANOVA of corrected hits

Table 5 shows mean hit rates, false alarm rates, and corrected hit rates, separately for the five response criteria and the single item types. Regarding the effects of retrieval practice and reexposure 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, $F(1,47) = 27.519$, $MSE = 0.068$, $p < .001$, indicating that retrieval practice was successful. Like in Experiments 1 and 2, the effect varied with criterion, $F(2,94) = 4.537$, $MSE = 0.004$, $p = .013$, but was present for all three criteria, all $ts > 4.103$, all $ps < .001$. An analogous analysis contrasting *re+* and *c+* items also showed a main effect of ITEM TYPE, $F(1,47) = 136.946$, $MSE = 0.049$, $p < .001$, and a reliable interaction between the two factors, $F(2,94) = 13.220$, $MSE = 0.005$, $p < .001$. Again, however, the practice effect arose for all three criteria, all $ts > 9.395$, all $ps < .001$. Corrected hits for *re+* items were higher than for *crp+* items, $F(1,47) = 39.239$, $MSE = 0.039$, $p < .001$, indicating that reexposure enhanced recognition more than competitive retrieval practice.

Regarding the 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, $F(1,47) = 7.283$, $MSE = 0.076$, $p = .010$, 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

Table 5
Hit rates, false alarm rates, and corrected hit rates for Experiment 3.

Item type		Response criteria				
		"1"	"2"	"3"	"4"	"5"
crp+	Hits	.838	.896	.924	.958	.982
	False alarms	.058	.115	.191	.329	.604
	Corrected hits	.780	.781	.733	.629	.378
re+	Hits	.993	.998	.998	.998	.998
	False alarms	.041	.077	.142	.259	.516
	Corrected hits	.952	.921	.856	.739	.482
c+	Hits	.634	.709	.759	.831	.924
	False alarms	.040	.094	.158	.294	.552
	Corrected hits	.594	.615	.601	.537	.372
crp−	Hits	.530	.671	.752	.843	.935
	False alarms	.058	.115	.191	.329	.604
	Corrected hits	.472	.556	.561	.514	.331
re−	Hits	.593	.711	.792	.864	.921
	False alarms	.041	.077	.142	.259	.516
	Corrected hits	.552	.634	.650	.605	.405
c−	Hits	.614	.727	.803	.875	.942
	False alarms	.040	.094	.158	.294	.552
	Corrected hits	.574	.633	.645	.581	.390

Note: (Corrected) hit and false alarm rates are shown as a function of item type and response criterion. *crp+* = retrieval practiced items; *re+* = reexposed and visualized items; *c+* = unpracticed items from unpracticed categories; *crp−* = unpracticed items from retrieval practiced categories; *re−* = unpracticed items from reexposed 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.

[7] Saunders et al. (2009) employed a stronger version of mental imagery than we did in the present experiment. 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 our 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 reexposure 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, we therefore employed a weaker version of mental imagery using only three cycles of practice, a constant feature to visualize, and no distractors between cycles. As can be seen in the Results section below, even with this reduced version, reexposure supplemented with mental imagery led to higher recognition of practiced items than retrieval practice did.

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

Item type	Parameter estimates		Goodness of fit		
	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+ = retrieval practiced items; *re+* = reexposed and visualized items; *c+* = unpracticed items from unpracticed categories; *crp−* = unpracticed items from retrieval practiced categories; *re−* = unpracticed items from reexposed 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 ($p < .05$).

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 reexposure did not induce any RIF-like forgetting. Consistently, there was a main effect of ITEM TYPE when contrasting *crp−* and *re−* items, $F(1,47) = 8.540$, $MSE = 0.056$, $p = .005$, but no interaction between ITEM TYPE and RESPONSE CRITERIA, $F(2,94) < 1$. 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

In the next step, we employed the unequal-variance signal detection model to analyze hits and false alarms for the single response criteria. The ROCs in Figs. 1c and 2c 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.

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 reexposed items gained more strength through practice than retrieval practiced items. Memory strength of *c−* items significantly exceeded memory strength of *crp−* items, $\chi^2(1) = 11.004$, $p < .001$, but did not exceed memory strength 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 after reexposure of the practiced items. Like in Experiments 1 and 2, (i) 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, $\chi^2s(1) > 52.488$, $ps < .001$, (ii) the placement of the five confidence criteria varied across item type, $\chi^2s(10) > 42.514$, $ps < .001$, and (iii) the order in which participants were provided with retrieval practice and reexposure in the practice phase did not affect the results, all $ps > .069$.

Discussion

Like in Experiments 1 and 2, we found retrieval practice to improve recognition of practiced items but to reduce recognition of unpracticed items. Analogous to Experiments 1 and 2, we also found reexposure supplemented with mental imagery to improve recognition of practiced items but to leave recognition of unpracticed items unaffected. Like the results of Experiments 1 and 2, these findings suggest forgetting of unpracticed items after retrieval practice but not after reexposure supplemented with mental imagery. The findings thus indicate retrieval specificity of RIF in item recognition, challenging strength-based accounts of RIF and supporting the view of a critical role of inhibition in RIF.

Strength-based accounts of RIF and inhibition also differ in the degree to which they can explain the pattern of strengthening of practiced items and forgetting of unpracticed items across Experiments 1–3. Indeed, in Experiment 1, competitive retrieval practice led to more strengthening of practiced items than did noncompetitive retrieval practice, raising the possibility that the greater strengthening of the practiced items in the competitive condition was at the core of the specificity finding. However, in Experiments 2 and 3, the strengthening effect was reversed and reexposure supplemented with pleasantness ratings and reexposure supplemented with mental imagery induced more strengthening of the practiced items than did competitive retrieval practice. Thus, across the three experiments, the effect of different kinds of practice on practiced items was dissociated from the effect on the unpracticed items, which also challenges strength-based accounts of RIF but is consistent with inhibition.

Experiments 1–3 were designed in a way that the materials and study and practice procedures were highly similar to the ones employed in the previous recall studies by Grundgeiger (2014), Raaijmakers and Jakab (2012), Saunders et al. (2009), and Verde (2013). This was done to largely exclude the possibility that a difference in results – the finding of reexposure-induced forgetting in the previous recall studies versus the finding of no reexposure-induced forgetting in the present item recognition study – would be due to differences in these methodological aspects. To completely rule out that differences in materials and study and practice procedures mediated the difference in results between studies, we ran another set of experiments (Experiments 4–6) to examine whether the same materials and study and practice procedures that were employed in Experiments 1–3 induced both RIF and RIF-like forgetting when using recall at test. Such finding would replicate the previous results by Grundgeiger (2014), Raaijmakers and Jakab (2012), Saunders et al. (2009), and Verde (2013). In particular, it would provide direct evidence that (competitive) retrieval practice induces RIF in both recall and item recognition, whereas the reexposure formats induce RIF-like forgetting in recall but not in item recognition.

Experiment 4

Raaijmakers and Jakab (2012) and Grundgeiger (2014) found that both competitive and noncompetitive retrieval practice can induce RIF when using recall as the final memory test. The goal of Experiment 4 was to replicate this finding with the same materials and study and practice procedures that were employed in Experiment 1 but usage of recall at test. On the basis of the two previous studies, we expected to find RIF after both competitive and noncompetitive retrieval practice.

Method

Participants

Thirty-six students of Regensburg University participated in the experiment (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

We used the same materials as in Experiment 1. However, substituting the recognition test for a cued recall test rendered the lures unnecessary, therefore only the target items of Experiment 1 were used as stimuli.

Design and procedure

Like Experiment 1, Experiment 4 had a 2×3 design with PRACTICE TYPE (competitive retrieval, noncompetitive retrieval) and ITEM TYPE (practiced, unpracticed, control) as within-subject factors.

Apart from the format of the final memory test, the same procedure was employed as in Experiment 1. A cued recall procedure was used at test. The category label and initial letter of each exemplar were provided as cues (e.g., INSECT – *t*____) for 5 s (ISI = 500 ms) and the participants were asked to generate the corresponding exemplar that they had learned in the study phase. The order of the items of the test list was blocked randomized: We arranged six blocks with nine items, each containing one member of each category. Three blocks consisted exclusively of unpracticed items from practiced categories (*crp*−, *ncrp*−) and their control counterparts (*c*−), the remaining three blocks were compiled of *crp*+, *ncrp*+, and their control counterparts (*c*+). Items within blocks and the blocks themselves were drawn randomly with the only restriction that the blocks compiled of *crp*−, *ncrp*−, and *c*− items were presented in the first half of the test, and the blocks compiled of *crp*+, *ncrp*+, and *c*+ items were presented in the second half, which was done to rule out output interference effects for the unpracticed items. Three buffer items were tested at the beginning of the cued recall test.

Results

Practice phase

In the competitive retrieval condition, participants recalled 54.6% of the items on the first practice cycle and 75.3% of the items in total. In the noncompetitive retrieval condition, participants recalled 94.4% of the category labels on the first practice cycle and 97.2% of the labels in total. All of these numbers are highly similar to those reported in Experiment 1.

Recall test

Figs. 3a and 4a show the percentages of correctly recalled practiced items, unpracticed items, and control items in the competitive and noncompetitive retrieval conditions. Regarding the beneficial effects of competitive and noncompetitive retrieval practice, participants recalled 86.7% of the *crp*+ items, 63.6% of the *ncrp*+ items, and 37.4% of the *c*+ items. Recall levels differed significantly across item type, $F(2, 70) = 113.298$, $MSE = .019$, $p < .001$. Planned comparisons showed significant recall enhancement of both *crp*+ and *ncrp*+ items relative to the *c*+ items, $t(35) = 15.707$, $p < .001$, and $t(35) = 7.517$, $p < .001$, respectively, and reliably higher recall for *crp*+ than *ncrp*+ items, $t(35) = 7.225$, $p < .001$. Thus, both types

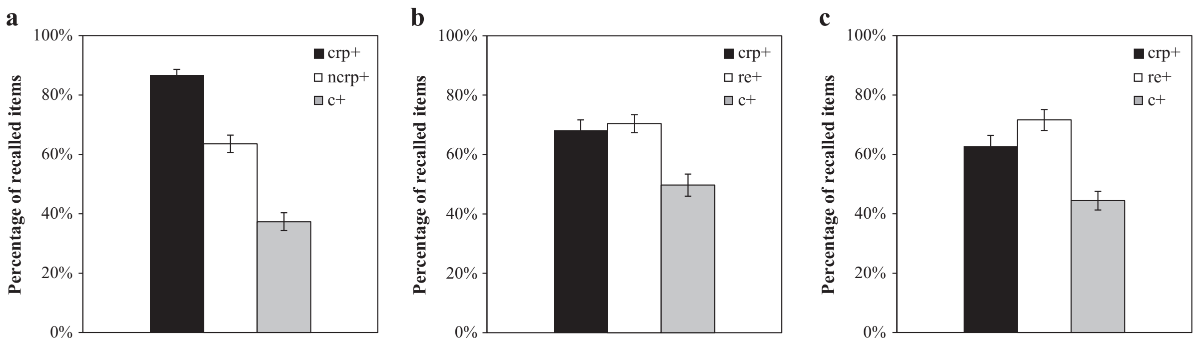

Fig. 3. Recall percentages for practiced item types and control items. Error bars represent standard errors. (a) Recall percentages for competitively retrieval practiced items (*crp*+), noncompetitively retrieval practiced items (*ncrp*+), and control items (*c*+) of Experiment 4. (b) Recall percentages for competitively retrieval practiced items (*crp*+), reexposed and rated items (*re*+), and control items (*c*+) of Experiment 5. (c) Recall percentages for competitively retrieval practiced items (*crp*+), reexposed and visualized items (*re*+), and control items (*c*+) of Experiment 6.

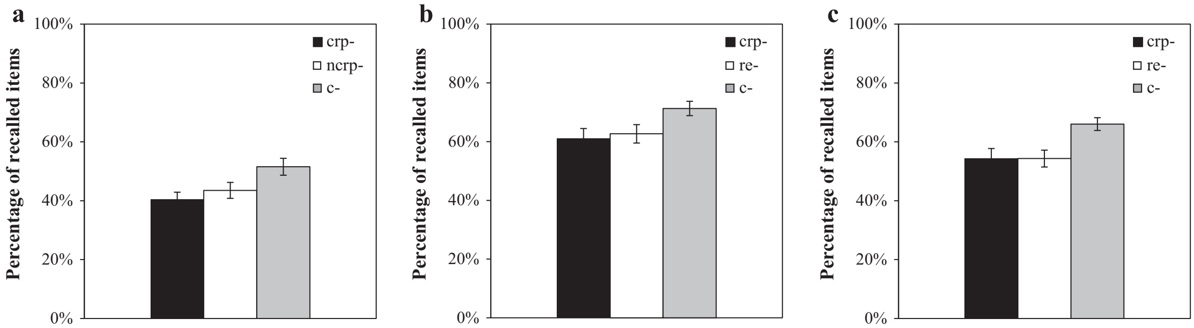

Fig. 4. Recall percentages for the unpracticed item types of practiced categories and control items. Error bars represent standard errors. (a) Recall percentages for unpracticed items of competitively retrieval practiced categories (*crp*−), unpracticed items of noncompetitively retrieval practiced categories (*ncrp*−), and control items (*c*−) of Experiment 4. (b) Recall percentages for unpracticed items of competitively retrieval practiced categories (*crp*−), unpracticed items of reexposed and rated categories (*re*−), and control items (*c*−) of Experiment 5. (c) Recall percentages for unpracticed items of competitively retrieval practiced categories (*crp*−), unpracticed items of reexposed and visualized categories (*re*−), and control items (*c*−) of Experiment 6.

of retrieval practice facilitated recall of practiced items, with competitive retrieval practice inducing a higher level of facilitation than noncompetitive retrieval practice. Intrusion rates for the three types of 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. Rates differed significantly across item type, $F(2,70) = 8.980$, $MSE = .004$, $p < .001$, indicating that the enhancement effect for the practiced items, when related to the intrusion rates, may have been slightly underestimated.

Regarding the detrimental effects of practice, participants retrieved 40.4% of the *crp*− items, 43.5% of the *ncrp*− items, and 51.5% of the *c*− items. Again, recall levels differed significantly across item type, $F(2,70) = 6.059$, $MSE = .020$, $p = .004$. Planned comparisons showed reliable forgetting of the unpracticed items following both competitive and noncompetitive retrieval practice, $t(35) = -3.550$, $p = .001$, and $t(35) = -2.548$, $p = .015$. Recall rates for *crp*− and *ncrp*− items did not differ significantly, $t(35) = -0.861$, $p = .395$. 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. Rates did not differ across item type, $F(2,70) = 1.430$, $MSE = .013$, $p = .246$.

A correlation analysis across participants revealed that enhancement in the competitive retrieval condition – defined as mean recall of *crp+* items minus mean recall of *c+* items – was not correlated to RIF – defined as mean recall of *c*− items minus mean recall of *crp*− items -, $r = .02$, $p = .908$, which replicates prior work (e.g., Hanslmayr et al., 2010; Hulbert et al., 2012; Staudigl et al., 2010). Similarly, there was no reliable correlation between enhancement in the noncompetitive retrieval condition – defined as mean recall of *ncrp+* items minus mean recall of *c+* items – and RIF-like forgetting – defined as mean recall of *c*− items minus mean recall of *ncrp*− items -, $r = -.08$, $p = .665$.

Discussion

The results show that both competitive and noncompetitive retrieval practice enhanced recall of the practiced items but reduced recall of the unpracticed items. The finding of RIF after both forms of retrieval practice replicates the previous recall results by Raaijmakers and Jakab (2012) and Grundgeiger (2014). Importantly, because exactly the same materials and study and practice procedures were employed as in Experiment 1, the results of Experiments 1 and 4 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.

Experiment 5

Verde (2013) reported that not only noncompetitive retrieval practice but also reexposure supplemented with pleasantness ratings can reduce recall of the not reexposed items. The goal of Experiment 5 was to replicate this finding with the same materials and study and practice procedures that were employed in Experiment 2 but usage of a recall test. On the basis of Verde's previous study, we expected to find RIF after (competitive) retrieval practice and RIF-like forgetting after reexposure supplemented with pleasantness ratings.

Method

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 material was identical to the one employed in Experiment 2 but omitting the lures.

Design and procedure

Apart from the format of the final memory test, we employed the same design and procedure as used in Experiment 2. The procedure in the recall test was identical to the one employed in Experiment 4.

Results

Practice phase

In the competitive retrieval condition, participants correctly retrieved 67.0% of the items on the first practice cycle and 68.7% of the items in total. These numbers are highly similar to those reported in Experiment 2.

Recall test

Percentages of correctly recalled practiced and unpracticed items in the retrieval practice and reexposure conditions and of the corresponding control items are depicted in Figs. 3b and 4b. Regarding the beneficial effects of practice, recall rates mounted up to 68.2% for the *crp+* items, 70.4% for the *re+* items, and 49.7% for the *c+* items. Recall levels differed significantly across item type, $F(2, 70) = 20.934$, $MSE = .022$, $p < .001$. Planned comparisons showed significant recall enhancement for both *crp+* and *re+* items when compared to *c+* items, $t(35) = 4.965$, $p < .001$, and $t(35) = 5.462$, $p < .001$. Recall performance of *crp+* and *re+* items did not differ reliably, $t(35) = -0.729$, $p = .471$. 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% for the *crp−* items, 62.7% for the *re−* items, and 71.3% for the *c−* items were observed. Recall rates varied across item type, $F(2, 70) = 5.504$, $MSE = .020$, $p = .006$. Planned comparisons revealed significant forgetting of *crp−* relative to *c−* items, $t(35) = -3.274$, $p = .002$, and significant forgetting of *re−* relative to *c−* items, $t(35) = -3.682$, $p = .001$. The numerical difference between *crp−* and *re−* items was not significant, $t(35) = -0.367$, $p = .716$. 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, $F(2, 70) = 1.170$, $MSE = .007$, $p = .317$. Analogous to Experiment 4, there was no reliable correlation between the enhancement effect in the competitive retrieval condition and RIF, $r = -.14$, $p = .404$, and there was no reliable correlation between the enhancement effect in the reexposure condition and RIF-like forgetting, $r = .14$, $p = .408$.

Discussion

The results demonstrate that both (competitive) retrieval practice and reexposure with pleasantness ratings enhanced recall of the practiced items but reduced recall of the unpracticed items. The finding of RIF after both practice conditions replicates the previous recall result by Verde (2013). Crucially, because exactly the same materials and study and practice procedures were employed as in Experiment 2, the results of Experiments 2 and 5 indicate that (competitive) retrieval practice reduces both recall and recognition of unpracticed items, whereas reexposure with pleasantness ratings reduces recall but not recognition of these items.

Experiment 6

Saunders et al. (2009) showed that imagining particular attributes of reexposed category exemplars during practice can reduce recall of the not reexposed items at test. The goal of Experiment 6 was to replicate this finding with the same materials and study and practice procedures that were employed in Experiment 3 but usage of recall at test. On the basis of the previous study, and analogous to Experiments 4 and 5, we expected to find RIF after (competitive) retrieval practice and RIF-like forgetting after reexposure supplemented with mental imagery.

Method

Participants

We tested 36 students of Regensburg University (*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

We used the target material of Experiment 3.

Design and procedure

Design and procedure of Experiment 6 were identical to Experiment 3 with the only exception that the final recognition test was replaced by a cued recall procedure. The procedure in the cued recall test was identical to the ones employed in Experiments 4 and 5.

Results

Practice phase

Recall performance in the (competitive) retrieval practice condition mounted up to 60.5% of the items on the first practice cycle, and to 63.7% of the items in total. These numbers are similar to those reported in Experiment 3.

Recall test

Figs. 3c and 4c show percentages of correctly recalled practiced and unpracticed items in the retrieval practice and reexposure conditions, and of the corresponding control items. Regarding the beneficial effects of practice, participants recalled on average 62.7% of the *crp+* items, 71.6% of the *re+* items, and 44.4% of the *c+* items. Recall levels differed significantly across item type, $F(2, 70) = 22.099$, $MSE = .031$, $p < .001$. When compared to the *c+* items, retrieval practice facilitated recall of both the *crp+* and the *re+* items, $t(35) = 4.595$, $p < .001$, and $t(35) = 6.215$, $p < .001$. Mental imagery boosted recall of the practiced items significantly more than retrieval practice did, $t(35) = -2.158$, $p = .038$. 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, $F(2, 70) = 1.99$, $MSE = .005$, $p = .144$.

Regarding the detrimental effects of practice, recall rates for *crp−* items, *re−* items, and *c−* items reached 54.3%, 54.3%, and 66.1%, respectively. Recall levels varied significantly across item type, $F(2, 70) = 9.304$,

$MSE = .018$, $p < .001$. Compared to *c*− items, both *crp*− and *re*− items showed a reliable reduction in recall, $t(35) = -4.245$, $p < .001$, and $t(35) = -3.952$, $p < .001$, 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, $F(2, 70) < 1$. There was no reliable correlation between the enhancement effect in the competitive retrieval condition and RIF, $r = .31$, $p = .065$, and there was no reliable correlation between the enhancement effect in the reexposure condition and RIF-like forgetting, $r = .22$, $p = .197$.[8]

Discussion

Both (competitive) retrieval practice and reexposure with mental imagery enhanced recall of the practiced items but reduced recall of the unpracticed items. The observed forgetting after both practice conditions replicates the previous recall result by Saunders et al. (2009). Critically, because exactly the same materials and study and practice procedures were employed as in Experiment 3, these results indicate that (competitive) retrieval practice reduces both recall and recognition of unpracticed items, whereas reexposure with mental imagery reduces recall but not recognition of these items.

General discussion

The main goal of the present study was to examine retrieval specificity of RIF in item recognition. Using recall at test and reexposure formats that supposedly lead to stronger cue-item associations than standard restudy does, several recent studies reported RIF-like forgetting after such reexposure formats, indicating that not only standard (competitive) retrieval practice but also reexposure of some studied items can lead to forgetting of the unpracticed items (Grundgeiger, 2014; Raaijmakers & Jakab, 2012; Saunders et al., 2009; Verde, 2013). Employing noncompetitive retrieval practice, reexposure supplemented with pleasantness ratings, and reexposure supplemented with mental imagery, the results of the present Experiments 4–6 replicate these findings, with each single experiment showing equivalent recall impairment for unpracticed items after (competitive) retrieval practice and one of the three reexposure formats. All of these findings challenge retrieval specificity, indicating that the strengthening of the cue-item associations of the practiced items may be sufficient to induce RIF and active retrieval may not be necessary to induce the effect.

Going beyond the prior work, the present Experiments 1–3 investigated whether the same reexposure formats that can induce RIF-like forgetting in recall tests also cause RIF-like forgetting in item recognition. Again noncompetitive retrieval practice, reexposure supplemented with pleasantness ratings, and reexposure supplemented with mental imagery were employed as reexposure formats, and exactly the same material and the same study and practice procedures were used as in Experiments 4–6. All three experiments reported improved recognition of practiced items and impaired recognition of unpracticed items after (competitive) retrieval practice, thus replicating the results of the many previous studies that reported RIF in item recognition (e.g., Hicks & Starns, 2004; Román et al., 2009; Spitzer & Bäuml, 2007). In contrast, however, the three experiments reported improved recognition of practiced items but unaffected recognition of unpracticed items after each of the three reexposure formats. This pattern of results indicates retrieval specificity of RIF in item recognition, demonstrating that (competitive) retrieval practice, but not reexposure of studied items, is sufficient to reduce recognition of unpracticed items. Together, the results of the six experiments suggest that whether retrieval specificity of RIF holds or not, can depend on the employed memory test, and that results on retrieval specificity of RIF in recall do not simply generalize to results on retrieval specificity of RIF in item recognition.

Implications for the strength-based account and the inhibitory account of RIF

According to strength-based accounts of RIF, RIF arises because (competitive) retrieval practice strengthens the cue-item associations of the practiced items and these strengthened associations lead to blocking for the unpracticed items at test, thus inducing RIF. The same accounts also assume that some reexposure formats strengthen cue-item associations similarly to how (competitive) retrieval does, thus also inducing forgetting of the unpracticed items at test (Raaijmakers & Jakab, 2012, 2013; Verde, 2013). Following this rationale, RIF should not be retrieval specific. In particular, failures to find retrieval specificity should not be restricted to certain recall formats but arise over a wide range of memory tests. By showing retrieval specificity in item recognition the present results challenge these strength-based accounts of RIF. They indicate that the strengthening of the cue-item associations of the practiced items without active retrieval of the items is not sufficient to cause RIF in item recognition, and retrieval of the practiced items is necessary to induce this effect.

Raaijmakers and Jakab (2012) suggested a variant of strength-based accounts, which assumes that retrieval practice is a much more effective way of strengthening memory traces than noncompetitive retrieval practice or other reexposure formats. Accordingly, competitive retrieval practice should induce a higher level of strengthening of cue-item associations than the reexposure formats and thus induce a higher amount of interference at test, i.e., a higher amount of RIF. At least four points challenge this proposal, however. First, because a high level of strengthening of practiced items may not only increase the items' interference potential but may also improve differentiation of the unpracticed items during item recognition (e.g., Shiffrin & Steyvers, 1997), it remains unclear why RIF in

[8] When pooling the recall data of Experiments 4–6, separately for the three (competitive) retrieval practice conditions and the three reexposure conditions, a nonsignificant correlation of $r = .051$ between (competitive) retrieval-induced enhancement and RIF, and a nonsignificant correlation of $r = .104$ between reexposure-induced enhancement and RIF-like forgetting arose, $ps > .286$. The two correlations did not differ significantly, $p = .358$.

item recognition should reflect strength-based forgetting. Second, across Experiments 1–3, the effect of different kinds of practice on practiced items was dissociated from the effect on the unpracticed items, indicating that strength level of the practiced items did not mediate the RIF effect (see Discussion of Experiment 3). Third, the results of Experiment 4 show a higher level of recall enhancement for practiced items after competitive than noncompetitive retrieval practice, but the results do not show a higher amount of RIF after competitive retrieval practice (for related results, see the meta analysis of Murayama et al., 2014). Finally, if RIF was caused by the degree of strengthening of the practiced items, then amount of RIF in Experiments 4–6 may be expected to correlate with amount of retrieval-induced enhancement for the practiced items, which, however, is not the case (for comparable results, see again Murayama et al., 2014). It is therefore unclear how the suggested high levels of strengthening of the competitively retrieval practiced items should be at the core of the present RIF findings.

According to the inhibition account of RIF, RIF arises because, during (competitive) retrieval practice, the related unpracticed items interfere and are inhibited to reduce the interference (Anderson et al., 1994; Anderson & Spellman, 1995). Since noncompetitive retrieval practice, reexposure supplemented with pleasantness ratings, and reexposure supplemented with mental imagery should not induce such interference during practice, and thus should also not induce inhibition, RIF should arise after (competitive) retrieval practice but not after these reexposure formats and therefore be retrieval specific, at least in tests that supposedly reduce or eliminate strength-based interference effects, like item recognition (e.g., Ratcliff et al., 1990; Shiffrin et al., 1990; see also Kinnell & Dennis, 2011). The present finding that, in each single recognition experiment, forgetting of unpracticed items arose after (competitive) retrieval practice but not after any of the reexposure formats, indicates retrieval specificity of RIF and thus suggests a critical role of inhibition in RIF. More generally, the finding suggests that the strengthening of cue-item associations through reexposure does not induce the same processes as (competitive) retrieval practice. Rather, reexposure-induced strengthening effects in the retrieval practice paradigm may be more similar to strengthening effects in the list strength paradigm.

A two-factor account of RIF

Although results on RIF during the past two decades have mostly been evaluated with regard to the question of whether they were mediated by inhibition or blocking, there has also been research suggesting that both inhibition and blocking may contribute to RIF (Anderson & Levy, 2007; Aslan & Bäuml, 2010; Bäuml, 2008; Grundgeiger, 2014; Schilling, Storm, & Anderson, 2014; Storm & Levy, 2012; see also Anderson et al., 1994, p. 1080). According to such a two-factor account of RIF, inhibition operates during retrieval practice: it reduces the memory strength of not-to-be-practiced interfering items and thus reduces these items' accessibility on a later memory test. Such inhibition-based forgetting is supposed to be retrieval specific and to arise over a wide range of memory tests. In addition to inhibition, the strengthening of the cue-item associations of the practiced items during retrieval practice may lead to blocking at test and induce forgetting of the unpracticed items. Such blocking should play a role primarily in tests in which item-specific cues are absent and be gradually reduced as item-specific cues are included and the strength of these cues is increased. The relative contribution of blocking to RIF should therefore follow a gradient and, for categorized material, be larger in category-cued recall, in which no item-specific cues are provided, than in category-plus-stem-cued recall, in which the items' initial letters are provided as item-specific cues, and be largely absent in item recognition, in which the items themselves are presented as cues. Thus, although both inhibition and blocking may contribute to RIF in general, the relative contribution of these mechanisms may vary depending on how RIF is assessed at test.

This two-factor account of RIF is consistent with a number of findings from the literature. For instance, in their large-scale meta analysis, Murayama et al. (2014) reported evidence that retrieval-induced enhancement of practiced items correlates with RIF on category-cued recall tests, but does not correlate with RIF on tests employing item-specific cues, like category-plus-stem-cued recall or item recognition, which is in line with the view that blocking contributes to RIF mainly when no item-specific cues are provided and, in such case, degree of strengthening of the practiced items reduces recall of the unpracticed items. Similarly, Schilling et al. (2014) reported evidence that with category-plus-stem-cued recall and item recognition, in which blocking is supposedly reduced, better motor response inhibition predicts greater RIF, whereas with category-cued recall motor response inhibition has the opposite relationship with RIF. Finally, individual differences work with schizophrenic patients (Soriano, Jiménez, Román, & Bajo, 2009), ADHD patients (Storm & White, 2010), and young children (Aslan & Bäuml, 2010) showed that tests employing item-specific cues are sensitive to individual differences in inhibitory control, whereas tests employing no item-specific cues are much less so, if at all. In fact, whereas schizophrenic patients, ADHD patients, and young children failed to show RIF with category-plus-stem-cued recall or item recognition, they did show RIF when using category-cued recall.

The two-factor account can also explain the main results of the present study. Following the view that strength-based interference effects are largely absent in item recognition (e.g., Dennis & Humphreys, 2001; Shiffrin & Steyvers, 1997), the account attributes the finding of RIF in item recognition to inhibition and explains the finding of no RIF-like forgetting in item recognition through the absence of inhibition in noncompetitive retrieval practice, reexposure supplemented with pleasantness ratings, and reexposure supplemented with mental imagery (Experiments 1–3). In addition, the account attributes the finding of RIF-like forgetting in category-plus-stem-cued recall to blocking and the finding of RIF in category-plus-stem-cued recall to a combination of the effects of inhibition and blocking (Experiments 4–6). While the two-factor account can thus explain the basic pattern of results in the present experiments, two questions remain. The one question is why, in Experiments 4–6, amount of RIF was not larger than amount of RIF-like recall impairment (for comparable results, see Grundgeiger, 2014; Verde, 2013). Indeed, if both inhibition and blocking contributed to RIF, whereas blocking only contributed to RIF-like forgetting, one may expect amount of RIF to be larger than amount of RIF-like forgetting. The other question is why there was no positive relationship between reexposure-induced enhancement and RIF-like forgetting in Experiments 4–6, and the correlational results rather mimicked those for RIF. In fact, if RIF-like forgetting was caused by blocking, reexposure-induced enhancement of practiced items may correlate with amount of forgetting of the unpracticed items (see Murayama et al., 2014).

Although this study cannot provide a fully satisfying answer to these questions, some suggestions arise. Regarding the first question, the suggestion arises that, in contrast to Raaijmakers and Jakab's (2012) view, blocking effects may be larger after reexposure than (competitive) retrieval practice. Indeed, in a recent series of experiments, Kliegl and Bäuml (2015) provided evidence that retrieval practice after study can insulate items against intralist interference more than restudy does (for related results, see Abel & Bäuml, 2014; Halamish & Bjork, 2011), indicating that blocking-induced forgetting may be smaller in RIF than RIF-like forgetting. This may explain why, in recall (Experiments 4–6), the amount of RIF was not significantly larger than amount of RIF-like forgetting. Regarding the second question, the suggestion may arise that, to some degree, inhibitory processes also contribute to reexposure-induced forgetting. Reexposure may engage retrieval, either because participants recognize the practiced exemplars and recognition triggers inhibition, or participants, on the second or third practice cycle, remember their judgments from the first or second cycle, and this episodic retrieval is sufficient to induce some level of inhibition.[9] Although such view can explain why recall enhancement and forgetting were unrelated, not only after retrieval practice but also after reexposure cycles, an alternative explanation would be that inhibition did not contribute to reexposure-induced forgetting but the null relation between reexposure-induced enhancement and RIF-like forgetting arose because individual differences masked the true relation (Raaijmakers & Jakab, 2013; see also Hintzman, 1972). In such case, both inhibition and blocking may contribute to RIF, whereas RIF-like forgetting may be mediated by blocking only, which is consistent with the present two-factor account.

The suggested two-factor account of RIF together with the present finding that reexposure supplemented with mental imagery does not induce RIF in item recognition suggests that mental imagery induces RIF (primarily) by blocking and the effect is largely restricted to category-cued and category-plus-stem-cued recall tests. This view contrasts with Saunders et al.'s (2009) original explanation of the effect, who attributed it to inhibition, arguing that mental imagery generates images, which requires access to semantic knowledge in long-term memory and thus reflects a visual-based form of semantic generation. Because semantic generation of extra-list items during retrieval practice has been shown to initiate RIF of related studied items and such forgetting was attributed to inhibition (Bäuml, 2002; Storm, Bjork, Bjork, & Nestojko, 2006), imagery-induced forgetting may also reflect inhibition. If so, however, such imagery-induced forgetting should not be restricted to certain recall formats but should arise also in item recognition, which contrasts with the present results. The present results therefore suggest that reexposure supplemented with mental imagery primarily increases the strength of the cue-item associations of the reexposed items, which is not sufficient to induce RIF in item recognition.

From item recognition to independent-probe testing

The two-factor account of RIF suggests a way how to separate the inhibitory effect of retrieval practice from possible blocking effects, namely by using item recognition testing. However, this may not be the only way to separate inhibitory from noninhibitory influences. An analogous separation of inhibitory and noninhibitory effects may result from the use of independent-probe testing, i.e., the use of retrieval cues at recall that differ from the cues employed during encoding and practice (for instance, cuing the studied item *banana* by the new, independent probe MONKEY when the item was originally studied with the cue FRUIT; see Anderson, 2003). Indeed, item recognition has repeatedly been regarded a variant of independent-probe testing, with both procedures providing cues that supposedly avoid strength-based interference at test (e.g., Anderson, 2003; Ortega, Gómez-Ariza, Román, & Bajo, 2012; Román et al., 2009; Schilling et al., 2014; Spitzer & Bäuml, 2007).

However, some researchers criticized the view that independent-probe testing may avoid strength-based interference at test, arguing that blocking effects may not be absent with this type of test (e.g., Camp, Pecher, & Schmidt, 2005, 2007; Perfect et al., 2004; Raaijmakers & Jakab, 2013). The proposal has been that when participants receive an independent-probe test, they do not limit themselves to those probes but instead recall study cues covertly to aid performance, which may induce blocking. In contrast, Anderson (2003) argued that covert cuing may mask cue-independent forgetting by providing a compound cuing advantage. Weller, Anderson, Gómez-Ariza, and Bajo (2013) addressed the issue directly and found that explicitly instructing participants to engage in covert cuing during independent-probe testing decreases RIF, thus challenging the view that RIF, as measured with independent-probe testing, is the result of blocking induced by covert cuing processes. On the basis of this finding, the two-factor account of RIF predicts that item recognition and independent-probe testing behave similarly and both tests can be used to separate the inhibitory effect of retrieval practice from possible blocking effects. The results of the present Experiments 1–3 should

[9] This possibility was raised by one of the reviewers.

therefore generalize from item recognition to independent-probe testing.

Implications for the context account of RIF

The goal of the present study was to examine the inhibition and strength-based accounts of RIF by investigating the effects of certain reexposure formats on unpracticed items in item recognition. Although the present experiments were not designed to examine the more recent context account of RIF (Jonker et al., 2013), some of the results are of relevance for this account as well. According to the context account, retrieval practice induces a shift of context between study and the practice phase, thus creating two distinct experimental contexts. If category-exemplar pairs are learned in the study phase and the items' category labels are provided as retrieval cues at test, then, on the final test, the items' category labels are supposed to trigger the reinstatement of a particular context: the study context for control categories and the practice context for retrieval practiced categories. Because unpracticed items from retrieval practiced categories are not linked to the practice context, a contextual mismatch may arise for these items at test and RIF may be observed.

The context account predicts RIF whenever a retrieval process occurs during the practice phase (Jonker et al., 2013, p. 866). The account can therefore explain why both competitive and noncompetitive retrieval practice have been found to induce RIF in recall tests (Grundgeiger, 2014; Jonker & MacLeod, 2012; Raaijmakers & Jakab, 2012; present Experiment 4). However, the account cannot explain why competitive, but not noncompetitive retrieval practice induces RIF in item recognition (Grundgeiger, 2014; present Experiment 1). Jonker et al. argued that context effects should be observable in item recognition – at least if the experiment favors the use of context during the final test, as is typically the case in RIF experiments -, which could explain the well-documented finding of RIF in item recognition. If so, however, it remains unclear why noncompetitive retrieval practice does not induce RIF in item recognition. The results of Grundgeiger (2014) together with the findings of the present Experiments 1 and 4 thus challenge the context account, indicating that it cannot capture the whole range of RIF findings and cannot provide a general explanation of RIF (see also Anderson & Bjork, 1994).

Conclusions

In this series of experiments we showed that (competitive) retrieval practice impairs both recall and recognition of unpracticed items, whereas reexposure formats can impair recall but leave recognition of unpracticed items unaffected. These findings demonstrate retrieval specificity of RIF in item recognition, which challenges strength-based accounts of RIF and indicates a critical role of inhibition in RIF. Together with the results from other recent studies the findings are consistent with a two-factor account of RIF, which assigns a role for both inhibition and strength-based blocking to RIF. Whereas both inhibition and blocking may contribute to RIF in certain recall formats, only inhibition may induce RIF in item recognition.

Acknowledgments

This research is part of J. Rupprecht's dissertation. We thank M.C. Anderson, A. Criss, J.G.W. Raaijmakers, B. Spitzer, and J.T. Wixted for helpful comments and critics on a previous version of the manuscript.

References

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

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

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 265–325). San Diego, CA: Academic Press.

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

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., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 107–132). New York, NY: Psychology Press.

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, and Cognition, 37*, 264–269.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms. *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. Roediger, III (Ed.), *Cognitive psychology of memory. Learning and memory: A comprehensive reference* (Vol. 2, 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., Pastötter, B., & Hanslmayr, S. (2010). Binding and inhibition in episodic memory – Cognitive, emotional, and neural processes. *Neuroscience & Biobehavioral Reviews, 34*, 1047–1054.

Camp, G., Pecher, D., & Schmidt, H. G. (2005). Retrieval-induced forgetting in implicit memory tests: The role of test awareness. *Psychonomic Bulletin & Review, 12*, 490–494.

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, and Cognition, 33*, 950–958.

Chandler, C. C. (1989). Specific retroactive interference in modified recognition tests: Evidence for an unknown cause of interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*. 256–265.

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

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory & Language, 55*, 461–478.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452–478.
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.
Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*, 524–542.
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.
Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.
Gómez-Ariza, C. J., Lechuga, M. T., & Pelegrina, S. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition, 33*, 1431–1441.
Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and 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, and 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.
Healey, M. K., Ngo, K. W. J., & Hasher, L. (2014). Below baseline suppression of competitors during interference resolution by younger but not older adults. *Psychological Science, 25*, 145–151.
Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11*, 125–130.
Hintzman, D. L. (1972). On testing the independence of associations. *Psychological Review, 79*, 261–264.
Hockley, W. E., & Christi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition, 24*, 202–216.
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.
Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497–514.
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. (2013). Putting retrieval-induced forgetting into context: An inhibition-free, context-based account. *Psychological Review, 120*, 852–872.
Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and 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.
Kliegl, O., & Bäuml, K.-H. T. (2015). 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*. http://dx.doi.org/10.1037/xlm0000172.
Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science, 18*, 29–34.
Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). London, NJ: Lawrence Erlbaum 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.
Migueles, M., & García-Bajos, E. (2014). The role of exemplar typicality and encoding strategies in category retrieval-induced forgetting. *Memory, 22*, 212–221.
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 the 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.
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, and Cognition, 38*, 178–186.
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.
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.
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.
Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 163–178.
Román, P., Soriano, M. F., Gómez-Ariza, C. J., & Bajo, M. T. (2009). Retrieval-induced forgetting and executive control. *Psychological Science, 20*, 1053–1058.
Saunders, J., Fernandes, M., & Kosnes, L. (2009). Retrieval-induced forgetting and mental imagery. *Memory & Cognition, 37*, 819–828.
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.
Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and 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.
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, and Cognition, 33*, 863–875.
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. (2010). Overcoming fixation: Creative problem solving and retrieval-induced forgetting. *Psychological Science, 21*, 1263–1265.
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.
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-induced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 315–318.
Verde, M. F. (2013). Retrieval-induced forgetting in recall: Competitor interference revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1433–1448.
Weller, P., Anderson, M. C., Gómez-Ariza, C. J., & Bajo, M. T. (2013). On the status of cue-independence as a criterion for memory inhibition:

Evidence against the covert blocking hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1232–1245.
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., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*, 616–641.