

Retrieval-induced forgetting in a social context: Do the same mechanisms underlie forgetting in speakers and listeners?

Magdalena Abel & Karl-Heinz T. Bäuml

Memory & Cognition

ISSN 0090-502X

Volume 48

Number 1

Mem Cogn (2020) 48:1-15

DOI 10.3758/s13421-019-00957-x

Your article is protected by copyright and all rights are held exclusively by The Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Retrieval-induced forgetting in a social context: Do the same mechanisms underlie forgetting in speakers and listeners?

Magdalena Abel¹ · Karl-Heinz T. Bäuml¹

Published online: 8 July 2019
© The Psychonomic Society, Inc. 2019

Abstract

Selectively retrieving details from memory can result in forgetting related information, a finding known as retrieval-induced forgetting (RIF). The effect has mostly been examined in individuals, but RIF can also be socially transmitted and arise in listeners who are exposed to a speaker's selective memory retrieval. Whether within-individual RIF (WI-RIF) in speakers and socially shared RIF (SS-RIF) in listeners arise on the basis of the same cognitive mechanisms is unclear, however. In four experiments, we assessed both WI-RIF and SS-RIF while varying final test format to examine the potential involvement of output interference, strength-based blocking, and inhibition. WI-RIF and, to a similar degree, SS-RIF were observed on cued-recall tests with and without controlled output order at test, indicating that output interference cannot account for the observed forgetting. In contrast, SS-RIF was reduced relative to WI-RIF on tests of item recognition. These findings are consistent with the view that inhibition and blocking contribute to both WI-RIF and SS-RIF, but that the contribution of inhibition is reduced in listeners relative to speakers.

Keywords Retrieval-induced forgetting · Socially shared retrieval-induced forgetting · Social memory · Retrieval practice · Inhibition · Blocking

Imagine you meet with friends and end up reminiscing about old times and events from your shared past. Does listening to somebody else recall specific details of a previously experienced episode affect your own memory of the event? Several studies on memory retrieval in a social context indicate that this is indeed the case, and that being exposed to another person's recall can change the listener's memory. Critically, the simple act of listening to somebody else retell the past may not just re-emphasize the recounted events in memory, but induce forgetting of related, but unmentioned details (Cuc et al., 2007). Such socially mediated forgetting effects can be relevant in conversations with friends, but potentially also in other applied scenarios in daily life, like when listening to news anchors, politicians, teachers, or other types of speakers (for a review, see Hirst & Echterhoff, 2012).

The focus of the present study is not on such applied aspects of socially mediated forgetting but on two more

basic issues that have largely been ignored in prior work. The one issue is the question of which cognitive mechanisms underlie this form of forgetting. While there is considerable knowledge about the cognitive mechanisms contributing to this type of forgetting in individuals, not much is yet known about the mechanisms driving socially mediated forgetting. The other issue is the possible role of testing format, asking whether the forgetting is restricted to recall tests—which is what has been employed as testing format in prior work—or whether it generalizes to item recognition tests. Because many types of episodic forgetting arise primarily on recall tests and much less on tests of item recognition (e.g., Baddeley et al., 2015), it is important to examine whether socially mediated forgetting is present in item recognition as well. In fact, the role of testing format is also relevant for the question which cognitive mechanisms mediate the forgetting. Before addressing these issues in more detail, we will first introduce the experimental task and report some of the main findings on this form of forgetting.

The task typically used to examine socially mediated forgetting is borrowed from research on individual remembering and called the retrieval-practice task (Anderson et al., 1994). In this task, participants are often asked to study

✉ Magdalena Abel
magdalena.abel@ur.de

¹ Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany

category-exemplar pairs with several exemplars from several different categories (e.g., FLOWER-*daffodil*, FLOWER-*tulip*, INSTRUMENT-*tuba*, INSTRUMENT-*horn*). During a subsequent practice phase, subjects receive category-plus-wordstem cues for some of the exemplars from some of the categories (e.g., FLOWER-*tu—?*) and are asked to recall the appropriate items. Such selective retrieval creates three item types: practiced items (called Rp+ items; here FLOWER-*tulip*), unpracticed items from the same categories (called Rp- items; here FLOWER-*daffodil*), and unpracticed control items from categories that were not practiced at all (called Nrp items; here INSTRUMENT-*tuba*, INSTRUMENT-*horn*). On the final test, on which subjects are asked to recall all initially studied items, recall of the practiced Rp+ items is typically enhanced and recall of the unpracticed Rp- items typically impaired, relative to recall of the Nrp control items. The two effects are referred to as retrieval-induced enhancement and retrieval-induced forgetting (RIF) in the following (for recent reviews, see Bäuml & Kliegl, 2017; Storm & Levy, 2012).

RIF has mostly been studied in single individuals who engage in retrieval practice by themselves (which results in *within individual RIF*, or WI-RIF), but the task has recently been adapted to also capture social situations (Cuc et al., 2007; for a review, see Hirst & Coman, 2018). In the adapted version of the task, two subjects are tested together. Initial study and final test phase are still completed individually, but the intermediate retrieval-practice phase is completed together. Here, one subject is asked to act as speaker, i.e., to engage in overt retrieval practice and recall the to-be-retrieved items aloud. The other subject is asked to act as listener and monitor the speaker's performance for accuracy. Several studies have applied this adapted version of the task and reported forgetting of unpracticed Rp- items relative to Nrp items not only in speakers, but also in listeners; not only for semantically categorized word lists, but also for coherent stories and even autobiographical memories (e.g., Barber & Mather, 2012; Cuc et al., 2007; Stone et al., 2010, 2013; see also Abel & Bäuml, 2015). The speaker's forgetting is consistent with regular WI-RIF observed in the standard retrieval-practice task. The listener's forgetting is termed *socially shared RIF* (SS-RIF) and suggests that the consequences of a speaker's selective retrieval can be socially transmitted to listeners (Cuc et al., 2007).

In the literature on SS-RIF, it has generally been proposed that WI-RIF and SS-RIF may arise on the basis of the same cognitive mechanism(s). This view rests on the finding that SS-RIF emerges only when listeners are motivated to engage in concurrent (covert) retrieval along with speakers. Consistently, when listeners are not asked to monitor a speaker's retrieval for accuracy, but for smoothness and fluidity of recall instead, no SS-RIF emerges (Cuc et al., 2007; for related findings, see Coman

& Hirst, 2015; Koppel et al., 2014). In the literature on WI-RIF, there are different views on exactly which cognitive mechanisms may cause the forgetting (e.g., Anderson, 2003; Jonker et al., 2013; Raaijmakers & Jakab, 2012). While none of these accounts seems to be able to explain the whole range of RIF findings (see Bäuml & Kliegl, 2017), a two-factor account that assumes a role of inhibition *and* blocking may be able to explain most of the findings.

The two-factor account assumes that inhibition is recruited during retrieval practice to reduce the interference that arises due to coactivation of the not-to-be-retrieved (Rp-) items. This inhibition is supposed to impair the item representation of these items and reduce retrieval of the items over a wide range of memory tasks (e.g., Anderson, 2003). The account also assumes that blocking can operate at test. Retrieval practice strengthens the category-exemplar associations of the Rp+ items and this strengthening may lead to blocking of the (weaker) Rp- items (e.g., Raaijmakers & Jakab, 2012; Verde, 2013). Critically, the contribution of the two mechanisms is assumed to vary with testing format: Whereas inhibition is supposed to impair retrieval of Rp- items in more or less all testing formats, blocking is supposed to impair retrieval of Rp- items mainly in test formats in which no item-specific cues are provided, like in free recall or category-cued recall. In contrast, when item-specific cues are provided—as, for instance, occurs when items' unique initial letters are provided as retrieval cues or in item recognition—the role of blocking should be reduced, if not eliminated, with inhibition mainly contributing to the forgetting. Results from several recent studies support this two-factor account (Anderson & Levy, 2007; Rupperecht & Bäuml, 2016, 2017; Schilling et al., 2014).

Following the two-factor account, it seems likely that blocking does not only contribute to WI-RIF in speakers, but also to SS-RIF in listeners. Prior studies examining RIF in a social context have consistently shown that retrieval practice strengthens Rp+ relative to Nrp items not only in speakers, but also in listeners. Even though some prior studies reported comparable, and others reported different degrees of retrieval-induced enhancement in speakers and listeners (e.g., Cuc et al., 2007; Koppel et al., 2014; Stone et al., 2010), the Rp+ items seem to benefit from strengthening in listeners as well, suggesting that blocking can play a role for SS-RIF in listeners. It is less clear, however, if inhibition contributes to SS-RIF in listeners as well.

There are at least two reasons for why inhibition might not be critically involved in SS-RIF in listeners. The one reason is that, in contrast to speakers, listeners during the practice phase are not only exposed to the provided retrieval cues (e.g., FLOWER-*tu—?*), but they are additionally asked to monitor the speaker's retrieval practice, e.g., for accuracy, and to provide ratings on corresponding Likert scales for

every single item (see Cuc et al., 2007). Effectively, this may create a dual-task situation for listeners. Because prior work on WI-RIF has shown that a secondary task during retrieval practice can decrease cognitive capacities and result in a reduction in inhibition and, consequently, in RIF (Ortega et al., 2012; Román et al., 2009), arguably, inhibition might still contribute to WI-RIF in speakers, but not, or to a lesser degree, to SS-RIF in listeners. The other reason is that listeners are assumed to engage in concurrent retrieval along with speakers, but possibly such covert retrieval might boil down to a rather passive item recognition task. Instead of trying to retrieve the correct answer themselves, listeners may simply wait for the speaker's response and then try to recognize if the speaker's response corresponds to an "old" item or not. Because strength-based interference is reduced on recognition tests (e.g., Ratcliff et al., 1990; Rupperecht & Bäuml, 2016) and because inhibitory processes are assumed to be recruited only when interference is detected (e.g., Anderson, 2003), the involvement of inhibition might be reduced, if not eliminated, in listeners that engage in a monitoring task like this.

If SS-RIF in listeners depends primarily on blocking, then the forgetting should vary with final test format. Following the two-factor account of RIF, forgetting due to blocking should arise mainly on final tests that do not include item-specific cues (like category-cued recall tests), and should be reduced, if not eliminated, on final tests that provide strong item-specific cues (like item recognition). Yet, if SS-RIF in listeners relies on blocking *and* inhibition, there should be not much difference between speakers and listeners in terms of RIF, and both speakers and listeners should show RIF regardless of final test format. Studies that compared WI-RIF and SS-RIF mostly relied on final category-cued recall tests in the absence of any item-specific retrieval cues (e.g., Barber & Mather, 2012; Cuc et al., 2007; Stone et al., 2013),¹ so that, so far, it is unclear if SS-RIF generalizes to, say, item recognition tests. Similarly, it is unclear if RIF in speakers and listeners are caused by the same cognitive mechanisms or if the critical contribution of inhibition is indeed reduced, or eliminated, in listeners.

The goal of the present study was to address the two issues. Four experiments were conducted: All experiments used the social version of the retrieval-practice task as introduced by Cuc et al. (2007), and pairs of subjects completed the task together, acting as speakers and listeners during the retrieval-practice phase. The four experiments

differed in which type of final test was applied and whether item-specific cues were present or absent at test. In Experiment 1, a category-cued recall test without any item-specific retrieval cues was applied to replicate prior findings from the literature on SS-RIF. In Experiment 2, items' initial letters were provided as additional retrieval cues, which offers item-specific information as well as the opportunity to control output order at test, i.e., to exclude the possibility that the prior recall of (stronger) Rp+ items at test induces RIF for (weaker) Rp- items (e.g., Smith, 1971; Tulving & Arbuckle, 1963). Finally, Experiments 3 and 4 applied item recognition tests. Experiment 3 used old/new recognition testing, whereas Experiment 4 applied a recognition test with confidence ratings that allowed construction of Receiver Operating Characteristic (ROC) curves. Both recognition tests comprise strong item-specific cues and allow for a control of output order at test.

The results of the four experiments will show for the first time if SS-RIF in listeners is limited to category-cued-recall tests or is present irrespective of final test format. If RIF is found to be present across all final test formats not only for speakers, but also for listeners, this would indicate that similar cognitive mechanisms cause RIF in speakers and listeners, and blocking and inhibition contribute to RIF in both participant groups. In contrast, if RIF is present across all final test formats only in speakers, but is reduced or even eliminated in listeners when strong item-specific cues are included, this would indicate that inhibition plays a reduced role in listeners, suggesting that WI-RIF and SS-RIF may rely on (partially) different cognitive mechanisms.

Experiment 1

The goal of Experiment 1 was to first replicate previous studies on WI-RIF and SS-RIF that mostly applied category-cued recall tests without any item-specific cues (e.g., Cuc et al., 2007; Stone et al., 2013). Subjects completed the experiment in pairs, studying and practicing category-exemplar word pairs (e.g., FLOWER-*daffodil*, INSTRUMENT-*tuba*). At test, category labels were provided as retrieval cues and subjects were asked to recall all previously studied items belonging to the category labels. On the basis of the prior work on regular WI-RIF on the one hand and SS-RIF on the other hand, we expected to observe RIF effects in both speakers and listeners.

Method

Participants Sample sizes in Experiments 1 and 2 were determined on the basis of prior recall experiments on retrieval-induced forgetting from our lab (e.g., Rupperecht & Bäuml, 2016). 80 subjects participated in pairs. Mean age

¹To the best of our knowledge, the only exception is a study by Coman et al. (2009) in which a final old/new recognition test was applied. The reported results do however only include response times (for hits) and further analyses for misses, instead of hit and false alarm rates for the different item types. Therefore, it remains unclear if WI-RIF and SS-RIF were in fact intact on this item-specific test or not.

was 20.55 years ($SD = 2.27$); 17 subjects were male, 63 female.

Material Item material was created by selecting six exemplars from each of eight semantic categories, taken from published category norms (Mannhaupt, 1983; Scheithe & Bäuml, 1995; Van Overschelde et al., 2004). Within categories, all items had unique initial letters. Across subjects, each category served equally often as to-be-practiced and as control category. Within categories, however, we kept constant which items were used as Rp+ and as Rp- items. Following Anderson et al. (1994), items with a lower frequency in the published category norms were used as Rp+ items, items with a higher frequency as Rp- items.

Design The experiment had a 2×3 mixed-factorial design. The factor role during retrieval practice (speaker, listener) was manipulated between-subjects. During the retrieval-practice phase, one subject in each pair was asked to act as speaker and to practice some of the initially studied items by trying to recall them out loud. The other subject was asked to act as listener and to monitor the speaker's overt practice for accuracy. The factor item type (Rp+, Rp-, Nrp) was manipulated within-subjects. During retrieval practice, only some items from some of the categories were presented for practice, thus creating the three different item types: practiced Rp+ items, related, but unpracticed Rp- items, and unrelated and unpracticed Nrp items.

Procedure. Study phase Subjects in each pair were seated in front of the same computer. During study, items were presented one at a time, together with their semantic category, and for 5 s each centrally on a computer screen. Subjects were asked to try to memorize them for a later test. Sequence of items was controlled via blocked randomization. Item material was organized into six lists, each containing one item from each semantic category. During study, the sequence of lists as well as the sequence of items within lists were set to random, so that a maximum of two items from the same category could be presented back to back.

Retrieval-practice phase During retrieval practice, subjects in both experiments were asked to practice half of the items from half of the categories, on two consecutive practice cycles. For each to-be-practiced item, the items' initial letters were presented below the corresponding category name. Items were again presented via blocked randomization and were shown one at a time and for 8 sec each. When all to-be-practiced items had been presented once, the second retrieval-practice cycle began. Importantly, one of the two subjects was asked to act as speaker, and

to practice some of the studied items out loud, by trying to complement the retrieval cues presented on the screen. The other subject was asked to act as listener. In particular, the listener's task was to monitor the speaker's practice, and to judge for each item how accurate the speaker's response was (for details, see Cuc et al., 2007).

Test phase Subjects completed the final test alone and worked at separate computers. The test was a category-cued recall test, i.e., subjects received category cues only for recall. Category names were presented one at a time, in random sequence, and for 48 sec each. Subjects received a small test booklet, with one page for recall of each of the eight semantic categories. On each page, subjects were asked to write down the category name and to then try to list as many of the studied category exemplars as they could remember.

Results

Success rates during retrieval-practice cycles On average, speakers were able to correctly recall 90.22% ($SD = 10.74$) of Rp+ items on the first retrieval-practice cycle, and even improved to 92.34% ($SD = 9.75$) on the second retrieval-practice cycle, $t(39) = 3.20$, $p = .003$, $d = 0.51$.

Final test performance Figure 1 shows mean recall performance on the final test, separately for speakers and listeners.

Retrieval-induced forgetting A 2x2 ANOVA with the factors of ITEM TYPE (Rp-, Nrp) and ROLE DURING PRACTICE (speaker, listener) showed a significant main effect of ITEM TYPE, $F(1, 78) = 14.66$, $MSE = 0.03$, $p < .001$, $\eta_p^2 = .16$, but no significant main effect of ROLE DURING PRACTICE, $F(1, 78) < 1.0$, $p = .418$, $\eta_p^2 = .008$, and also no significant interaction between the two factors, $F(1, 78) < 1.0$, $p = .949$, $\eta_p^2 < .001$. Recall did not differ between speakers and listeners (51.70% vs. 54.64%), but was higher for Nrp than Rp-items (58.76% vs. 47.58%). Follow-up t tests confirmed that this retrieval-induced forgetting was significant for both speakers (57.38% vs. 46.01%), $t(39) = 2.58$, $p = .014$, $d = 0.41$, and listeners (60.14% vs. 49.14%), $t(39) = 2.87$, $p = .007$, $d = 0.45$.

Retrieval-induced enhancement A 2x2 ANOVA with the factors of ITEM TYPE (Rp+, Nrp) and ROLE DURING PRACTICE (speaker, listener) revealed a significant main effect of ITEM TYPE, $F(1, 78) = 129.78$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = .63$, reflecting higher recall rates for Rp+ than Nrp items (83.22% vs. 58.76%). The main effect of

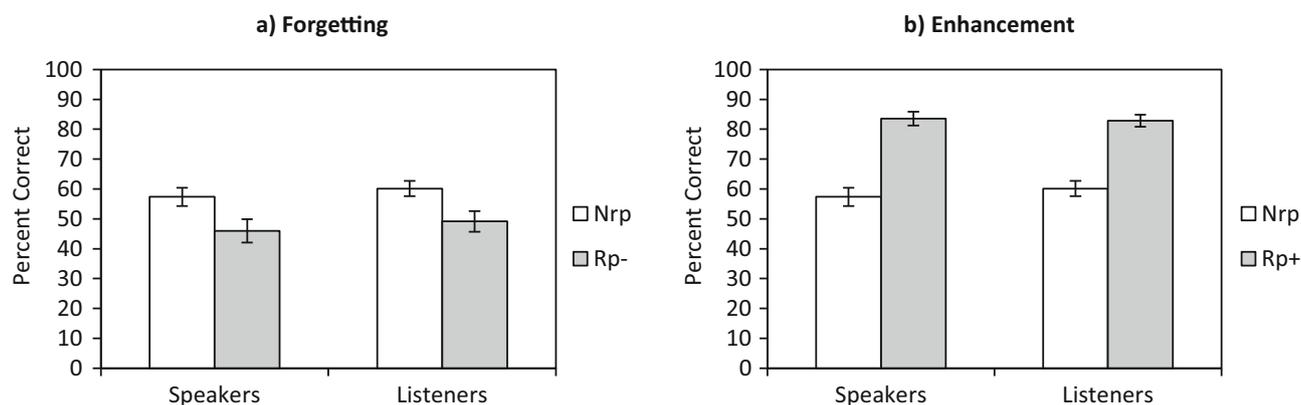


Fig. 1 Mean recall in Experiment 1, plotted separately for speakers and listeners. Panel **a** contrasts Rp- and Nrp items, whereas panel **b** contrasts Rp+ and Nrp items. Error bars represent ± 1 standard errors

ROLE DURING PRACTICE was not significant, $F(1, 78) < 1.0$, $p = .714$, $\eta_p^2 = .002$, suggesting that recall did not differ between speakers and listeners (70.47% vs. 71.52%). The ANOVA also returned a non-significant interaction of the two factors, $F(1, 78) < 1.0$, $p = .426$, $\eta_p^2 = .008$, indicating that retrieval-induced enhancement did not differ between speakers and listeners. Follow-up t tests confirmed that retrieval-induced enhancement was indeed present in both speakers (83.56% vs. 57.38%), $t(39) = 8.45$, $p < .001$, $d = 1.34$, and listeners (82.87% vs. 60.14%), $t(39) = 7.65$, $p < .001$, $d = 1.21$.

Discussion

The results of Experiment 1 replicate prior work that reported WI-RIF and SS-RIF on a category-cued recall test (e.g., Coman & Hirst, 2015; Cuc et al., 2007; Stone et al., 2013). Additionally, both forgetting of Rp- items and enhancement of Rp+ items were comparable in size across speakers and listeners, which is per se consistent with the view that similar cognitive mechanisms may mediate WI-RIF and SS-RIF. Category-cued recall tests come with two problems, however. First, output order at test is not controlled on such tests, leaving the possibility that forgetting of the (weaker) Rp- items arises due to output interference, i.e., because the (stronger) Rp+ items are recalled first at test (e.g., Smith, 1971; Tulving and Arbuckle, 1963). Second, blocking can play a major role for RIF on such tests. As a consequence, results relying on such tests are largely silent regarding the degree to which inhibition may have contributed to the forgetting – and whether this contribution differed between speakers and listeners. Experiment 2 therefore revisited the issue, using a cued-recall test with item-specific cues.

Experiment 2

Experiment 2 provided the studied items' category labels plus the items' unique initial-letters as item-specific retrieval cues at test. Such item-specific cues allow to control the output order of the single items and thus prevent that the prior recall of the (stronger) Rp+ items at test induces RIF for the (weaker) Rp- items; Rp- items were therefore tested prior to Rp+ items. Moreover, such item-specific cues can decrease the relative contribution of blocking to RIF and increase the relative contribution of inhibition (e.g., Anderson et al., 1994; Bäuml, 1998). If WI-RIF in speakers and SS-RIF in listeners are caused by the same cognitive mechanisms and these mechanisms contribute to about the same degree to the forgetting in the two participant groups, then the results of Experiment 2 should replicate those of Experiment 1, with similar amounts of RIF in speakers and listeners. In contrast, if SS-RIF in listeners is primarily caused by blocking and inhibition contributes to WI-RIF only, then RIF may be reduced in listeners relative to speakers.

Method

Participants As in Experiment 1, the sample comprised 80 subjects, tested in pairs. Mean age was 23.15 years ($SD = 3.89$); 29 subjects were male, 51 female.

Material The same item material was applied as in Experiment 1.

Design The experiment had a 2×4 mixed-factorial design. As in Experiment 1, the factor role during retrieval practice (speaker, listener) was manipulated between-subjects. The factor item type (Rp+, Rp-, Nrp+, Nrp-)

was manipulated within-subjects. Selective retrieval in the intermediate practice phase again created practiced categories (containing Rp+ and Rp- items) as well as unpracticed control categories (containing Nrp items). In Experiment 2, due to the adapted final-test format, it was possible to further divide Nrp items into Nrp+ items (serving as control items for Rp+ items) and Nrp- items (serving as control items for Rp- items; see below for details).

Procedure The procedure was identical to the one applied in Experiment 1, the only difference being the final-test format. In Experiment 2, the final test was a category-plus-initial-letter-cued recall test, i.e., subjects did not only receive category names as cues but also the items' initial letters. All items from one semantic category were tested back to back, but sequence of items within categories was controlled. Each item's initial letter was presented together with the name of the category, for 8 s each. For practiced categories, Rp- items were always tested first, in random order, to reduce potential output interference effects; subsequently, Rp+ items were tested in the same way. To ensure adequate baselines for (tested first) Rp- and (tested second) Rp+ items, Nrp categories were tested in the same manner. The first half of items tested in each Nrp category was used as baseline for Rp- items and labeled Nrp- items; the second half of tested items was used as baseline for Rp+ items and labeled Nrp+ items.

Results

Success rates during retrieval-practice cycles Speakers' practice performance was again excellent, with 88.95% ($SD = 13.07$) of Rp+ items being correctly recalled on the first retrieval-practice cycle, and 89.57% ($SD = 14.00$) on the second retrieval-practice cycle, $t(39) < 1.0$, $p = .462$, $d = 0.12$.

Final test performance Figure 2 shows mean recall performance on the final test, separately for speakers and listeners.

Retrieval-induced forgetting A 2x2 ANOVA with the factors of ITEM TYPE (Rp-, Nrp) and ROLE DURING PRACTICE (speaker, listener) detected a significant main effect of ITEM TYPE, $F(1, 78) = 21.40$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = .22$, but no significant main effect of ROLE DURING PRACTICE, $F(1, 78) < 1.0$, $p = .741$, $\eta_p^2 = .001$, and also no significant interaction between the two factors, $F(1, 78) < 1.0$, $p = .823$, $\eta_p^2 = .001$. Recall did not differ between speakers and listeners (72.16% vs. 71.03%), but was higher for Nrp- than Rp- items (76.48% vs. 66.70%). Follow-up t tests confirmed significant retrieval-induced forgetting for both speakers (77.27% vs. 67.02%), $t(39) =$

3.64, $p = .001$, $d = 0.58$, and listeners (60.14% vs. 49.14%), $t(39) = 2.95$, $p = .005$, $d = 0.47$.

Retrieval-induced enhancement A 2x2 ANOVA with the factors of ITEM TYPE (Rp+, Nrp+) and ROLE DURING PRACTICE (speaker, listener) revealed a significant main effect of ITEM TYPE, $F(1, 78) = 17.21$, $MSE = 0.03$, $p < .001$, $\eta_p^2 = .18$, reflecting higher recall rates for Rp+ than Nrp+ items (83.55% vs. 71.90%). The main effect of ROLE DURING PRACTICE was not significant, $F(1, 78) < 1.0$, $p = .606$, $\eta_p^2 = .003$, suggesting that recall did not differ between speakers and listeners (77.00% vs. 78.46%). The interaction of ITEM TYPE and ROLE DURING PRACTICE was also not significant, $F(1, 78) < 1.0$, $p = .944$, $\eta_p^2 < .001$, which indicates that the size of the enhancement effect was not affected by ROLE DURING PRACTICE. Follow-up t tests confirmed significant retrieval-induced enhancement in both speakers (82.93% vs. 71.08%), $t(39) = 3.03$, $p = .004$, $d = 0.48$, and listeners (84.18% vs. 72.73%), $t(39) = 2.84$, $p = .007$, $d = 0.45$.

Discussion

The results of Experiment 2 show intact WI-RIF and SS-RIF with controlled output order at test, indicating that the forgetting of the Rp- items is not just a consequence of output interference arising from the fact that the (stronger) Rp+ items are recalled first. Moreover, the results provide a first demonstration that (i) SS-RIF in listeners can also arise when item-specific cues are presented at test, and (ii) SS-RIF can be equivalent in size to WI-RIF in speakers under such conditions. On the basis of the two-factor account of RIF, this demonstration is consistent with the view that SS-RIF and WI-RIF are caused by the same cognitive mechanisms and inhibition may not only contribute to RIF in speakers, but also in listeners. Still, although items' unique initial letters served as item-specific retrieval cues in this experiment, these letters constitute relatively weak item-specific cues, which do not always reduce or eliminate blocking processes. In fact, whereas some RIF studies provided evidence that such cues can eliminate blocking (e.g., Anderson et al., 1994; Bäuml, 1998), other RIF studies reported contrasting results (e.g., Raaijmakers & Jakab, 2012; Rupperecht & Bäuml, 2016; Verde, 2013). It was therefore the goal of Experiments 3 and 4 to use stronger item-specific cues at test in order to enable stronger examination of the question if inhibition also contributes to SS-RIF in listeners.

Experiment 3

There is a good amount of evidence in the literature that strength-based interference effects - as they may occur in

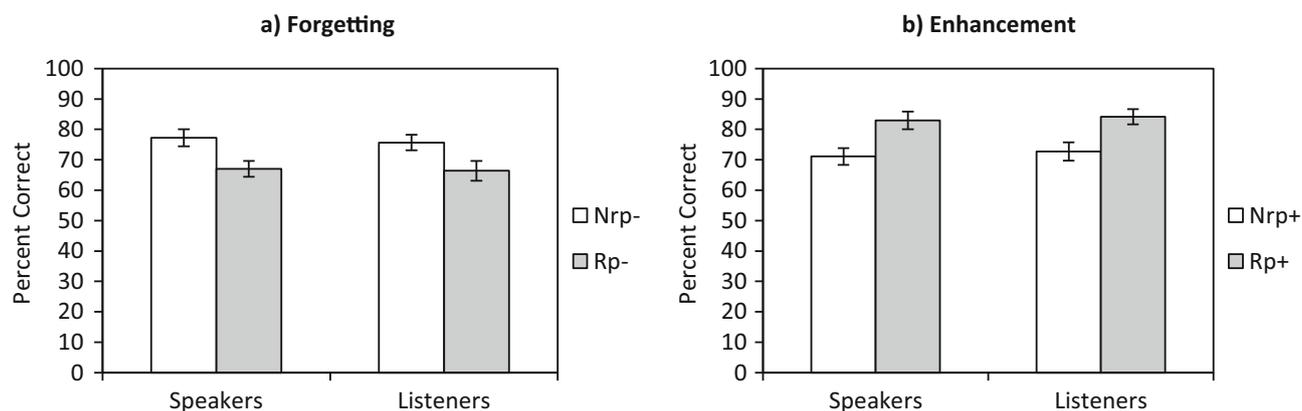


Fig. 2 Mean recall in Experiment 2, plotted separately for speakers and listeners. Panel **a** contrasts Rp- and Nrp- items, whereas panel **b** contrasts Rp+ and Nrp+ items. Error bars represent ± 1 standard errors

the retrieval-practice task but also in other experimental tasks, like the list-strength effect - can be present in recall but do not arise in item recognition (Ratcliff et al., 1990; Rupperecht & Bäuml, 2016; Shiffrin et al., 1990), a finding well captured by models of recognition memory (e.g., Dennis & Humphreys, 2001; Shiffrin & Steyvers, 1997). Experiment 3 therefore employed a yes/no recognition test to reduce blocking at test and thus to examine if inhibition does not only play a role for WI-RIF in speakers, but also for SS-RIF in listeners. The basic experiment was similar to Experiments 1 and 2, but at test, subjects were presented with all items from the study phase – intermixed with new items from the same semantic categories. The subjects' task was to judge for each single item whether it was an old item from the study phase or a new item not studied earlier.

Based on prior work that showed RIF in individuals on such item recognition tests (e.g., Gómez-Ariza et al., 2005; Hicks & Starns, 2004; Spitzer & Bäuml, 2007) we expected WI-RIF in speakers in this experiment. For listeners, expectations depend on whether inhibition is assumed to play a critical role for SS-RIF. If this is the case, then WI-RIF and SS-RIF should both be present and be largely equivalent in size. In contrast, if inhibition plays a reduced role in listeners, then RIF should be present in speakers, but be reduced or eliminated in listeners.

Method

Participants Sample sizes in Experiments 3 and 4 were determined on the basis of prior recognition experiments on RIF from our lab (e.g., Rupperecht & Bäuml, 2016, 2017; Spitzer & Bäuml, 2007, 2009). Ninety-six subjects participated in pairs. Mean age was 21.32 years ($SD = 3.05$); 16 subjects were male, 80 female.

Material Item material again consisted of exemplars from semantic categories, but we used new semantic categories in

Experiment 3. Moreover, the number of items per category was doubled and 12 exemplars were selected from each of 8 semantic categories. These materials were divided into two sets of items, with each set containing 6 exemplars per category. The items of one of the two sets were used to be studied and practiced in the experiment, but the items of the other set served as lures during the recognition test. Sets of items were counterbalanced across participants and were equally often used as study and lure items.

Design The experiment had the same 2×4 mixed-factorial design as Experiment 2, with the between-participants factor of role during retrieval practice (speaker, listener) and the within-participants factor of item type (Rp+, Rp-, Nrp+, Nrp-).

Procedure The procedures for study and retrieval-practice phases were identical to those applied in Experiments 1 and 2. After retrieval practice, however, subjects were asked to engage in unrelated distractor tasks for 8 min; this delay was added to prevent ceiling effects that might emerge on an immediate recognition test.

At test, all initially studied items were presented one at a time and intermixed with the same number of lure items. To control output interference, the first half of the test only comprised Rp- items and half of the Nrp items (called Nrp- items, because they were tested intermixed with Rp- items and corresponding “new” lure items). The second half of the test then comprised the remaining Rp+ items and the other half of the Nrp items (called Nrp+ items, because they were tested intermixed with Rp+ items and corresponding lures). Within each half of the test, we again used blocked randomization such that old (or new) items from the same semantic category could appear back to back at most two times in succession. For each single item, subjects were asked to judge whether it was old (i.e., studied in the experiment) or new (i.e., not previously encountered

in the experiment); subjects entered their responses via the computer keyboard. The test was subject-paced, i.e., the next item appeared on the screen as soon as subjects entered a response. Mean test duration was 137.48 sec ($SD = 30.43$).

Results

Success rates during retrieval practice Speakers' performance was excellent on both practice cycles, with 89.24% ($SD = 11.91$) of Rp+ items being recalled on the first retrieval-practice cycle and 90.45% ($SD = 12.98$) on the second retrieval-practice cycle, $t(47) = 1.55$, $p = .128$, $d = 0.22$.

Final test performance Table 1 shows hits and false alarms, separately for the different item types and for speakers and listeners. Initial analyses indicated that false alarm rates were not significantly different for Rp- vs. Nrp- items or for Rp+ vs. Nrp+ items; all $t_s(95) \leq 1.95$, all $p_s \geq .054$, all $d_s \leq 0.20$. For each single item type, false alarm rates were also comparable between speakers and listeners; all $t_s(94) \leq 1.85$, all $p_s \geq .067$, all $d_s \leq 0.38$. Figure 3 shows mean corrected hit rates (i.e., hits - false alarms) for all item types.

Retrieval-induced forgetting A 2x2 ANOVA with the factors of ITEM TYPE (Rp-, Nrp-) and ROLE DURING PRACTICE (speaker, listener) only revealed a significant main effect of ITEM TYPE, $F(1, 94) = 15.14$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = .14$, which suggests that recognition was generally reduced for Rp- compared with Nrp- items (52.68% vs. 61.27%). There was no significant main effect of ROLE $F(1, 94) = 2.28$, $MSE = 0.06$, $p = .134$, $\eta_p^2 = .02$, and also no significant interaction, $F(1, 94) = 1.68$, $MSE = 0.02$, $p = .198$, $\eta_p^2 = .02$, suggesting that the difference between item types was not affected by role during practice. Follow-up tests, however, confirmed significantly reduced recognition of Rp- vs. Nrp- items only

in speakers (53.85% vs. 65.31%), $t(47) = 4.05$, $p < .001$, $d = 0.58$. The numerical pattern in listeners indicated forgetting, too (51.50% vs. 57.23%), but statistically this difference did not reach significance, $t(47) = 1.69$, $p = .098$, $d = 0.24$.

Retrieval-induced enhancement A 2x2 ANOVA with the factors of ITEM TYPE (Rp+, Nrp+) and ROLE DURING PRACTICE (speaker, listener) showed a significant main effect of ITEM TYPE, $F(1, 94) = 176.61$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = .65$, reflecting higher recognition accuracy for Rp+ than Nrp+ items (83.48% vs. 56.16%). The ANOVA also returned a significant main effect of ROLE, $F(1, 94) = 4.88$, $MSE = 0.04$, $p = .030$, $\eta_p^2 = .05$, which was accompanied by a significant interaction with ITEM TYPE, $F(1, 94) = 7.63$, $MSE = 0.02$, $p = .007$, $\eta_p^2 = .08$, suggesting that the recognition difference between speakers and listeners depended on item type. Follow-up tests showed that recognition of Rp+ items did not differ between speakers and listeners (83.69% vs. 83.27%), $t(94) < 1.00$, $p = .877$, $d = 0.03$, but Nrp+ recognition was higher in speakers than in listeners (62.04% vs. 50.27%), $t(94) = 2.90$, $p = .005$, $d = 0.59$. Consistently, recognition was enhanced for Rp+ relative to Nrp+ items in speakers (83.69% vs. 62.04%), $t(47) = 7.05$, $p < .001$, $d = 1.02$, but even more so in listeners (83.27% vs. 50.27%), $t(47) = 12.07$, $p < .001$, $d = 1.74$.²

Discussion

On the one hand, the ANOVA results replicate those of Experiments 1 and 2 by showing RIF that did not differ statistically between speakers and listeners. On the other hand, the results deviate somewhat from those of the two previous experiments, because the effect size of RIF reported for listeners in this experiment was only half the size of that for speakers - as is also reflected by the fact that significant RIF in single comparisons arose in speakers but not listeners. Sample sizes for Experiment 3 were determined on the basis of prior work from our lab, but a sensitivity analysis (with 48 subjects per condition, an alpha level of .05 and power of .80) indicates that the smallest detectable effect size for within-group comparisons

Table 1 Hits and false alarms in Experiment 3

Item type	Speakers		Listeners	
	Hits	False alarms	Hits	False alarms
Rp-	.730	.191	.713	.198
Nrp-	.800	.147	.765	.193
Rp+	.957	.120	.972	.139
Nrp+	.757	.137	.685	.182

²Recognition performance was also analyzed using d' rather than hits-false alarms to estimate enhancement and forgetting effects. Usage of d' in these types of experiments is suboptimal because it relies on the assumption of equal variances of old and new items, which is not supported by typical data (see Experiment 4, for an example). Nonetheless, the pattern of results reported above when using the hits-false alarm measure did not change when using the d' measure.

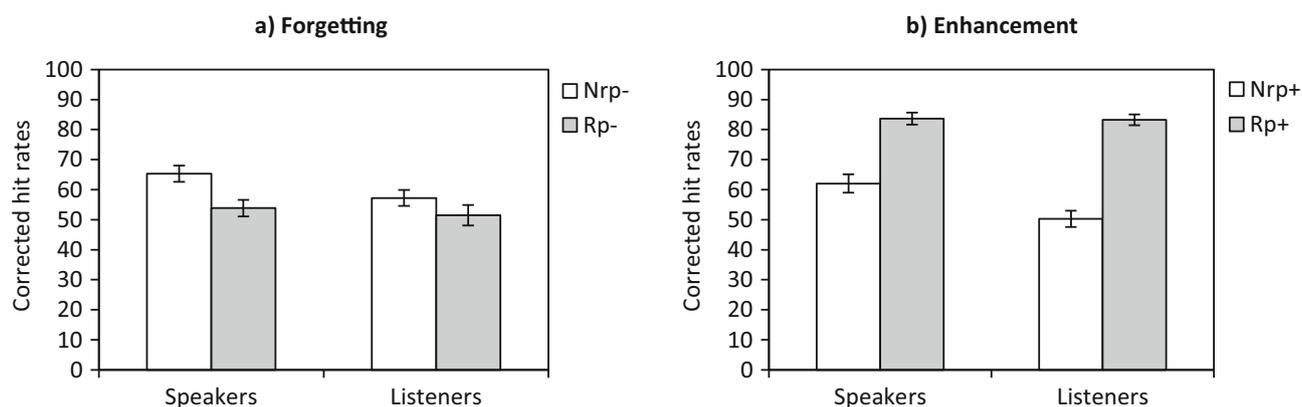


Fig. 3 Mean corrected hit rates (hits-false alarms) in Experiment 3, plotted separately for speakers and listeners. Panel **a** contrasts Rp- and Nrp- items, whereas panel **b** contrasts Rp+ and Nrp+ items. Error bars represent ± 1 standard errors

was approximately $d = .41$. Because smaller-sized effects were therefore difficult to detect, the single comparisons in this experiment may have found RIF to be significant for speakers but not for listeners.

Retrieval-induced enhancement was stronger in listeners than speakers, which makes it particularly unlikely that the numerical difference in amount of RIF was related to differences in blocking processes. Rather, on the basis of the view that mainly inhibition contributes to RIF in item recognition, the pattern of numerically reduced RIF in listeners may suggest that the contribution of inhibition is somewhat reduced in listeners relative to speakers. Because this pattern is present numerically but not statistically in Experiment 3, the goal of Experiment 4 was to reexamine the issue, this time by means of an item recognition test that applies confidence ratings and enables ROC analyses. ROC analysis can overcome problems that may be present with yes/no recognition testing (e.g., Macmillan & Creelman, 2004; Wixted, 2007b; see also below) and thus may provide more accurate information regarding differences between WI-RIF in speakers and SS-RIF in listeners.

Experiment 4

Experiment 4 was largely identical to Experiment 3, the only difference being that the applied item recognition test was based on confidence ratings instead of old/new judgments. At test, old items from the study phase were again presented intermixed with new items from the same semantic categories, and subjects provided a rating for each item reflecting how confident they were that the item was old or new. On the basis of prior work on individual RIF that used the same type of memory test (e.g., Rupprecht & Bäuml, 2016, 2017; Spitzer & Bäuml, 2007, 2009), we again expected WI-RIF in speakers. Depending on the reading of the results of Experiment 3, one may expect a similar SS-RIF effect or a reduced SS-RIF effect in listeners.

Method

Participants

96 subjects participated in pairs. Mean age was 21.95 years ($SD = 2.62$); 38 subjects were male, 58 female.

Material We used the same semantic categories as in Experiments 1 and 2, but as in Experiment 3, 12 exemplars were chosen from each of the 8 semantic categories. Materials were divided into two sets of items, with each set containing 6 exemplars per category. Sets of items were counterbalanced across participants and equally often used as study and lure items.

Design The experiment again followed a 2×4 mixed-factorial design with the between-participants factor of role during retrieval practice (speaker, listener) and the within-participants factor of item type (Rp+, Rp-, Nrp+, Nrp-).

Procedure The procedures for study, retrieval-practice, and distractor phases were identical to those applied in Experiment 3. At test, all initially studied items were presented one at a time and intermixed with the same number of lure items. In Experiment 4, a schematic rating scale was depicted below each item, and subjects were asked to rate for each single item how confident they were that the item had previously been studied (old) or not (new); the rating scale ranged from 1 to 6 (1=definitely old; 6=definitely new). Participants entered their answers via the computer keyboard. Before starting the test, subjects were instructed to try to make use of the whole range of the rating scale. The test was completely subject-paced, i.e., as soon as subjects entered a response, the next item was shown on the screen. Mean test duration was 209.26 sec ($SD = 47.98$).

We again controlled sequence of items at test. Unbeknownst to participants, the first half of the test only comprised Rp- items and half of the Nrp items (called

Nrp- items), tested intermixed with corresponding lures. The second half of the test comprised the remaining Rp+ items and the other half of the Nrp items (called Nrp+ items), also mixed with lures. Within each half of the test, we again used blocked randomization such that items from the same semantic category could appear back to back at most two times in succession.

ROC curves and statistical analysis Proportion of initially studied items correctly recognized as old (i.e., hits) and proportion of lure items incorrectly recognized as old (i.e., false alarms) were cumulated across the rating scale response options, beginning with the most confident response option (definitely old, “1”). This approach allows to plot ROC curves, relating hits and false alarms 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. We applied the unequal-variance signal detection model to analyze the ROC data (e.g., Dunn, 2004; Wixted, 2007a), because the assumption of unequal variance for the distributions of old and new items can accommodate the typically asymmetrical shape of the ROC. According to this model, recognition responses are made on the basis of a single source of information, namely the items' general memory strength, which may reflect the additive or nonadditive combination of familiarity and recollection codes (e.g., Kelley & Wixted, 2001; Wixted & Stretch, 2004). Using this model, discriminability of old relative to new items can be derived from the distance between the means of the underlying strength distributions of those old and new items (d_a). When applied to 5-point ROC data, the model results in seven free parameters (discriminability d_a , variance of the distribution of old items σ , and five criterion points c_1 – c_5); when testing the model's goodness of fit, three degrees of freedom are left. To estimate model parameters, we used maximum-likelihood methods which allow statistical testing (for further details on the general analysis approach, see Rupperecht & Bäuml, 2016, 2017).

Results

Success rates during retrieval practice Speakers' performance was similar on the two practice cycles, with 91.85% ($SD = 10.95$) of Rp+ items being recalled on the first retrieval-practice cycle and 92.80% ($SD = 10.06$) on the second retrieval-practice cycle, $t(47) = 1.07$, $p = .291$, $d = 0.15$.

Recognition test: overall recognition performance To see if the general numerical recognition pattern was consistent

with that observed in Experiment 2, we determined overall hit and false alarm rates by collapsing subjects' responses across the three most confident response levels of the rating scale (i.e., rating levels 1–3) and then compiled corrected hits (hits–false alarms) separately for the single item types. Corrected hits were reduced for Rp- relative to Nrp- items, and this reduction was larger for speakers (47.4% vs. 54.6%) than listeners (59.0% vs. 61.7%). Moreover, corrected hits were enhanced for Rp+ relative to Nrp+ items, for both speakers (78.0% vs. 56.1%) and listeners (82.4% vs. 63.9%). Yet, because an analysis of corrected hits assumes that the underlying ROC function is linear (e.g., Wixted, 2007b), and because ROC functions are typically curvilinear and asymmetric (see below), our main analysis relied on a signal detection approach, which can account for the curvilinear and asymmetric form of the ROC.

Recognition test: unequal-variance signal detection model

Figure 4 shows the ROCs for Rp- vs. Nrp- items and for Rp+ vs. Nrp+ items, separately for speakers and listeners. In the first step, we examined if the unequal-variance signal detection model described the data sufficiently for each item type and in both speakers and listeners. Table 2 shows the statistics of goodness-of-fit and maximum-likelihood estimates of the model's parameters d_a and σ . The model provided a good fit to the data for all item types, $\chi^2s(3) \leq 1.21$, $ps \geq .752$. The variance of the old items' distribution, as estimated by the parameter σ , was larger than 1.0 for both speakers and listeners, $\chi^2s(1) \geq 12.29$, $ps < .001$, indicating that the model's assumption of unequal variances for old and new items improved the description of the data significantly. For Rp- and Nrp- items, σ was constant across item types, $\chi^2s(1) \leq 0.23$, $ps \geq .634$. For Rp+ and Nrp+ items, it varied with item type for speakers, $\chi^2(1) = 5.71$, $p = .017$, though not listeners, $\chi^2(1) = 2.89$, $p = .089$. Placement of the five response criteria never varied across item types, $\chi^2s(4) \leq 2.69$, $ps \geq .611$.

In the second step, we evaluated retrieval-induced forgetting and enhancement and tested potential differences in discriminability d_a across item types, separately for speakers and listeners. For speakers, Rp- compared with Nrp- items showed reduced discriminability d_a (1.53 vs. 1.85), $\chi^2(1) = 29.62$, $p < .001$, reflecting RIF for Rp- items. For listeners, discriminability d_a also differed between Rp- and Nrp- items (1.87 vs. 2.06), $\chi^2(1) = 4.40$, $p = .036$. Therefore, SS-RIF was present, although the reduction in discriminability for Rp- items was smaller in listeners than speakers, $\chi^2(1) = 6.11$, $p = .013$. Rp+ items showed enhanced discriminability d_a relative to Nrp+ items for both speakers and listeners, $\chi^2s(1) \geq 35.60$, $ps \leq .001$. This retrieval-induced enhancement was larger in speakers (5.80 vs. 1.98) than listeners (3.06 vs. 2.37), $\chi^2(1) = 8.00$, $p = .005$.

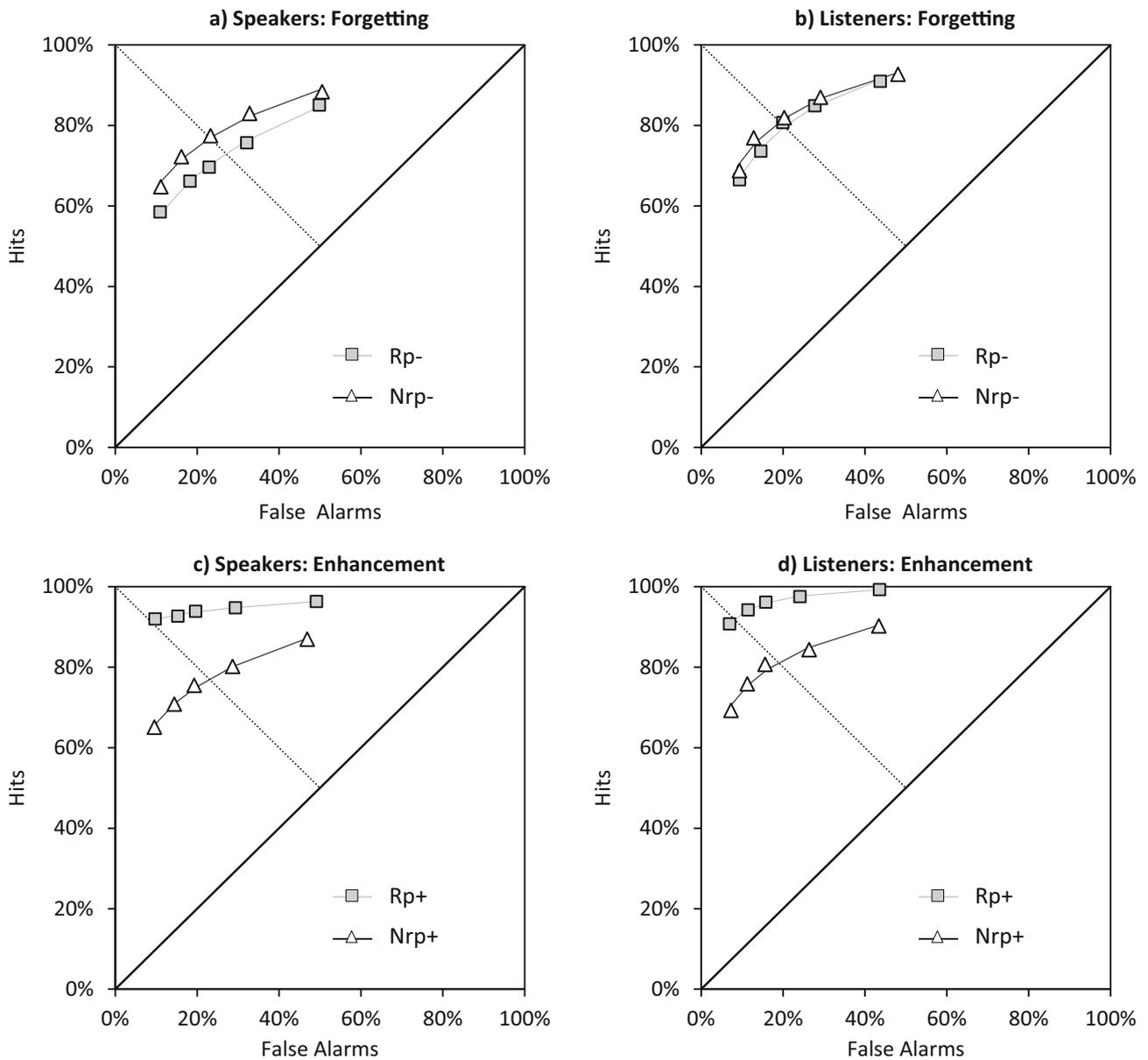


Fig. 4 ROC curves based on cumulative hit and false alarm rates in Experiment 4, plotted separately for speakers (panels **a** and **c**) and listeners (panels **b** and **d**). Panels **a** and **b** show ROC curves for Rp- and

Nrp- items, whereas panels **c** and **d** show ROC curves for Rp+ and Nrp+ items. The lines between data points reflect theoretical ROCs predicted by the unequal-variance signal detection model

Discussion

The results of the experiment show significant RIF for both speakers and listeners, but with a larger RIF effect for speakers than listeners, which is numerically consistent with the results of Experiment 3. On the basis of the view that RIF in item recognition reflects mainly inhibition, this finding suggests that inhibition contributes not only to WI-RIF but does also contribute to SS-RIF. This holds while

the contribution to RIF seems to be reduced in listeners relative to speakers. Retrieval-induced enhancement of Rp+ items was larger for listeners in Experiment 3, but larger for speakers in Experiment 4. On its own, Experiment 4 could thus be seen as consistent with an alternative blocking explanation, which may suggest a larger RIF effect when the enhancement effect is increased. However, as emphasized above, blocking does not seem to contribute much to RIF in item recognition and there is also no evidence that

Table 2 Unequal-variance signal detection model for Experiment 4

Item type	Parameter estimates		Goodness of fit		
	d_a	σ	χ^2	df	p
Speakers					
Rp-	1.53	1.49	0.21	3	0.976
Nrp-	1.85	1.52	0.76	3	0.859
Rp+	5.80	3.23	0.16	3	0.984
Nrp+	1.98	1.68	0.22	3	0.975
Listeners					
Rp-	1.87	1.26	0.62	3	0.892
Nrp-	2.06	1.35	1.21	3	0.752
Rp+	3.06	1.18	0.13	3	0.988
Nrp+	2.37	1.68	1.20	3	0.753

the size of the enhancement effect is correlated with the size of the RIF effect in item recognition (see Rupperecht & Bäuml, 2016, 2017; see also Murayama et al., 2014). Blocking thus should not have mediated the results of this experiment.

General discussion

The present results concerning WI-RIF in speakers replicate typical findings from the literature on RIF in individuals, confirming that RIF is present in category-cued recall, initial-letter-cued recall, and item recognition. The results concerning SS-RIF in listeners also replicate prior work by showing that SS-RIF is present in category-cued recall (Experiment 1). However, they also extend the prior work by showing that SS-RIF can also arise on initial-letter cued recall (Experiment 2) and item recognition (Experiments 3 and 4). RIF was similar in size for speakers and listeners when category or initial-letter cues were applied at test (Experiments 1 and 2), but the results obtained on item recognition tests indicated a reduced size of SS-RIF in listeners relative to WI-RIF in speakers. In Experiment 3, this reduction was present numerically only, in Experiment 4 it was also present statistically.

According to the two-factor account of RIF, there are mainly two mechanisms that contribute to RIF, inhibition and blocking. Inhibition is assumed to be involved irrespective of which final test is applied, whereas the additional contribution of blocking is assumed to arise mainly when item-specific retrieval cues are absent at test and to be reduced or even eliminated when such cues are present. On the basis of this account, the results of the present experiments suggest that inhibition and blocking are involved in both WI-RIF in speakers and SS-RIF in

listeners. While this equivalence holds qualitatively, the results also suggest subtle quantitative differences between the two participant groups, with inhibition contributing more in speakers and less in listeners.

At least two factors could be responsible for why the involvement of inhibitory processes is reduced in listeners. First, the additional completion of a monitoring task may create a dual-task situation for listeners. For individuals, a secondary task during retrieval practice has been shown to take up cognitive resources, thus reducing the involvement of inhibition and resulting in reduced RIF (see Ortega et al., 2012; Román et al., 2009). The same reasoning can also be applied to explain why the involvement of inhibition is reduced in SS-RIF in listeners. This reasoning is not in conflict with the finding that retrieval practice still entailed enhancement for Rp+ items in listeners. Indeed, prior work on the positive effects of retrieval practice in both the RIF literature and the testing-effect literature has shown that the benefits of retrieval practice can still be reaped when secondary tasks are introduced (Buchin & Mulligan, 2017; Mulligan & Picklesimer, 2016; Ortega et al., 2012; Román et al., 2009).

Second, the monitoring task may also have prompted listeners to engage in a recognition task rather than in more effortful retrieval practice. Because the involvement of inhibition is assumed to depend on interference (Anderson, 2003) and because strength-based interference is reduced in item recognition (Ratcliff et al., 1990; Rupperecht & Bäuml, 2016), this could explain why inhibition is less involved in SS-RIF than WI-RIF. Using the same basic speaker-listener setup as applied in the present experiments, with listeners engaging in accuracy monitoring, a recent study examined non-selective retrieval practice and resulting testing effects in a social context (Abel & Roediger, 2018). Across two-day delays, this study found regular benefits of retrieval practice relative to restudy in speakers, but only reduced benefits in listeners; equivalent testing effects only emerged when accountability was enhanced and listeners were asked to monitor their own retrieval (instead of the speaker's retrieval). Although these findings for testing effects arose on the basis of a more demanding memory task, they are generally consistent with the idea that monitoring a speaker's response for accuracy might prompt recognition judgments rather than effortful retrieval practice in listeners.

Clearly, more work is needed to better understand and tease apart the exact processes involved during retrieval practice in speakers and listeners. Although both factors described above may have contributed to the present findings, their importance may vary within and across individuals. For instance, subjects may differ (generally or on specific trials) in whether they manage or even try to retrieve the correct answer themselves before being

exposed to the speaker's responses. Alternatively, it may be possible to passively wait for the speaker's response and then provide an accuracy rating on the basis of a quick recognition judgment. The retrieval practice plus monitoring task as used in both the present and prior studies does not enable an investigation of what type of retrieval listeners engage in, or of whether judging the speaker's accuracy on rating scales takes up cognitive resources and creates a dual-task situation for listeners. Consequently, the present findings must remain silent on the extent to which the two factors contributed to the results. Moreover, it also remains unclear if similar findings would be obtained if free-flowing conversations were used instead of experimenter-controlled retrieval-practice phases (e.g., Cuc et al., 2007). For instance, listeners might be more motivated to engage in effortful retrieval along with speakers under such more naturalistic conditions. Future studies, potentially even with novel experimental setups, are necessary to examine these issues.

On the basis of the present proposal that inhibition contributes more to RIF in speakers than listeners, the question arises of why RIF was not also larger in speakers than listeners in Experiments 1 and 2. In fact, if speakers and listeners differed in degree of inhibition but were equivalent in degree of blocking, then speakers should show larger RIF than listeners regardless of testing format. Following a finding by Kliegl and Bäuml (2016), who reported reduced intralist interference after retrieval practice relative to restudy, Rupperecht and Bäuml (2016) recently argued that retrieval practice may reduce possible blocking effects at test. If so, and if speakers engaged in effortful retrieval practice but listeners in less effortful recognition (see above), the contribution of blocking to RIF might be reduced in speakers relative to listeners. In such case, both inhibition and blocking may contribute to RIF in speakers and listeners, but inhibition may contribute more to RIF in speakers than listeners, and blocking contribute more to RIF in listeners than speakers, which is consistent with the present results. Clearly, more work is needed to examine this proposal in depth.

The conclusion arising on the basis of the present experiments is that differences between WI-RIF and SS-RIF are not qualitative, but mainly quantitative in nature, i.e., both effects arise due to blocking and inhibition, but the two mechanisms differ in relative contributions between speakers and listeners. This perspective is also consistent with other recent work on selective retrieval. Although many studies on RIF in individuals demonstrate that selective retrieval of some studied items can cause forgetting of the nonretrieved items, some other recent studies indicate that, under certain circumstances, selective retrieval can also facilitate recall of the nonretrieved information (for reviews, see Bäuml, 2019; Bäuml et

al., 2017). Corresponding studies often applied lists of unrelated items as study material and induced a contextual change between study and retrieval practice, for instance, by manipulating delay between study and selective retrieval. Typically, RIF was present when delay was short, but recall facilitation arose when delay was long. This facilitation was explained by context retrieval, assuming that the initial learning context becomes less accessible with delay and that selective retrieval then reactivates the study context, thereby facilitating recall of the other information (see also Polyn & Kahana, 2008). Corresponding facilitation effects were also observed in social settings, such that the delayed selective retrieval of speakers also improved recall of the listeners (Abel & Bäuml, 2015; for related work on collaborating groups, see Abel & Bäuml, 2017). Similar to the present findings, these results suggest that, qualitatively, the cognitive mechanisms active during selective retrieval are equivalent in speakers and listeners.

In sum, the results of the present study show that selective retrieval carried out by speakers can induce forgetting in listeners over a variety of memory tests. For test formats without item-specific cues, like category-cued recall, the forgetting can be equivalent in size between speakers and listeners, whereas for test formats with strong item-specific cues, like item recognition, the forgetting in listeners can be reduced. On the basis of the two-factor account of RIF, these findings suggest that the involvement of inhibitory processes in RIF is reduced in listeners relative to speakers, even though inhibitory processes still seem to contribute to the observed forgetting.

Open practices statement

The data and materials for all experiments are available on the Open Science Framework (<https://osf.io/y9q37/>). None of the experiments were preregistered.

Authors' Note We thank C. Hansbauer, L. Holzer, L. Leue and A. Nickl for their help with data collection.

References

- Abel, M., & Bäuml, K.-H. T. (2015). Selective memory retrieval in social groups: When silence is golden and when it is not. *Cognition*, *140*, 40–48.
- Abel, M., & Bäuml, K.-H. T. (2017). Collaborative remembering revisited: Study context access modulates collaborative inhibition and later benefits for individual memory. *Memory & Cognition*, *45*, 1319–1334.
- Abel, M., & Roediger, H. L. III. (2018). The testing effect in a social setting: Does retrieval practice benefit a listener? *Journal of Experimental Psychology: Applied*, *24*, 347–359.

- Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415–445.
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063–1087.
- Anderson, M. C., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In Benjamin, A. S. (Ed.) *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork*, (pp. 107–132). New York: Psychology Press.
- Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2015). *Memory*. London: Psychology Press.
- Barber, S. J., & Mather, M. (2012). Forgetting in context: The effects of age, emotion, and social factors on retrieval-induced forgetting. *Memory & Cognition*, *40*, 874–888.
- 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. T. (2019). Context retrieval as a critical component in selective memory retrieval. *Current Directions in Psychological Science*, *28*, 177–182.
- Bäuml, K.-H. T., & Kliegl, O. (2017). Retrieval-induced remembering and forgetting. In Wixted, J. T., & Byrne, J. H. (Eds.) *Cognitive psychology of memory, vol. 2 of learning and memory: A comprehensive reference*. (2nd ed., pp. 27–51). Oxford: Academic Press.
- Bäuml, K.-H. T., Aslan, A., & Abel, M. (2017). The two faces of selective memory retrieval - cognitive, developmental, and social processes. In Ross, B. (Ed.) *The psychology of learning and motivation*, Vol. 66: Academic Press: Elsevier Inc.
- Buchin, Z., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *43*, 1934–1947.
- Coman, A., & Hirst, W. (2015). Social identity and socially shared retrieval-induced forgetting: The effects of group membership. *Journal of Experimental Psychology: General*, *144*, 717–722.
- Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable through conversation: Socially shared retrieval induced forgetting of 9/11 memories. *Psychological Science*, *20*, 627–633.
- Cuc, A., Koppel, J., & Hirst, W. (2007). Silence is not golden – a case for socially shared retrieval-induced forgetting. *Psychological Science*, *18*, 727–733.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–478.
- Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, *111*, 524–542.
- Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition*, *33*, 1431–1441.
- Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review*, *11*, 125–130.
- Hirst, W., & Coman, A. (2018). Building a collective memory: The case for collective forgetting. *Current Opinion in Psychology*, *23*, 88–92.
- Hirst, W., & Echterhoff, G. (2012). Remembering in conversations: The social sharing and reshaping of memories. *Annual Review of Psychology*, *10*, 55–79.
- 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.
- Kliegl, O., & Bäuml, K.-H. T. (2016). Retrieval practice can insulate items against intralist interference: Evidence from the list-length effect, output interference, and retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 202–214.
- Koppel, J., Wohl, D., Meksin, R., & Hirst, W. (2014). The effect of listening to others remember on subsequent memory: The roles of expertise and trust in socially shared retrieval-induced forgetting and social contagion. *Social Cognition*, *32*, 148–180.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*, (2nd ed.). London: Lawrence Erlbaum Assoc Inc.
- Mannhaupt, H.-R. (1983). Produktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien. *Sprache und Kognition*, *2*, 264–278.
- Mulligan, N. W., & Picklesimer, M. E. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *42*, 938–950.
- 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.
- Ortega, A., Gómez-Ariza, C. J., Román, P., & Bajo, 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 Roediger, H. L. III. (Ed.) *Cognitive psychology of memory of learning and memory: A comprehensive reference*, (Vol. 2, pp. 389–416). Oxford: Elsevier.
- Polyn, S. M., & Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences*, *12*, 24–30.
- 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.
- 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.
- Rupprecht, J., & Bäuml, K.-H. T. (2016). Retrieval-induced forgetting in item recognition: Retrieval specificity revisited. *Journal of Memory and Language*, *86*, 97–118.
- Rupprecht, J., & Bäuml, K.-H. T. (2017). Retrieval-induced versus context-induced forgetting: Can restudy preceded by context change simulate retrieval-induced forgetting? *Journal of Memory and Language*, *93*, 259–275.
- Scheithe, K., & Bäuml, K.-H. (1995). Deutschsprachige Normen für Vertreter von 48 Kategorien. *Sprache und 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., & Steyvers, M. (1997). A model for recognition memory: REM - retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.
- 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.
- Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*, 400–408.

- 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.
- Spitzer, B., & Bäuml, K.-H. (2009). Retrieval-induced forgetting in a category recognition task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 286–291.
- Stone, C. B., Barnier, A. J., Sutton, J., & Hirst, W. (2010). Building consensus about the past: Schema consistency and convergence in socially shared retrieval-induced forgetting. *Memory*, *18*, 170–184.
- Stone, C. B., Barnier, A. J., Sutton, J., & Hirst, W. (2013). Forgetting our personal past: Socially shared retrieval-induced forgetting of autobiographical memories. *Journal of Experimental Psychology: General*, *142*, 1084–1094.
- Storm, B. C., & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition*, *40*, 827–843.
- Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *1*, 321–334.
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*, 289–335.
- Verde, M. F. (2013). Retrieval-induced forgetting in recall: Competitor interference revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1433–1448.
- 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 Nairne, J. S. (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 and Review*, *11*, 616–641.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.