Mem Cogn
DOI 10.3758/s13421-012-0224-2

Less we forget: Retrieval cues and release from retrieval-induced forgetting

Tanya R. Jonker · Paul Seli · Colin M. MacLeod

Abstract Retrieving some items from memory can impair the subsequent recall of other related but not retrieved items, a phenomenon called *retrieval-induced forgetting* (RIF). The dominant explanation of RIF—the inhibition account—asserts that forgetting occurs because related items are suppressed during retrieval practice to reduce retrieval competition. This item inhibition persists, making it more difficult to recall the related items on a later test. In our set of experiments, each category was designed such that each exemplar belonged to one of two subcategories (e.g., each BIRD exemplar was either a *bird of prey* or a *pet bird*), but this subcategory information was not made explicit during study or retrieval practice. Practicing retrieval of items from only one subcategory led to RIF for items from the other subcategory when cued only with the overall category label (BIRD) at test. However, adapting the technique of Gardiner, Craik, and Birtwistle (Journal of Learning and Verbal Behavior 11:778–783, 1972), providing subcategory cues during the final test eliminated RIF. The results challenge the inhibition account's fundamental assumption of cue independence but are consistent with a cue-based interference account.

Keywords Memory · Recall · Interference/inhibition in memory retrieval

T. R. Jonker (✉) · P. Seli · C. M. MacLeod
University of Waterloo,
Waterloo, Canada
e-mail: trjonker@uwaterloo.ca

P. Seli
e-mail: pseli@uwaterloo.ca

C. M. MacLeod
e-mail: cmacleod@uwaterloo.ca

Published online: 26 June 2012

Less we forget: Retrieval cues and release from retrieval-induced forgetting

When we remember, does the output directly reflect the internal representation? Research suggests that it does not, demonstrating that the process of retrieval can shape the product of retrieval (e.g., reproducing visual forms, Hanawalt & Demarest, 1939; release from proactive interference, Gardiner, Craik, & Birtwistle, 1972; context reinstatement in directed forgetting, Sahakyan & Kelley, 2002). In fact, slight changes to the retrieval context can produce dramatically different outcomes, even providing access to memories that would otherwise seem forgotten. As a consequence, addressing the state of a memory itself is challenging, because the retrieval context during test influences the output.

To demonstrate the complexity of the relation between the information in memory and the retrieval outcome, consider the classic study by Carmichael, Hogan, and Walter (1932). Their participants studied ambiguous shapes, each paired with one of two labels (e.g., two circles connected by a line were labeled either "eye-glasses" or "dumbbells"). Participants' later reproductions of the forms were biased by the studied label, such that the label "eye-glasses" resulted in a form that looked more like glasses, whereas the label "dumbbells" resulted in a form that looked more like free weights. On the basis of this study, one could conclude—as Carmichael et al. did—that the label given during study shaped the encoded memory representation of the ambiguous form. However, a less well-known article by Hanawalt and Demarest (1939) challenged this interpretation. They provided the labels during test, rather than encoding, and demonstrated precisely the same results. Thus, participants' reproductions of the forms were altered by the test cues and, therefore, were not a pure reflection of the internal memory representation. In this case, a

premature conclusion about the state of the memory representation was made before the retrieval process was fully considered.

Might this same interpretational problem occur in other well-known paradigms within the realm of memory? Our argument is that it does. Specifically, we argue that it applies to the retrieval-induced forgetting (RIF) paradigm. RIF is somewhat of a paradoxical finding: Retrieving a subset of items can result in a cost to related but nonretrieved items on a subsequent memory test (Anderson, Bjork, & Bjork, 1994). For example, retrieving the item *peach* from memory might impair later recall of other fruit items, like *banana*. Thus, retrieval is simultaneously beneficial for the retrieved items (*peach*) and detrimental for the related nonretrieved items (*banana*).

Investigations of RIF use a standard procedure that typically involves study of category–exemplar word pairs (e.g., FRUIT– *peach,* FRUIT–*banana,* INSECT–*wasp*), retrieval practice of a subset of these pairs, and a final category-cued recall test for all studied exemplars. During the intervening retrieval practice phase, participants practice half of the items from half of the categories by completing category-cued word stems (e.g., FRUIT–*pe*___; no practice for *banana* or *wasp*). This manipulation produces three types of items: (1) practiced items (*peach*), labeled Rp+; (2) unpracticed items from practiced categories (*banana*), labeled Rp−; and (3) items from categories for which no items are practiced (*wasp* from the category INSECT), labeled Nrp and serving as a baseline for category recall.

During the later recall test, participants are given category cues to prompt recall of all studied exemplars. This test typically reveals two things. The first is a benefit for Rp+ items relative to Nrp items. This finding is intuitive: The additional practice of Rp+ items improves later recall of these items. Interestingly, however, the test also reveals a cost for Rp− items relative to Nrp items. Thus, practice of some items (Rp+) impairs later recall of nonpracticed related items (Rp−).

Many have argued that this cost to Rp− items occurs because the representations of the Rp− items themselves are altered during retrieval practice (see Anderson, 2003, for a review). Specifically, the dominant theory asserts that when a category label is presented during retrieval practice, strongly related exemplars are routinely activated and compete for retrieval, creating substantial retrieval interference. These competing exemplars are then suppressed to reduce the retrieval interference. Critically, the suppression of these competing exemplars is enduring, making them less accessible on the later recall test. This is the inhibition account of RIF.

The inhibition account is the dominant explanation of RIF because it can explain a number of unique findings. For example, Anderson et al. (1994) demonstrated that Rp− items with high taxonomic frequency showed RIF but that Rp− items with lower taxonomic frequency showed no RIF. This finding is supportive of the inhibition account because, ostensibly, Rp− items with lower taxonomic frequency are not strongly related to the category cue and, therefore, are less likely to compete during practice, decreasing the need for and, hence, the likelihood of inhibition. In fact, Aslan and Bäuml (2010) suggested that an executive control mechanism may be recruited to control this retrieval interference based on their finding that working memory was positively correlated with RIF (see also Román, Soriano, Gómez-Ariza, & Bajo, 2009).

Although the abovementioned findings (and others) support the inhibition account, there is one key finding that has been cited as the definitive evidence that inhibition underlies RIF. This is the finding of cue independence. The inhibition account assumes that the representations of Rp− items are suppressed during retrieval practice and that any cue—not just the studied category cue—used on a later test to access that suppressed representation will reveal RIF (Anderson & Spellman, 1995). So, if a participant studies FRUIT–*banana* and it is an Rp− items, this exemplar could be tested using MONKEY– *b*___ and still be expected to show reduced memory. Similarly, RIF should occur on a recognition test for the items alone (i.e., without the category cue), and some have shown that it does (e.g., Hicks & Starns, 2004; Spitzer & Bäuml, 2007; but see Koutstaal, Schacter, Galluccio, & Stofer, 1999). Put simply, the inhibition account asserts that an earlier experience shaped the memory representation, and the later recall test reflects this change independently of the cue.

The present study

In the present article, we focus on the assumption of cue independence, because it is cited as the key support for the inhibition account as an explanation for RIF. Indeed, Anderson and Levy (2007)—strong proponents of the inhibition account in RIF—argued that "to make a strong claim in any study about the presence or absence of inhibition, or about variations in the magnitude of inhibition as a function of condition or population, it is necessary to include an independent probe of the impaired items' accessibility" (p. 82).

Although cue independence is arguably the "gold standard" for assessing the presence of inhibition, it is, in fact, only rarely reported in the now quite extensive RIF literature. Furthermore, in cases where independent cues have been used, results have been mixed (e.g., Aslan, Bäuml, & Pastötter, 2007; Camp, Pecher, & Schmidt, 2007; Williams & Zacks, 2001). This has led some researchers to question the actual independence of nominally "independent" cues, arguing that participants likely use covert cuing during these tests by essentially self-providing the originally studied category cues as mediators (Camp, Pecher, Schmidt, &

Zeelenberg, 2009; Perfect et al., 2004; but see Huddleston & Anderson, 2012). Furthermore, Perfect et al. demonstrated cue *dependence* by assigning two different cues to the exemplar during study (i.e., a category cue and a face cue) and then using either one cue or both during retrieval practice and, again, one cue or both during final test. Using this method, Perfect et al. found that the occurrence of RIF depended on overlap of the cues used during retrieval practice and during test.

Thus, if cue independence is the lynchpin for the inhibition account, it is a shaky one. With this in mind, we decided to test the cue independence of RIF by introducing a new type of cuing to the RIF paradigm. Our methods were inspired by a classic study by Gardiner et al. (1972), in which they investigated the buildup and release of proactive interference when multiple short lists were studied. In the prototypical version of the release from proactive interference task, participants saw three or four related items on each trial and then tried to recall them after brief distraction. Recall performance declined as proactive interference built up across three related trials. Then, on the fourth trial, a fourth set of related items could be presented or a set of unrelated items could be presented. Performance on the fourth trial continued to fall for the related items but improved sharply for the unrelated items. This was held to be evidence that a particular dimension, or category, was being encoded (Wickens, 1970).

To test this encoding idea, Gardiner et al. (1972) presented four successive lists of to-be-recalled items from the same category (e.g., *games*). Critically, items presented on the first three lists belonged to one subcategory (e.g., *indoor games*), whereas those on the fourth list belonged to another subcategory (e.g., *outdoor games*). Some participants were informed of the subtle subcategory change prior to studying the fourth list, others were informed after studying the fourth list and prior to recall, while others were not informed of the subcategory change. Gardiner et al. found a significant and equivalent release from proactive interference for the two informed groups but no release for the uninformed group. Because performance benefited equally from notifying participants of the subcategory change before *and* after, Gardiner et al. concluded that the buildup and release from proactive interference resulted not from differential encoding, but from the availability of an effective retrieval cue at test.

Drawing on this logic, we provided some participants in our experiments with discriminatory subcategory cues at test. If the representations of Rp− items are indeed inhibited, RIF should be observed irrespective of the cues (i.e., cue independence). Alternatively, Rp− items might be forgotten because the strengthened Rp+ items interfere with Rp− recall. If this is the case, RIF should be cue dependent. Gardiner et al. (1972) found that the broad category cue was not helpful for recall of the items from the fourth and final list. However, when participants were provided with a cue that was specific to the items that would otherwise be forgotten (i.e., the fourth-list items), they were able to overcome the interference from the other lists of related items.

In our experiments, some participants received only the category cue on the final test (BIRD), some received the category cue and subcategory information that was specific to the Rp− items, and others received the category cue and subcategory information that cued both Rp+ and Rp− items. We predicted that participants' recall of Rp− items would improve only when additional cues discriminated between Rp+ and Rp− items. When subcategory information was provided at test in the study by Gardiner et al. (1972), proactive interference still occurred for the different lists from the same subcategory; only the list from a unique subcategory benefited from the subcategory information. Similarly, providing participants with subcategory information during the final test should eliminate RIF only when the subcategory cues are specific to the Rp− items, because these cues would not be contaminated by any interference from the Rp+ items.

Subcategory information has been employed in the RIF paradigm before by Bäuml and Hartinger (2002) and by Bäuml, Zellner, and Vilimek (2005), but using a different approach. They presented subcategory information during both the study phase (e.g., ANIMAL–*predator*–*tiger*) and the test phase, a manipulation intended to encourage integration between similar items during encoding. This manipulation emphasized the similarity of exemplars from the same subcategory, while emphasizing the dissimilarity of exemplars from different subcategories. Their experiments revealed that when Rp− items were highly similar to Rp+ items through shared subcategory membership, RIF did not occur, whereas when Rp− items were dissimilar from Rp+ items, RIF did occur.

Our approach differs from that of Bäuml and colleagues (Bäuml & Hartinger, 2002; Bäuml et al., 2005) because we employed subcategory information at test only; during study, participants in our experiments were not made explicitly aware of the subcategorization of exemplars. Thus, our participants and Bäuml and colleagues' participants likely encoded the stimuli quite differently. Indeed, it has been shown repeatedly that encouraging semantic or episodic similarity during encoding makes similar items resistant to RIF (e.g., Anderson, Green, & McCulloch, 2000). So, whereas Bäuml and colleagues were interested in the effects of encoding differences on later recall, analogous to Carmichael et al. (1932), we were interested in test cues and retrieval processes, analogous to Hanawalt and Demarest (1939).

Indeed, our testing approach also highlights the difference between our experiments and previous tests of the cue independence assumption. In previous experiments, this assumption has been tested by substituting a novel test cue

for the studied cue. In our experiments, we *added* a novel cue to the study cue. Thus, our paradigm circumvents arguments about covert cuing (e.g., Camp et al., 2009; Huddleston & Anderson, 2012) because the studied cue was provided. Furthermore, our approach differs from previous experiments demonstrating the cue dependence of RIF because we made no changes to the cue–exemplar association or to the standard retrieval practice phase; our test of cue dependence was restricted to the final test phase (cf. Camp et al., 2009; Perfect et al., 2004). Thus, in our experiments, we can ask the following: When all else is equal to the standard RIF paradigm, can the cues provided at test determine the presence or absence of RIF?

Experiments 1 and 2

To investigate the effect of subcategory cuing on RIF, we designed a stimulus set of categories in which half of the exemplars each belonged to one of two subcategories. These subcategories were not identified during study or during retrieval practice.

During the retrieval practice phase, some participants practiced items from only one of the two subcategories (*standard* and *pure subcategory* conditions), whereas others practiced items from both subcategories (*mixed subcategory* condition). On the later test, some participants were given explicit subcategory cues (pure subcategory and mixed subcategory conditions). The inhibition account predicts RIF in all three conditions, because RIF should be independent of the test cues. We hypothesized, however, that participants would show RIF only in the standard and mixed subcategory conditions, because both Rp+ and Rp− would be equivalently cued. In the pure subcategory condition, on the other hand, we predicted that participants would show release from RIF, because one of the two subcategory cues would provide access to the Rp− items uncontaminated by the stronger Rp+ items. This would be a demonstration of cue dependence.

Method

Participants Students were recruited from the University of Waterloo and received bonus course credit for their participation. All had normal or corrected-to-normal vision and English as their most fluent language. A total of 116 (34 males, 82 females) participated in Experiment 1, with ages ranging from 18 to 27 years (M = 19.5, SE = 0.16). A total of 89 (35 males, 54 females) participated in Experiment 2, with ages ranging from 18 to 45 years (M = 20.6, SE = 0.47).

Materials We first compiled a list of 11 categories, each with 12 exemplars. These stimuli were selected from the Battig and Montague's (1969) category norms or were generated by the authors. The resulting list of 132 category–exemplar pairs was presented to 28 undergraduate students who were recruited in the same manner as the aforementioned participants but who did not later participate in either experiment. These participants were asked to classify each exemplar as belonging to one of two subcategories or as belonging to neither given subcategory. For example, upon seeing the exemplar *liver*, from the category BODY PART, participants were provided three response options: (1) *joint*, (2) *organ*, or (3) *neither*.

On the basis of these classification judgments, we selected categories that met two criteria: (1) A category consisted of four exemplars from each of its two subcategories, with a minimum of 75% classification agreement, and (2) all exemplars within a category had a unique initial letter. When more than four exemplars met these criteria, we selected the four exemplars that were most frequently classified as belonging to a subcategory. Eight categories met these criteria and were included in the final stimulus set (see Table 1 for example stimuli). Each participant studied from six randomly selected categories (of the eight), with eight exemplars in each category, for a total of 48 word pairs.

Procedure Because their procedures were highly similar, Experiments 1 and 2 are presented together. They differed only with respect to the final test.

During the study phase, participants saw category–exemplar word pairs; the subcategory information was not mentioned. Each pair was displayed individually for 5 s, followed by a blank screen for 500 ms. Presentation order was randomized for each participant; however, exemplars from the same category could not appear in succession. To limit the influence of primacy and recency effects (Murdock, 1962), six filler pairs from two different categories were included, three before and three following the 48 experimental stimuli.

Following study, participants practiced retrieval of half of the exemplars from half of the categories (12 items total),

Table 1 Exemplars and subcategories for the categories BODY PART and BIRD

Subcategory	Body Part		Bird	
	Joint	Organ	Bird of Prey	Pet Bird
Exemplar	Ankle	Heart	Eagle	Budgie
	Elbow	Liver	Falcon	Canary
	Knee	Pancreas	Hawk	Lovebird
	Wrist	Stomach	Vulture	Parrot

three times each. Retrieval practice involved category-cued word stem completions (e.g., BODY PART–*li*___). The word stems were presented in a constrained random order such that exemplars from the same category were never practiced in succession.

The four practiced exemplars were selected on the basis of condition. Participants in the pure subcategory condition and the standard condition practiced exemplars from the same subcategory (e.g., all of the *organ* exemplars but none of the *joint* exemplars). Consequently, for these participants, one subcategory consisted purely of Rp+, and the other of Rp−. In contrast, participants in the mixed subcategory condition practiced two exemplars from each subcategory (e.g., two *organ* and two *joint* exemplars); each subcategory was a mix of both Rp+ and Rp− items. During practice, participants were not made aware of the subcategorization of the exemplars. Following retrieval practice, participants completed a 5 min distractor task, during which they made a list of countries. This task was borrowed from Macrae and Roseveare (2002).

Experiment 1 To test the recall of the studied items in Experiment 1, participants were given a category-cued free recall test. They were provided either with the category cue only (standard condition; e.g., BODY PART) or with the category cue plus both subcategory cues (pure subcategory and mixed subcategory conditions; e.g., BODY PART, *organ*, *joint*). In the pure subcategory and mixed subcategory conditions, participants were told that they would see the category along with some extra cues related to the exemplars and were instructed to use these extra cues to assist their recall. The cue(s) remained on the screen for 30 s, during which time participants were to write a list of studied exemplars. Category test order was randomized across participants.

Experiment 2 Some investigators have suggested that, in a category-cued free recall test, RIF may occur because Rp+ items are output before Rp− items; Nrp items, on the other hand, are recalled across all output positions (Anderson et al., 1994). Therefore, any reduction in Rp− recall could be the result of these items suffering from output interference (see Roediger, 1974). To address this possibility, the test in Experiment 2 was designed to force recall of Rp− items prior to recall of Rp+ items. A one-letter word stem was presented along with the category cue to test each item individually. In those categories where retrieval practice had taken place, all of the Rp− exemplars were tested before any of the Rp+ exemplars (for similar procedures, see Bäuml & Hartinger, 2002; Raaijmakers & Jakab, 2012). Testing was blocked by category, such that all exemplars from one category were tested together. Both the order of the categories and the order of the items of each type within the category were randomized across participants.

Subcategory cues were presented along with the category cues in the pure subcategory and mixed subcategory conditions. The subcategory cue was item specific, so for the studied item *liver*, participants would see BODY PART–*organ*–*l*___; they would not see the subcategory cue *joint* as well for this word stem. Participants in the standard condition saw only the category cue and first letter (e.g., BODY PART–*l*___).

Results

An alpha level of .05 was used to evaluate all reported statistical outcomes. Recall data for each condition were analyzed using repeated measures analyses of variance and two follow-up comparisons. Follow-up analyses employed one-tailed repeated measures *t*-tests to examine the benefit of retrieval practice (Rp+ > Nrp) and any cost to related but unpracticed items (Rp− < Nrp).

Experiment 1

Participants had high success rates for retrieval practice. The proportions completed were .84 (*SE* = .02), .81 (*SE* = .02), and .83 (*SE* = .02) in the standard, mixed subcategory, and pure subcategory conditions, respectively.

Standard condition Overall, recall of Rp+, Rp−, and Nrp items differed, $F(2, 82) = 56.74$, $MSE = 0.02$, $p < .05$, $\eta_p^2 = .58$. As is clear in Fig. 1, retrieval practice aided the later recall of Rp+ items relative to Nrp items, $t(41) = 7.44$, $SE = .03$, $p < .05$, $d = 1.15$, but hurt the later recall of Rp− items relative to Nrp items, $t(41) = 2.47$, $SE = .03$, $p < .05$, $d = 0.38$.

Mixed subcategory condition Overall, recall of Rp+, Rp−, and Nrp items differed, $F(2, 68) = 55.04$, $MSE = 0.01$, $p < .05$, $\eta_p^2 = .62$. Paralleling the pattern in the standard condition, retrieval practice aided the later recall of Rp+ items relative to Nrp items, $t(34) = 8.90$, $SE = .02$, $p < .05$, $d = 1.50$, but hurt the later recall of Rp− items relative to Nrp items, $t(34) = 2.15$, $SE = .03$, $p < .05$, $d = 0.36$.

Pure subcategory condition Overall, recall of Rp+, Rp−, and Nrp items differed, $F(2, 76) = 29.78$, $MSE = 0.02$, $p < .05$, $\eta_p^2 = .44$. Retrieval practice aided the later recall of Rp+ items relative to Nrp items, $t(38) = 6.36$, $SE = .03$, $p < .05$, $d = 1.02$, but in this condition there was no difference in the recall of Rp− and Nrp items, $t(38) = 0.18$, $SE = .03$, $p = .86$.

Experiment 2

Participants had high success rates for retrieval practice. The proportions completed were .84 (*SE* = .03), .81 (*SE* = .02),

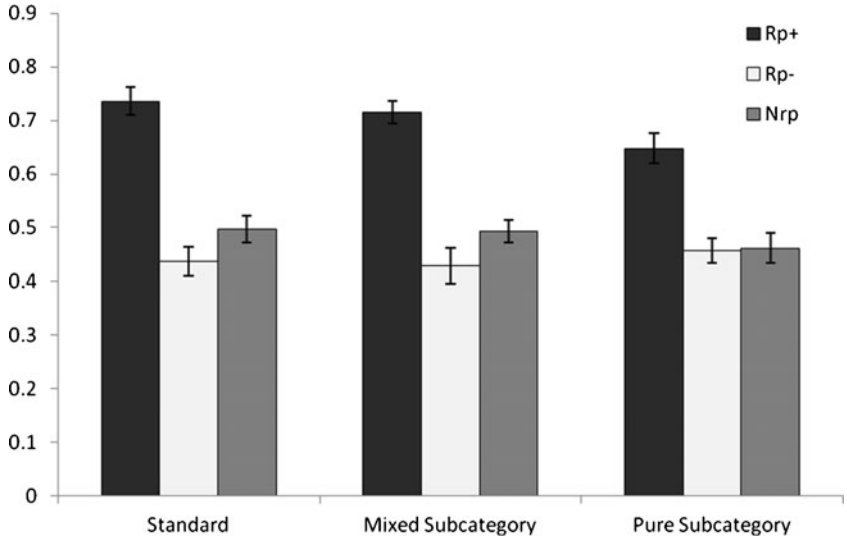

Fig. 1 Experiment 1: Mean correct recall proportions of exemplars on the final cued recall test as a function of type of cue. The error bars represent one standard error of their respective means

and .87 (*SE* = .02) in the standard, mixed subcategory, and pure subcategory conditions, respectively.

Because the Rp− items were tested before the Rp+ items in each practiced category, the Nrp items from each unpracticed category were separated to allow for comparison across similar test order position. The first four Nrp items from each category (Nrp1) were compared with Rp−, and the remaining four Nrp items (Nrp2) were compared with Rp+. Nrp2 recall was expected to be lower than Nrp1 recall because Nrp2 items are tested in the latter half of their category and should suffer from a buildup of output interference.

Standard condition Overall, the recall of Rp+, Rp−, Nrp1, and Nrp2 items differed, $F(3, 81) = 8.07$, $MSE = 0.02$, $p < .05$, $\eta_p^2 = .23$. As is clear in Fig. 2, retrieval practice aided the later recall of Rp+ items relative to Nrp2 items, $t(27) = 4.37$, $SE = .03$, $p < .05$, $d = 0.83$, but hurt the recall of Rp− items relative to Nrp1 items, $t(27) = 1.73$, $SE = .05$, $p < .05$, $d = 0.33$.

Mixed subcategory condition Overall, the recall of Rp+, Rp−, Nrp1, and Nrp2 items differed, $F(3, 141) = 21.78$, $MSE = 0.02$, $p < .05$, $\eta_p^2 = .32$. Paralleling the pattern of the standard condition, retrieval practice aided the later recall of Rp+ items relative to Nrp2 items, $t(47) = 7.64$, $SE = .03$, $p < .05$, $d = 1.10$, but hurt the recall of Rp− items relative to Nrp1 items, $t(47) = 1.67$, $SE = .03$, $p = .05$, $d = 0.24$.

Pure subcategory condition Overall, the recall of Rp+, Rp−, Nrp1, and Nrp2 items differed, $F(3, 78) = 7.28$, $MSE = 0.02$, $p < .05$, $\eta_p^2 = .22$. Retrieval practice aided the later recall of Rp+ items relative to Nrp items, $t(26) = 4.48$, $SE = .04$, $p < .05$, $d = 0.86$, but there was no difference in the recall of Rp− and Nrp items, $t(26) = 0.66$, $SE = .04$, $p = .51$.

The subcategory cues also helped recall overall; Nrp recall in the mixed subcategory and pure subcategory conditions ($M = .72$) was better than that in the standard condition ($M = .64$), $t(101) = 2.31$, $SE = .03$, $p < .05$, $d = 1.99$.

Discussion

A clear pattern of results emerged from Experiments 1 and 2: When provided with no subcategory information, participants had poorer recall of the Rp− items (standard condition); when provided with subcategory information on the final test, participants still showed poorer recall of Rp− items if those subcategory cues referenced both Rp+ and Rp− items (mixed subcategory condition); however, if the subcategory cues referenced only Rp− items, participants no longer showed RIF (pure subcategory condition).

Of critical interest is the finding that subcategory information itself is not beneficial. If subcategory information alone were beneficial, we should see a release from RIF in both the mixed subcategory and pure subcategory conditions. Instead, subcategory cues were helpful only when they discriminated between Rp− and Rp+ items. To ensure that this novel finding was not spurious, we replicated the mixed subcategory and pure subcategory conditions.

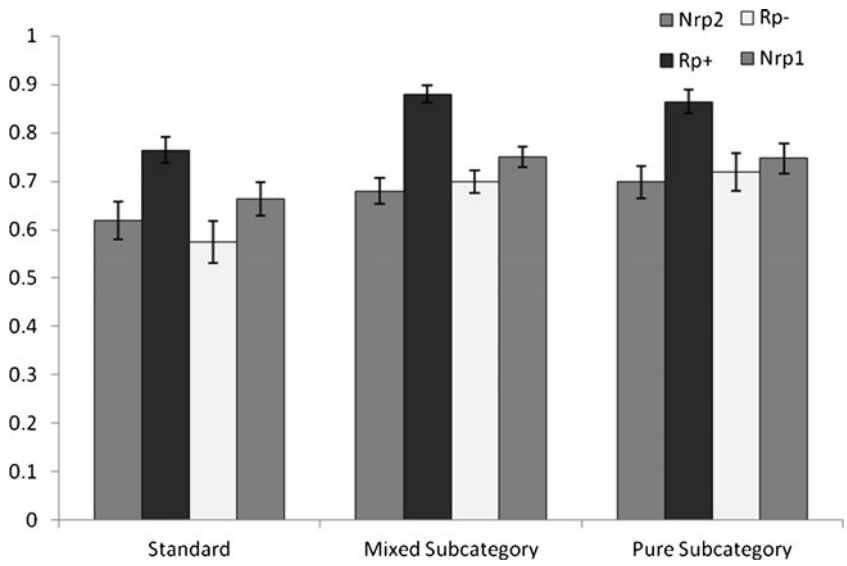

Fig. 2 Mean correct recall proportions of exemplars on the final recall test in Experiment 2. The error bars represent one standard error of their respective means. Nrp1 and Rp− exemplars come from comparable recall positions in the test list, while Nrp2 and Rp+ exemplars come from comparable positions in the test list

Replication

Method

The mixed subcategory and pure subcategory conditions from Experiment 2 were replicated with 85 students from the University of Waterloo (32 males, 53 females; age, M = 20.2 years, SE = 0.30). None had participated in Experiment 1 or 2.

Results

Success rates for retrieval practice were appropriate, with correct word stem completion proportions of .87 (SE = .02) and .83 (SE = .02) in the mixed subcategory and pure subcategory conditions, respectively.

Mixed subcategory condition Overall, the recall of Rp+, Rp−, Nrp1, and Nrp2 items differed, $F(3, 99) = 4.77$, $MSE = 0.02$, $p < .05$, $\eta_p^2 = .13$. Replicating the pattern of Experiments 1 and 2, participants recalled more Rp+ items (M = .79) than Nrp2 items (M = .70), $t(33) = 2.61$, $SE = .03$, $p < .05$, $d = 0.45$, and fewer Rp− items (M = .70) than Nrp1 items (M = .79), $t(33) = 3.00$, $SE = .03$, $p < .05$, $d = 0.51$.

Pure subcategory condition Overall, the recall of Rp+, Rp−, Nrp1, and Nrp2 items differed, $F(3, 108) = 12.14$, $MSE = 0.02$, $p < .05$, $\eta_p^2 = .25$. Replicating the pattern of Experiments 1 and 2, participants recalled more Rp+ items (M = .85) than Nrp2 items (M = .69), $t(36) = 5.80$, $SE = .03$, $p < .05$, $d = 0.95$, but there was no difference in the recall of Rp− items (M = .71) and Nrp1 items (M = .71), $t(36) = 0.06$, $SE = .03$, $p = .95$.

Interaction analysis

In this set of experiments, we consistently found RIF in the standard and mixed subcategory conditions and a lack of RIF in the pure subcategory condition, demonstrating that this pattern is replicable. We also sought to assess the differential RIF pattern statistically. To provide ourselves with the necessary power for this between-subjects analysis, we combined the data from Experiment 2 and the replication for the mixed subcategory and pure subcategory conditions. Difference scores were calculated for each participant (Nrp minus Rp−) to indicate the degree of forgetting. A one-tailed independent samples t-test comparing these difference scores from the mixed subcategory (M = .01, SE = .03) and pure subcategory (M = .07, SE = .02) conditions revealed a significant difference in RIF, $t(144) = 1.64$, $SE = .03$, $p = .05$, $d = 0.27$. This analysis effectively indicates an interaction such that there was RIF in the mixed subcategory condition but not in the pure subcategory condition, despite the fact that both of these conditions featured subcategory cues on the final test.

Discussion

In explaining RIF, the most influential theory asserts that forgetting on the final test occurs because an earlier

process—inhibition—changed the memory representation (see Anderson, 2003); thus, the final test measures the status of the memory representation and should be cue independent.

Our findings challenge this theory. In three cases, the standard RIF effect occurred when participants were given either category information alone (standard condition) or additional subcategory cues that referred to both Rp+ and Rp− items (mixed subcategory condition), but RIF was absent when participants were provided with subcategory cues that referred to the Rp− items alone (pure subcategory condition). Thus, much like Gardiner et al. (1972), we found a release from a standard forgetting effect when participants were given discriminatory cues at test, highlighting the cue-dependent nature of RIF and, hence, the pivotal role of the retrieval context.

In a recent article, Huddleston and Anderson (2012) argued that—in spite of a demonstration of covert cuing by Camp et al. (2009)—the independent cuing method supports the inhibition account. Specifically, they argued that covert cuing will occur only when the studied cue and the independent cue are semantically related but will not occur when semantic relatedness is controlled. Although Camp et al. (2009) found support for covert cuing using independent cues, Huddleston and Anderson showed that the independent cues used by Camp et al. (2009) were judged as being more semantically related than their own set of cues. On the basis of their findings, Huddleston and Anderson argued for the preservation of independent cues as a "diagnostic tool in research on inhibition" (p. 8). However, our experiments employed an entirely different type of cuing—by providing additional cues that occurred only at final retrieval, rather than by substituting a studied cue for a novel cue—and yet our results converge with the findings of cue dependence by Camp et al. (2009) and Perfect et al. (2004). Therefore, we conclude that it is difficult to maintain the argument that RIF is caused by cue-independent inhibition.

It is informative to contrast our study, where subcategorization was used only at test, with the studies of Bäuml and colleagues (Bäuml & Hartinger, 2002; Bäuml et al., 2005), where subcategorization was used during both study and test. In our experiments, participants were not privy to the subcategorization of the items during encoding. If participants had inferred subcategorization during study, we should have seen the absence of RIF in the mixed subcategory condition, which parallels the dissimilar condition of Bäuml and colleagues and the presence of RIF in the pure subcategory condition, which parallels their similar condition. Thus, when contrasting our results with those from previous research, we can conclude that the similarity results obtained by Bäuml and colleagues depend on the emphasis of similarity during encoding, rather than on any similarity that is inherent in the semantic representations of the items themselves.

Alternative accounts of retrieval-induced forgetting

As a strongly supported theory with few contenders, it is important to examine the central assumptions of the inhibition account closely. In doing so, we found that subcategory cuing—a procedure borrowed from Gardiner et al. (1972)—produced results that cannot be accounted for by the inhibition account.

Although our goal in this study was not to provide and test an alternative account, we think that it is worthwhile to acknowledge other explanations of RIF. Indeed, a growing body of literature challenges the dominant inhibition account (for a review, see Verde, 2012; e.g., Dodd, Castel, & Roberts, 2006; Jakab & Raaijmakers, 2009; Jonker & MacLeod, in press; Raaijmakers & Jakab, 2012; Verde & Perfect, 2011; Williams & Zacks, 2001), and call for a new theoretical explanation.

To explore possible alternatives, we look at the effect of subcategory cues in our experiments. In our experiments, subcategory cues only benefited the recall of Rp− items in the pure subcategory condition. The presence of RIF in the mixed subcategory condition suggests that subcategory information alone is not sufficient to eliminate RIF. Instead, subcategory cues eliminated RIF when they served as a retrieval cue that was uncontaminated by the history of the Rp+ items (unlike the usual superordinate category cue). With this in mind, we will outline two possible alternative explanations for RIF.

The first is a strength-based competition model (e.g., Raaijmakers & Jakab, 2012; search of associative memory model [SAM], Raaijmakers & Shiffrin, 1981). According to this account, strengthening the association between the category and some exemplars can make it more difficult to recover other, unstrengthened exemplars on a later test, due to interference caused by the category label. Specifically, when a subset of items is practiced, the category–exemplar association is strengthened for those items. During a later test, the practiced items interfere with the Rp− items because the category cue is more strongly associated to these Rp+ items, perhaps even modifying the meaning of the cue in favor of the Rp+ items. In the mixed subcategory condition, the category label and the mixed subcategory cues reference both Rp+ and Rp− items. Thus, the strong Rp+ items likely produce interference when these cues are used, leading to difficult retrieval of the Rp− items. In our pure subcategory condition, on the other hand, the subcategory cue for the Rp− items is uncontaminated by the strengthened Rp+ items, so the Rp− items are free from interference and more easily retrieved.

The second possible explanation addresses the role of context change in RIF (for speculations on the role of context in RIF, see Perfect et al., 2004; Verde & Perfect, 2011). Indeed, many researchers have speculated on the

importance of context in recall paradigms (e.g., the SAM model; Raaijmakers & Shiffrin, 1981). A context change account of RIF acknowledges the role that episodic memory plays during recall. Specifically, when performing the final test, a participant might treat the study and retrieval practice phases as distinct learning contexts. The participant then might rely on memory for these learning contexts when searching through memory. If memory focuses on the retrieval practice phase, access to the Rp− items might be relatively difficult, thereby leading to forgetting. In contrast, the uncontaminated cues used in the pure subcategory condition might help the participant to access the earlier study context, facilitating recall of the otherwise difficult-to-retrieve Rp− items.

Conclusion

Our findings are problematic for the inhibition account because they directly challenge the cue independence assumption, which is at the heart of the theory. Our results instead emphasize the *cue-dependent* nature of RIF. After demonstrating that providing a label such as "dumbbells" at test altered the output of the ambiguous form, Hanawalt and Demarest (1939) stated that "we can no longer assume that a change in reproduction is a direct representation of an identical change in the trace" (p. 159), a retrieval interpretation that challenges the encoding interpretation put forward by Carmichael et al. (1932). On the basis of our findings, we see a similar assertion with regard to RIF as warranted: We cannot assume that impaired recall of Rp− items is a direct indicator of earlier suppression of the memory representation. Our results demonstrate that RIF is a retrieval-dependent phenomenon, a finding that is in direct conflict with the inhibition account.

Author Note This research was supported by postgraduate scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC) to T.R.J. and P.S. and by an NSERC discovery grant to C.M.M. We thank Gus Craik for providing the stimuli used in the Gardiner, Craik, and Birtwistle (1972) experiments, which facilitated the development of our stimulus set. We also thank Trish Varao Sousa, Sukhdip Grewal, Nivedita Sharma, Melinda Recchia, Christine Li, and Sheldon Frey for their assistance with data collection and Wayne Podrouzek and Andrea Hughes for their comments on the work.

References

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

Anderson, M. C., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 81–102). Washington, DC: American Psychological Association.

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. (2010). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/a0021324. Advance online publication.

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

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1–46.

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

Bäuml, K.-H., Zellner, M., & Vilimek, R. (2005). When remembering causes forgetting: Retrieval-induced forgetting as recovery failure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1221–1234.

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.

Camp, G., Pecher, D., Schmidt, H. G., & Zeelenberg, R. (2009). Are independent probes truly independent? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 934–942.

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology, 15,* 73–86.

Dodd, M. D., Castel, A. D., & Roberts, K. E. (2006). A strategy disruption component to retrieval-induced forgetting. *Memory & Cognition, 34,* 102–111.

Gardiner, J. M., Craik, F. I. M., & Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Learning and Verbal Behavior, 11,* 778–783.

Hanawalt, N. G., & Demarest, I. H. (1939). The effect of verbal suggestion in the recall period upon the reproduction of visually perceived forms. *Journal of Experimental Psychology, 25,* 159–174.

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

Huddleston, E., & Anderson, M. C. (2012). Reassessing critiques of the independent probe method of studying inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/a0027092. Online first publication.

Jakab, E., & Raaijmakers, J. G. W. (2009). The role of item strength in retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 607–617.

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

Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology and Aging, 14,* 220–237.

Macrae, C. N., & Roseveare, T. A. (2002). I was always on my mind: The self and temporary forgetting. *Psychonomic Bulletin & Review, 9,* 611–614.

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

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 and Language, 51,* 399–417.

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

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 95,* 434–455.

Roediger, H. L., III. (1974). Inhibiting effects of recall. *Memory & Cognition, 2,* 261–269.

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

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

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.

Verde, M. F. (2012). Retrieval-induced forgetting and inhibition: A critical review. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 56, pp. 47–80). New York: Academic Press.

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

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15.

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