

Is retrieval success a necessary condition for retrieval-induced forgetting?

BENJAMIN C. STORM, ELIZABETH L. BJORK, ROBERT A. BJORK, and JOHN F. NESTOJKO
University of California, Los Angeles, California

When information is retrieved from memory, it becomes more recallable than it would have been otherwise. Other information associated with the same cue or configuration of cues, however, becomes less recallable. Such retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994) appears to reflect the suppression of competing nontarget information, with this suppression facilitating the selection of target information. But is success at such selection a necessary condition for retrieval-induced forgetting? Using a procedure in which some cues posed an impossible retrieval task for participants, we report evidence that the attempt to retrieve, even if unsuccessful, can produce retrieval-induced forgetting. This finding, we believe, supports and refines a suppression/inhibitory account of retrieval-induced forgetting.

Forgetting is as underappreciated as it is annoying. As we (e.g., E. L. Bjork & R. A. Bjork, 1988; R. A. Bjork, 1989) and others have argued, forgetting is critical for the efficient and adaptive functioning of human memory. Without some means to suppress or set aside information that has become outdated or irrelevant, humans would become vulnerable to a devastating accumulation of proactive interference, greatly impeding the ability to learn and access current information. Stated more broadly, a memory system, to be efficient and adaptive, must have mechanisms that render desired or target information more accessible and potentially interfering information less accessible.

One such mechanism of recent interest is *retrieval-induced forgetting* (RIF; Anderson, Bjork, & Bjork, 1994)—the observation that selectively retrieving some information associated with a given cue or configuration of cues renders other information associated with the same cue(s) less accessible. The broad goal of the present research was to clarify more fully the processes that underlie such retrieval-induced forgetting. The more specific goal was to answer an important question: Is success at selective retrieval a necessary condition for retrieval-induced forgetting?

The Retrieval Practice Paradigm

The paradigm used to study RIF typically involves three phases: study, retrieval practice, and testing (An-

derson et al., 1994). First, participants study items such as category-exemplar pairs (e.g., *fruit-lemon*, *profession-accountant*), often consisting of six members of eight categories, presented one at a time in an interspersed order. The participants then retrieve half of the members from half of the categories via guided retrieval practice (e.g., *fruit-le _____*), with such practiced pairs called *Rp+* items. Members of the practiced category that do not receive retrieval practice (e.g., *orange*) are called *Rp-* items, and members of the nonpracticed categories are called *Nrp* items.

After a delay, the participants are asked to recall members of each category, typically via a category-cued recall task. Not surprisingly, *Rp+* items are better recalled than both *Rp-* and *Nrp* items. The more surprising and important finding is that *Rp-* items are often recalled less well than baseline *Nrp* items. Such a finding was and is surprising because there were and are reasons to expect *Rp-* items to profit, not suffer, from the retrieval of *Rp+* items. The spreading activation idea, for example, which is a prominent assumption in associative models of memory, seems to predict that strengthening some members of a given category should strengthen other members of that category as well. Covert rehearsal of other members of a given category, such as *orange* during the effort to retrieve *lemon*, might also be expected to increase the subsequent recall of *Rp-* items. The finding is important because it points to the role of active suppression processes in the updating and management of human memory.

Over the past decade, RIF has been demonstrated with a variety of materials and experimental contexts (e.g., visuo-spatial [Ciranni & Shimamura, 1999]; eyewitness [Shaw, Bjork, & Handal, 1995]; and social [Dunn & Spellman, 2003]). It is, therefore, well established that *Rp-* items suffer impaired recall, but how that impairment comes about remains not fully understood.

The authors thank the following undergraduate research assistants: Jennifer Storm, Gaby Pocasangre, Amin Saïar, and Susy Hassan. We greatly appreciate the constructive comments of Michael Anderson, Barbara Spellman, Timothy Perfect, an anonymous reviewer, and the members of Cogfog. Correspondence concerning this article should be addressed to B. C. Storm, Department of Psychology, University of California, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095 (e-mail: storm@ucla.edu).

Selection and Inhibition in Retrieval-Induced Forgetting

The most fully explored and best-supported explanation of RIF rests on an assumption that $Rp-$ items are actively inhibited during the retrieval of $Rp+$ items (e.g., Anderson, 2003; Anderson et al., 1994; Anderson & Spellman, 1995; E. L. Bjork, R. A. Bjork, & Anderson, 1998; E. L. Bjork, R. A. Bjork, & MacLeod, 2006; Levy & Anderson, 2002; MacLeod, Bjork, & Bjork, 2003). According to this view, the attempt to retrieve a target item also activates other items associated with the same retrieval cue(s), creating competition and requiring that the competing items be selected against. Intrinsic to such selection is the inhibition of competing items, which is then presumed to impair their subsequent recall. Thus, RIF can be thought of as an adaptive process that functions to resolve interference from competing information in order to facilitate the retrieval of target information and, in so doing, reduces access to the information that is selected against.

In contrast, noninhibitory interference-based accounts, such as blocking, argue that retrieval practice may simply strengthen the cue–target association between $Rp+$ exemplars and their respective categories. As a consequence, $Rp+$ exemplars may interfere with or block the participants' ability to recall $Rp-$ exemplars associated with those same categories (see Anderson & Bjork, 1994, for a discussion of such interference-based accounts). Accumulating evidence suggests, however, that RIF is not the passive side effect of interference or blocking; rather, it reflects the active inhibition of competing information.

One finding that particularly supports the inhibitory explanation is cue independence. According to most non-inhibitory accounts, RIF should be cue dependent because it is the strengthening, during practice, of the association between $Rp+$ items and a specific cue—namely, the cue presented at study—that then leads to interference with or blocking of $Rp-$ items in response to that cue on the final recall test. If, instead, another cue is presented, there is no reason to assume that $Rp+$ items will also be more strongly associated with that cue and, thus, able to interfere with or block retrieval of the $Rp-$ items. In contrast, Anderson and Spellman (1995) demonstrated that $Rp-$ items are recalled at a diminished rate not only when tested with the same category cue as that presented during study, but also when tested with a novel category cue. After participants learn *cherry* as an exemplar of the category *fruit*, for example, and then receive retrieval practice on *fruit–orange*, subsequent recall of *cherry* is impaired, not only when tested via *fruit–c_____*, but also when tested via *red–c_____*. RIF has now been shown, across a number of studies and laboratories, to be cue independent (see Anderson, 2003, for a summary), providing strong evidence against noninhibitory accounts (but see also Perfect et al., 2004, for a critique of the cue independence procedure).

Substantial evidence supporting the inhibitory account of RIF also exists outside the demonstration of cue independence. RIF has been shown to be recall specific, strength independent, and competition dependent. When, for example, participants read $Rp+$ items, rather than retrieve

them during what would typically be the retrieval practice phase, recall of related $Rp-$ items is not impaired—even though the $Rp+$ items are equally strengthened (e.g., Anderson, Bjork, & Bjork, 2000; Bäuml, 2002). If blocking is the mechanism responsible for impaired $Rp-$ recall, strengthening the $Rp+$ items—regardless of how such strengthening occurs—should lead to RIF. Instead, the extent to which recall of $Rp-$ items is impaired appears largely independent of the extent to which the $Rp+$ items are strengthened.

Furthermore, the extent to which information is forgotten seems to depend on the extent to which it interferes during attempts to retrieve target information. According to the inhibitory account, because weak items do not readily come to mind and, therefore, do not compete during the retrieval practice phase, they do not need to be selected against. Consistent with this idea, it is not the weaker, but the stronger items that are more impaired by RIF (e.g., Anderson et al., 1994; Storm, Bjork, & Bjork, 2005).

Additional evidence for the competition-dependent nature of RIF comes from the research of Anderson et al. (2000), in which the direction of retrieval practice (*fruit–le_____* vs. *fr_____lemon*) was manipulated during the retrieval practice phase. It was only in the former case, when *lemon* had to be selected from other studied fruits, such as *orange*, that RIF was observed, even though both directions of retrieval practice enhanced the recall of $Rp+$ items to the same degree. Taken together, these findings offer compelling evidence that RIF reflects the active suppression of competition during retrieval practice.

Logic of the Present Study

In the present experiment, we examined further the inhibitory account of RIF by asking whether it occurs in the face of unsuccessful target retrieval. Although it is typically assumed that recall during retrieval practice must be successful in order for RIF to occur, such an assumption is not required by the inhibitory account. According to this account, potentially interfering and competing information is suppressed in order to facilitate the retrieval of target information. Thus, RIF should occur whether retrieval of the target information is eventually successful or not. According to interference-based accounts, however, the impairment of competing information is a consequence of target information's being strengthened by retrieval practice. Thus, if a retrieval attempt is not successful, the target of the retrieval attempt will not be strengthened, and no RIF should occur.

The procedure in the present experiment was similar to that in typical RIF studies, with one key exception. After studying several members of each of several categories, the participants were given retrieval practice for half of the studied categories, but with category and letter fragment cues that either (1) triggered the recall (generation) of exemplars from that category that had not been presented in the prior study phase or (2) made retrieval impossible by presenting a two-letter stem that did not represent the initial letters of any exemplar associated with that category. This manipulation thus dictated whether retrieval practice

would be successful. If RIF reflects blocking or some other noninhibitory effect of interference, it should be seen only in conditions in which retrieval practice is possible. If, instead, RIF reflects the active inhibition of competing responses during the retrieval attempt, it should be seen in either condition.

METHOD

Participants

Sixty-four undergraduate students (20 men and 44 women, mean age = 19.2 years) from the University of California, Los Angeles received course credit for their participation.

Materials

Studied lists. The lists to be studied consisted of the 48 category–exemplar pairs (six members of eight categories) of high taxonomic frequency (e.g., *metals–silver*, *metals–brass*, *fruit–orange*, and *fruit–lemon*) used by Anderson et al. (1994, Experiment 3). For counterbalancing purposes, each category was divided into two subsets (A and B) of three exemplars. During study, the participants were exposed to three exemplars from each of the eight categories in semirandom order, with half studying Set A and half studying Set B.

Retrieval practice. Four of the eight categories received retrieval practice, with these categories counterbalanced across participants. For two practiced categories, the participants were guided to retrieve three exemplars that had not been presented during their original study phase, by presenting a category cue and a two-letter stem for each of those category members (possible retrieval practice). For the other two categories, the participants were shown a category name, as well as a realistic two-letter stem, but the two letters did not begin an actual exemplar associated with that category (impossible retrieval practice). To make the materials and procedures in the possible and impossible retrieval conditions as equivalent as possible, no category–exemplar pairs appeared in both the studied list and the retrieval practice task. The possible retrieval condition differed, therefore, from the typical RIF paradigm, but as Bäuml (2002) has demonstrated, RIF is also obtained when participants generate extra-list exemplars during the retrieval practice phase (see also Blaxton & Neely, 1983).

A total of 12 category-plus-stem pairs (6 possible and 6 impossible to retrieve) were presented for practice three times each in a semirandom order. The particular categories associated with possible or impossible retrieval practice were counterbalanced across participants; no two stems shared the same initial two letters, and no two exemplars in a category shared the same initial letter.

Procedure

The experiment consisted of three phases: study, retrieval practice, and testing. During study, 24 category–member pairs were presented one at a time on the computer for 5 sec each. The presentation order was random, with the constraint that no two consecutive pairs could be from the same category. Immediately following study, the retrieval practice phase began with the category-plus-stem cues appearing on the screen for 5 sec each. The participants were instructed to write down the particular exemplar that completed the two-letter stem in their response packet. They were told that there would be repetitions and that the pairs might or might not come from the list that they had just studied.

After an unrelated 10-min distractor task, the first test was administered. Recall of the same 24 category–exemplar pairs as those that had been presented during study was tested, using a category-plus-one-letter-stem cued recall test (e.g., *fruit–l_____* for *fruit–lemon*). Because none of these pairs had appeared in the retrieval practice phase, there were no Rp+ items on the final test. The category-plus-stem cues were listed on a single testing sheet in a semirandom

order, with the average position controlled across all the experimental conditions. The participants were given the testing sheet and 3 min to complete as many items as possible. The participants returned 1 week later and were given an identical test.

RESULTS AND DISCUSSION

Performance During the Retrieval Practice Phase

The participants were able to generate exemplars 62% ($SD = 27\%$) of the time during possible retrieval practice. Despite our efforts to create two-letter stems that did not correspond to an exemplar of a given category, the participants were able to generate correct exemplars 7% ($SD = 11\%$) of the time during impossible retrieval practice. One category (*drinks*) had a particularly high rate of generation in the impossible condition ($M = 29\%$, $SD = 27\%$), reflecting a variety of cocktail drinks that our pilot participants had failed to produce or know. Removing the drinks category from the analysis, successful generation rates become 62% ($SD = 28\%$) and 3% ($SD = 8\%$) for the possible and impossible conditions, respectively. It should be noted that the participants generated something 31% ($SD = 34\%$) of the time in the impossible retrieval practice condition. The vast majority of these words began with the cued letter stem but did not qualify as a member of the cued category.

Cued Recall Performance

The mean correct recall proportions for exemplars from all the categories (including *drinks*) as a function of their being Rp– or Nrp items and from categories receiving possible or impossible retrieval practice are presented in Table 1 for the 10-min and 1-week delays. When these data were analyzed in a 2 (Rp– vs. Nrp) \times 2 (possible vs. impossible) \times 2 (10 min vs. 1 week) repeated measures ANOVA, a significant RIF effect was found, with Rp– items being recalled significantly less often ($M = .63$, $SE = .02$) than their Nrp counterparts ($M = .71$, $SE = .02$) [$F(1,63) = 11.84$, $p < .001$]. Importantly for the present issue, this effect did not interact with whether retrieval practice was possible or impossible. As compared with the recall of studied exemplars from control categories that did not receive retrieval practice, studied exemplars from categories that were given impossible retrieval practice (M for Rp– = .62, $SE = .03$; M for Nrp = .71, $SE = .03$)

Table 1
Mean Category-Plus-Stem Cued Recall Rates for Rp– Versus Nrp Items From Categories Receiving Possible Versus Impossible Retrieval Practice at the 10-min and 1-Week Delays (With Standard Errors)

Delay	Retrieval Practice	Item Type			
		Rp–		Nrp	
		<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
10 min	Possible	.65	.03	.74	.03
	Impossible	.65	.03	.76	.03
1 week	Possible	.59	.03	.67	.03
	Impossible	.62	.03	.67	.03

were forgotten at essentially the same rate as were studied exemplars from categories that received possible retrieval practice (M for $Rp^- = .63$, $SE = .03$; M for $Nrp = .72$, $SE = .03$) [$F(1,63) < 1$].

Although the participants recalled significantly fewer words after a week's delay ($M = .64$, $SE = .02$) than at a 10-min delay ($M = .70$, $SE = .02$) [$F(1,63) = 13.76$, $p < .001$], RIF was observed on both tests and did not interact with delay [$F(1,63) = 2.20$, $p = .14$]. For Rp^- and Nrp items, mean correct recall proportions were .65 ($SE = .03$) versus .75 ($SE = .02$) at the 10-min delay and .61 ($SE = .02$) versus .67 ($SE = .03$) at the 1-week delay, respectively.

Given the high (and unexpected) generation rates of drinks during impossible retrieval practice, we repeated the analysis excluding all data (possible or impossible) from the *drinks* category. The pattern of results remained essentially identical—with, in fact, a trend toward the RIF effect in the impossible retrieval practice condition actually becoming larger.

Generation Rates and RIF

We also examined the relationship between successful generation rates during retrieval practice and the amount of RIF. Within each counterbalancing condition, the participants were ranked according to their performance during possible retrieval practice and then were assigned to either a high or a low generation group via a median split. The resulting groups of high and low generators produced correct exemplars 80% of the time ($SD = 14\%$) and 43% of the time, respectively. Whereas high generators demonstrated a relatively small RIF effect of -4% ($SD = 28\%$), low generators demonstrated a large RIF effect of -13% ($SD = .27$). Although not a difference that achieved statistical significance [$t(62) = 1.24$, $p = .22$], it is nonetheless interesting to consider why RIF might be greater when retrieval practice has failed than when it has succeeded. One problem in interpreting this trend, however, is that the participants who performed worse in the retrieval practice phase for a particular category might simply have been less familiar with that category, and the lower Rp^- rates when retrieval practice failed could, thus, reflect this category-participant interaction, rather than an increase in RIF.

The same analysis was conducted for retrieval practice performance in the impossible condition. For this analysis, however, we computed performance scores on the basis of the extent to which the participants generated anything during impossible retrieval practice, whether or not their response fit the category or the letter stem cues. By performing median splits on each counterbalancing condition, we created high- and low-generation groups, with high generators producing something in response to the impossible cues 53% of the time ($SD = .34$) and low generators producing something 9% of the time ($SD = 14\%$). Interestingly, although not a difference achieving statistical significance, the low generators ($M = -.16$, $SD = .26$) demonstrated a larger RIF effect than did the high generators ($M = -.07$, $SD = .28$) [$t(62) = 1.4$, $p = .17$]. Possibly, the participants who failed to generate a response spent more time actively engaged in the retrieval

search process, thus exacerbating the need to suppress competing items. Given the lack of statistical significance or an a priori prediction, however, this interesting trend should be interpreted with caution.

GENERAL DISCUSSION AND CONCLUSIONS

The present findings provide an important insight into why remembering causes forgetting. Specifically, they suggest that it is the retrieval attempt, and not the retrieval per se, that induces the forgetting of competing information. Even when the participants were provided with retrieval practice cues that precluded the successful retrieval of a category member, studied exemplars associated with that category became less recallable than they would have been otherwise. This finding is consistent with an inhibitory account because RIF—if it is the product of a process in which interfering information is suppressed in order to facilitate the retrieval of target information—should occur whether the eventual retrieval is successful or not.

In contrast, interference-based noninhibitory accounts of RIF seem to demand that Rp^+ items be strengthened in order for RIF to be observed. It could be argued, however, that the participants did generate something during the impossible retrieval practice phase and that these generations—despite their not conforming to the cues—nonetheless resulted in blocking the future recall of Rp^- items. Although this conjecture cannot be ruled out completely by the present data, the trend suggesting that the participants who recorded fewer responses during impossible retrieval practice demonstrated greater RIF appears to argue against it. An inhibitory account of this finding might be that once participants do generate a response, they cease further retrieval attempts, thus preventing additional RIF.

Another finding of theoretical interest, given MacLeod and Macrae's (2001) finding that RIF effects may dissipate within 24 h, is that the RIF effect remained after an additional week's delay. Because the same material was tested at both delays in the present experiment, however, the observed RIF after a week could be a kind of testing effect: Rp^- items were recalled less often than the Nrp items during the initial test and may, therefore, have benefited less from an additional retrieval event. Nevertheless, it is interesting that the effect of retrieval practice on nonpracticed items remained at a week's delay, even if mediated by differing recall rates on the intermediate test. That is, by affecting what information can and cannot be recalled in the near future, RIF might dictate which information also benefits most from subsequent learning and rehearsal. As such, even if the inhibition caused by RIF is relatively short-lived, the effects of that temporary inhibition could potentially be very long lasting.

Finally, the present findings suggest that researchers, who typically have been very careful to ensure that participants are able to complete the retrieval practice task with a high degree of success, have assumed dynamics that are not in evidence—at least in the present results. In fact, the present findings suggest that any effort to ensure high

rates of retrieval practice success by reducing competition may be more than misguided, because it appears to be the level of competition that arises during retrieval practice that determines whether related information is forgotten, not whether such retrieval practice succeeds or fails.

REFERENCES

- ANDERSON, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory & Language*, **49**, 415-445.
- ANDERSON, M. C., BJORK, E. L., & BJORK, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review*, **7**, 522-530.
- ANDERSON, M. C., & BJORK, R. A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 265-325). San Diego: Academic Press.
- ANDERSON, M. C., BJORK, R. A., & BJORK, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1063-1087.
- 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.
- BÄUML, K.-H. (2002). Semantic generation can cause episodic forgetting. *Psychological Science*, **13**, 356-360.
- BJORK, E. L., & BJORK, R. A. (1988). On the adaptive aspects of retrieval failure in autobiographical memory. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues. Vol. 1: Memory in everyday life* (pp. 283-288). New York: Wiley.
- BJORK, E. L., BJORK, R. A., & ANDERSON, M. C. (1998). Varieties of goal-directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 103-137). Hillsdale, NJ: Erlbaum.
- BJORK, E. L., BJORK, R. A., & MACLEOD, M. D. (2006). Types and consequences of forgetting: Intended and unintended. In L.-G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological perspectives* (pp. 141-165). Hove, U.K.: Psychology Press.
- BJORK, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309-330). Hillsdale, NJ: Erlbaum.
- BLAXTON, T. A., & NEELY, J. H. (1983). Inhibition from semantically related primes: Evidence of a category-specific inhibition. *Memory & Cognition*, **11**, 500-510.
- CIRANNI, M. A., & SHIMAMURA, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1403-1414.
- DUNN, E. W., & SPELLMAN, B. A. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology*, **39**, 420-433.
- LEVY, B. J., & ANDERSON, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, **6**, 299-305.
- MACLEOD, M. D., BJORK, R. A., & BJORK, E. L. (2003). The role of retrieval-induced forgetting in the construction and distortion of memories. In B. Kokinov & W. Hirst (Eds.), *Constructive memory* (NBU Series in Cognitive Science, pp. 55-68). Sofia: New Bulgarian University.
- MACLEOD, M. D., & MACRAE, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science*, **12**, 148-152.
- PERFECT, T. J., STARK, L.-J., TREE, J. J., MOULIN, C. J. A., AHMED, L., & HUTTER, R. (2004). Transfer appropriate forgetting: The cue-dependent nature of retrieval-induced forgetting. *Journal of Memory & Language*, **51**, 399-417.
- SHAW, J. S., III, BJORK, R. A., & HANDAL, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review*, **2**, 249-253.
- STORM, B. C., BJORK, E. L., & BJORK, R. A. (2005). Social metacognitive judgments: The role of retrieval-induced forgetting in person memory and impressions. *Journal of Memory & Language*, **52**, 535-550.

(Manuscript received September 2, 2005;
revision accepted for publication March 28, 2006.)