Semantic Integration as a Boundary Condition on Inhibitory Processes in Episodic Retrieval

Leilani B. Goodmon
Saint Leo University

Michael C. Anderson
MRC Cognition and Brain Sciences Unit, Cambridge, England

Recalling an experience often impairs the later retention of related traces, a phenomenon known as retrieval-induced forgetting (RIF). Research has shown that episodic associations protect competing memories from RIF (Anderson & McCulloch, 1999). We report 4 experiments that examined whether semantic associations also protect against RIF. In all experiments, robust RIF occurred when there were few associations between practiced and nonpracticed sets, but RIF was abolished when there were many. The benefits of semantic integration were independent of episodic integration strategies and were not mediated by intentional use of the associations. Rather, these results establish a new boundary condition on RIF—semantic integration—that has a potent impact on the magnitude of RIF and may explain variability in the RIF phenomenon.

Keywords: inhibition, forgetting, integration, interference, retrieval-induced forgetting

Supplemental materials: http://dx.doi.org/10.1037/a0021963.supp

Everyone has forgotten things that they were sure they knew. For example, the name of a colleague, the location of one's passport, or an event from our past, sometimes simply eludes us. When this occurs, we often abandon our efforts to remember, only to have the information return (to our consternation), on a later occasion, as though there was never a problem recalling it. In recent years, evidence has accumulated that some of these lapses arise from inhibitory processes. According to this idea, retrieval cues activate many items in memory, even when we are searching for a particular experience. The activation of so many traces unfortunately causes retrieval interference, because traces associated to a common cue compete for access to conscious awareness (for reviews, see Anderson & Neely, 1996; Crowder, 1976; Postman, 1971). To focus retrieval on the target, inhibition is thought to suppress competing traces, with the lingering effects causing difficulties in recalling those competitors. This phenomenon, known as retrieval-induced forgetting (RIF), has been observed in many contexts, including episodic memory (e.g., Anderson, Bjork, & Bjork, 1994; Ciranni & Shimamura, 1999), semantic memory (Blaxton & Neely, 1983; Johnson & Anderson, 2004), memory for word meaning (Johnson & Anderson, 2004; Shivde & Anderson, 2001), memory for factual knowledge (Anderson & Bell, 2001; Gómez-Ariza, Lechuga, Pelegrina, & Bajo, 2005; Macrae &

MacLeod, 1999; Phenix & Campbell, 2004), autobiographical memory (Barnier, Hung, & Conway, 2004; Harris, Sharman, Barnier, & Moulds, 2010), eyewitness memory (Garcia-Bajos, Migueles, & Anderson, 2009; MacLeod, 2002; Migueles & Garcia-Bajos, 2006; Shaw, Bjork, & Handal, 1995), visuo-spatial memory (Ciranni & Shimamura, 1999), and even memory for phonological information retrieved during language production (Levy, McVeigh, Marful, & Anderson, 2007). RIF, thus, is a general factor contributing to memory lapses in a broad range of contexts.

Although RIF appears to be a general phenomenon, there are boundary conditions under which it does not occur (see Anderson, 2003, for review). These boundary conditions are theoretically informative. For example, retrieving a memory only impairs the retention of competing traces when those competitors interfere during retrieval (Anderson et al., 1994; Anderson, Bjork, & Bjork, 2000; Bäuml, 1998; Blaxton & Neely, 1983; Shivde & Anderson, 2001), a property known as interference dependence (Anderson, 2003; for reviews, see Anderson & Levy, 2010; Storm, 2010). The dependency of RIF on interference suggests that RIF occurs as a response to interference, consistent with a role of inhibition in this phenomenon. Some investigators have also found that RIF dissipates after 24 hr (Baran, Wilson, & Spencer, 2010; Chan, 2009; MacLeod & Macrae, 2001; Saunders & MacLeod, 2002), although other studies have found RIF to be unaltered in magnitude after a day (Conroy & Salmon, 2005, 2006; Ford, Keating, & Patel, 2004; Migueles & Garcia-Bajos, 2007) or even after a week (Garcia-Bajos et al., 2009; Migueles & Garcia-Bajos, 2007; Storm, Bjork, Bjork, & Nestojko, 2006; Tandoh & Naka, 2007). Most importantly, for present purposes, the amount of RIF depends on associations linking the memory target to its competitors: When the associates of a cue are connected to one another, retrieving some of them no longer impairs the recall of their competitors, a phenomenon known as an *integration effect* (Anderson & McCulloch, 1999). Integration is an important boundary condition on RIF,

perhaps dictating whether nontarget associations of a cue are treated as distinct and competing traces, or instead as other components of a multipart representation (Anderson & McCulloch, 1999; see Radvansky & Zacks, 1991, for a similar view), the parts of which are to be preserved in memory.

Although prior research clearly identifies intentional episodic integration as a moderator of RIF, no work has determined whether preexisting semantic relationships serve a similar function. In the present investigation, we use the term *semantic integration* to refer to the extent to which the associates of a retrieval cue are themselves interlinked via preexisting semantic associations. To examine the influence of semantic integration on RIF, we conducted four experiments that addressed three questions. First, does the magnitude of RIF depend on the existence of a priori associations between the associates of a retrieval cue? Second, does this moderating influence depend on the pattern of a priori links between those associates? Finally, do preexisting semantic associations influence RIF outside of participants' intentions, irrespective of the amount of episodic integration that participants perform? We study these questions with materials tailored to manipulate the degree and pattern of semantic integration. By isolating a distinct role for semantic integration as a boundary condition on RIF, we hope to account for otherwise unexplained variability in the RIF phenomenon more generally.

Integration as a Boundary Condition on RIF

Evidence for integration as a boundary condition on RIF comes from several studies showing that RIF depends on whether participants intentionally interrelate the associates of a retrieval cue. For example, Anderson and McCulloch (1999) examined how encoding instructions affected RIF using the retrieval-practice paradigm of Anderson et al. (1994). Participants studied six exemplars from each of eight taxonomic categories, under either standard or integrative rehearsal study instructions. In the standard encoding condition, participants were asked to study the relation between the category and each exemplar so that they could recall the exemplars later. In the integrative encoding condition, participants were further asked to inter-relate the exemplars of each category. Participants then performed retrieval practice on half of the exemplars from half of the categories, via category-plus-stem cued recall tests (e.g., Fruit Or___ as cues for "Orange"). All of the exemplars were tested after a 20-min delay. Across three experiments, integrative encoding instructions reduced RIF, relative to standard encoding. Even when participants are not instructed to integrate items, but discover this strategy spontaneously, as assessed by postexperimental questionnaires, integration effects occur (Anderson & Bell, 2001; Anderson & McCulloch, 1999). Indeed, when Anderson and McCulloch (1999) increased study time from 5 to 10 s over two repetitions of the study material, self-reports of the use of integration as a strategy increased in the standard encoding condition, and RIF decreased correspondingly. These findings indicate that intentionally inter-relating the associates of a cue often eliminates RIF.

Episodic integration does not always eliminate RIF, however. Integration effects strongly depend on which items within a category are integrated with one another. Anderson, Green, and McCulloch (2000) examined this issue by manipulating the pattern of relationships that participants encoded during the study phase of the retrieval practice paradigm. They found that the *pattern of*

episodic integration matters. RIF was only eliminated when practiced items (Rp+) and their nonpracticed competitors (Rp−) were episodically associated (i.e., target–competitor similarity encoding condition). RIF was not reduced when the associations were formed among the nonpracticed competitors (Rp−; i.e., competitor–competitor similarity encoding condition). Thus, inter-item associations do not necessarily protect against RIF: The pattern of integration matters a great deal to whether RIF will be attenuated. This observation will feature heavily in the current studies on semantic integration.

Semantic Integration: A New Boundary Condition on RIF?

Thus far, all demonstrations of integration effects have required participants to explicitly encode relationships between the associates of a cue. Might similar effects occur when associates have preexisting semantic relationships? Several findings suggest this possibility. In classical studies of retroactive interference, many experiments using the A–B, A–B′ paradigm found greatly reduced retroactive interference when a second list of paired associates shared the same stimulus members as the first list but had a highly similar or associated response term (Kanungo, 1967; Osgood, 1946; Postman & Stark, 1964; Runquist & Marshall, 1963; Shapiro, 1970; Underwood, 1951; Young, 1955; for reviews, see Anderson & Neely, 1996; Crowder, 1976; Postman, 1971). Preexisting semantic relationships also moderate directed forgetting. Sahakyan and Goodmon (2007) found less directed forgetting when there were strong associations between the to-be-forgotten list and the to-be-remembered list. Reduced directed forgetting occurred even though participants were not informed about the relationships between the lists.

Although it has never been studied, retrospective analyses hint that semantic integration also moderates RIF. When we used free-association norms to examine published studies on RIF, we found that variability in this effect was predicted by a priori associative strengths between the practiced and nonpracticed sets (i.e., target–competitor associations). As shown in Table 1, the absence of RIF coincided with more and stronger associative links between the practiced items and their nonpracticed competitors, as assessed by the University of South Florida free association database (Nelson, McEvoy, & Schreiber, 2004). In a particularly striking example, an analysis of the stimuli used by Butler, Williams, Zacks, and Maki (2001)—who failed to find RIF effects with four types of item-specific cues—revealed 39 direct target–competitor associations, 24 of which were moderate-to-strong. There were also approximately 332 indirect associations (i.e., shared associates) between targets and competitors. In contrast, experiments that found RIF using item-specific cues (Anderson et al., 1994; Anderson, Green, & McCulloch, 2000; Anderson & McCulloch, 1999; Bäuml & Hartinger, 2002) had few associations between exemplars. Indeed, as reported in Anderson et al. (1994), semantic associations were intentionally avoided, out of concern that integration might moderate RIF. These contrasting findings indicate that uncontrolled variation in the degree of semantic integration between associates of a cue may significantly moderate the magnitude of RIF and may account for otherwise unexplained variability in the phenomenon. Such variability at times has influenced theoretical perspectives of researchers investigating RIF.

Table 1
Analysis of Associative Connections Between Practiced and Nonpracticed Competitor Sets Across Studies That Obtained Significant RIF (Top) and Studies That Did Not (Bottom) With Item Specific Cues on the Final Test

Study type	Study time (s)		Links	Strength	RIF effect
Studies that did obtain significant RIF					
Anderson et al. (1994)	5		9	.04	−7.60**
Anderson et al. (1994)	5		14	.01	−9.40**
Anderson & McCulloch (1999)	4		14	.04	−7.00**
Anderson, Green, & McCulloch (2000)	3		1	.02	−10.00**
Bäuml & Hartinger (2002)	6		3	.005	−8.28**
		M:	3.20	.02	−8.23
Studies that did not obtain significant RIF					
Bäuml & Hartinger (2002)	6		15	.14	−1.80
Butler et al. (2001)	8		24	.07	3.00
		M:	19.50	.11	0.60

Note.   Associative links refers to the total number of links, both forward and backward, between the practiced and nonpracticed sets across all categories. Associative strength refers to the average strength, both forward and backward, of the total number of associative links between the sets. Numbers in the last column indicate the amount of retrieval-induced forgetting (RIF). Negative numbers reflect amount of forgetting of Rp− items relative to baseline, and positive numbers reflect facilitation of Rp− relative to baseline.
** $p < .05$.

Indeed, Butler et al.'s repeated failure to find RIF with item specific cues led them to conclude that item-specific cuing was a boundary condition on RIF.

Table 1 also summarizes the results of a study by Bäuml and Hartinger (2002), who examined the influence of target–competitor similarity on RIF. Bäuml and Hartinger manipulated the similarity between the practiced and nonpracticed sets by varying whether the nonpracticed exemplars (e.g., *tree* conifer PINE) were drawn from the same subcategory (e.g., *conifers*) as the practiced items (e.g., *tree* conifer CYPRESS) or a different subcategory (e.g., *deciduous*, such as *tree* deciduous BIRCH). They found significant RIF when nonpracticed items were from different subcategories (conifers vs. deciduous), but they found reduced RIF when the nonpracticed items were from the same subcategory (conifers vs. conifers). Thus, like Butler et al. (2001), Bäuml and Hartinger found no RIF when there were target–competitor associations. Because Bäuml and Hartinger found RIF when a priori target–competitor associations were absent, but did not when they were present, this provides especially focused evidence for the possibility that the target–competitor associations present in Butler et al.'s stimulus set may have been responsible for their failure to observe RIF.

Although Bäuml and Hartinger's (2002) findings are compatible with semantic integration, interpretation is complicated because participants were made explicitly aware of target–competitor similarities during encoding: Participants were explicitly presented subcategory labels together with the category name (e.g., *tree* conifer PINE, *tree* deciduous BIRCH). As such, it is difficult to know whether incidental similarity would reduce RIF by itself. Given the presentation of a linking cue, Bäuml and Hartinger may have gotten a similar reduction in RIF if, instead of presenting "*tree* conifer CYPRESS" and "*tree* conifer PINE," they presented "*tree* one CYPRESS" and "*tree* one BIRCH." In the latter example, CYPRESS and BIRCH are from different subcategories but can be integrated episodically by the mediator, "one." If this change also reduced RIF, it would suggest that the presence of a verbal mediator is sufficient to explain reduced RIF (for reviews of

verbal mediation, see Horton & Kjeldergaard, 1961; Jenkins, 1963; Kjeldergaard, 1968; Postman, 1971). Nevertheless, Bäuml and Hartinger's study is compatible with the notion that preexisting target–competitor associations moderate RIF. Taken together, the studies summarized in Table 1 suggest that semantic integration moderates RIF.

The Current Experiments

The four experiments reported here used similar experimental designs to examine the influence of preexisting associations on RIF. All experimental lists, including association strengths among items, can be found in the appendices in the online supplemental materials. We focused in particular on the contrast between effects of target–competitor and competitor–competitor semantic integration. To achieve this, we developed four taxonomic categories, each with eight exemplars. Within each category, there were four pairs of exemplars. The exemplars in these pairs were strongly associated, but they were not directly associated to items from other exemplar pairs. For example, for the Animal category, we chose the pairs HORSE–PONY, LION–TIGER, WALRUS–SEAL, and DUCK–GOOSE, in which associations between exemplars within pairs were strong, but in which associations between exemplars of different pairs were not present (Nelson et al., 2004).

To distinguish the influences of target–competitor and competitor–competitor integration, we simply varied which exemplars within each category received retrieval practice. For example, in the *inter-set* condition, we ensured strong associations between items receiving retrieval practice and their competitors. To achieve this, participants performed retrieval practice on a single item from each of the four exemplar-pairs within the practiced category (e.g., HORSE, LION, WALRUS, and DUCK, in the above example). As illustrated in Figure 1, this ensures high a priori relatedness between the retrieval-practiced targets and their competitors (high target–competitor integration; see the top panel), very much like what was present in Butler et al.'s (2001)

Inter-set Associations Condition

To be Practiced Category – Animals

To be Practiced (Rp+)		Not to be Practiced (Rp-)
HORSE	↔	PONY
LION	↔	TIGER
WALRUS	↔	SEAL
DUCK	↔	GOOSE

Baseline Category – Professions

Not to be Practiced (Nrp)		Not to be Practiced (Nrp)
JUDGE	↔	LAWYER
NURSE	↔	DOCTOR
CHEF	↔	BAKER
SENATOR	↔	PRESIDENT

Intra-set Associations Condition

To be Practiced Category – Animals

To be Practiced (Rp+)			Not to be Practiced (Rp-)		
HORSE	↔	PONY	WALRUS	↔	SEAL
LION	↔	TIGER	DUCK	↔	GOOSE

Baseline Category – Professions

Not to be Practiced (Nrp)			Not to be Practiced (Nrp)		
JUDGE	↔	LAWYER	CHEF	↔	BAKER
NURSE	↔	DOCTOR	SENATOR	↔	PRESIDENT

Figure 1. Top panel: Inter-set condition design. Bottom panel: Intra-set condition design. *Note:* "↔" indicates the presence of a semantic association between the exemplars. In the inter-set condition, one item from each pair (e.g., HORSE, LION) is given practice so that everything that is practiced has an associated exemplar in the nonpracticed set. In the intra-set condition, both items from the pair (e.g., HORSE and PONY, WALRUS and SEAL) are given practice, so that everything that is practiced is not associated to the nonpracticed set. Rp+ = practiced items; Rp− = their nonpracticed competitors; Nrp = no retrieval practice.

stimulus set. We hypothesized that strong preexisting target–competitor associations would reduce RIF, mimicking the null RIF finding of Butler et al. Experiments 1–4 examine this prediction under different encoding conditions, and with different stimuli, designed to better understand the range of conditions under which such effects occur.

Of course, reduced RIF in the inter-set condition would not unambiguously support the importance of target–competitor associations because having any associations among exemplars (even if they were competitor–competitor associations) might reduce the effect. To attribute reduced RIF to target–competitor associations, we must show that RIF occurs as long as associations do not link targets to competitors. Thus, we compared the results from the inter-set condition (see the top panel of Figure 1) with those from an *intra-set* condition (see the bottom panel of Figure 1), in which participants encoded precisely the same exemplars as were encoded in the inter-set condition. However, during retrieval practice, instead of practicing one item from each of the four exemplar-pairs, participants practiced both items from each of two of the exemplar pairs of a category (e.g., HORSE, PONY, LION, TIGER). This ensured that there were no associations between the set of items receiving retrieval practice and their competitors (see the low target–competitor integration in the bottom panel of Figure 1). Because there were no target–competitor associations, there should be a typical RIF effect, according to Anderson and

Spellman's (1995) model (a point to which we return in the General Discussion section). Thus, even with the items held constant, and the associations within a category being identical, different patterns of RIF should emerge.

To ensure the relevance of the current findings to long-lasting RIF, we inserted a 20-min retention interval between retrieval practice and the final recall test, and we also controlled the influence of output interference. In support of the latter aim, we measured RIF using item-specific cues, cuing participants with the category name plus the first letter of each exemplar, one item at a time. To ensure that output interference from strong practiced items could not be a factor explaining RIF, all of the unpracticed members of practiced categories (Rp− items) were tested in the first four positions of the final block testing items from each category, whereas practiced items (Rp+ items) were tested in the last four positions. Items within the baseline (hereinafter, Nrp or "no retrieval practice") categories were correspondingly divided into those tested in the first four positions (Nrp1st) within a category block and those tested in the last four positions (Nrp2nd), to serve as baselines, matched for within-category serial position, against which to compare Rp− and Rp+, respectively. Because we tested the Rp− items first, any forgetting would have to be due to processes that occurred during retrieval practice.

Controlling output order also allows us to examine, within the same experiment, the effects of semantic integration on output interference and longer term RIF. Output interference is considered to be a manifestation of RIF, whereby retrieving items in the initial portion of the testing sequence causes forgetting of items cued later on (Anderson et al., 1994; Bäuml, 1998). Essentially, the effects of retrieval during the test should be comparable with those induced by retrieval practice. Thus, the more items that are tested, the greater the forgetting of later tested items should be. If so, we should find that our manipulations of preexperimental associations have a similar effect on output interference as they have on RIF. Specifically, when comparing the recall of items tested in the first and second halves of a category block, there should less output interference in the inter-set condition, in which associations between the halves are strong, than in the intra-set condition, in which there are few associations between sets. Moreover, we should be able to observe this in the baseline condition, the recall of which is uncontaminated by the effects of the earlier retrieval practice phase.

Experiment 1: The Effect of Semantic Integration on RIF

In Experiment 1, we manipulated the retrieval-practice pattern between subjects, creating separate inter-set and intra-set integration groups. If RIF occurs, recall should be lower for Rp− items than Nrp1st items; if output interference occurs, Nrp2nd items should be more poorly recalled than Nrp1st items. If semantic target–competitor integration reduces RIF, as shown to be the case for episodic integration, significantly less RIF should be observed in the inter-set condition than in the intra-set condition. A corresponding effect may be observed for output interference to the extent that this effect is also produced by the mechanisms underlying RIF.

Method

Participants.   Sixty-four undergraduate psychology students from the University of South Florida participated in exchange for extra credit. Participants ranged from 18 to 35 years of age. Thirty-two participants participated in each condition.

Design.   Integration pattern was manipulated on two levels: inter-set and intra-set integration. In the inter-set condition, participants performed retrieval practice on one exemplar from each of the four associative couples within the practiced category, whereas intra-set participants practiced both exemplars from each of two couples within a category. Retrieval practice status of an exemplar was manipulated within subjects on four levels (Rp+, Rp−, Nrp1st, and Nrp2nd). Exemplars either (a) were given retrieval practice (Rp+), (b) were unpracticed but from practiced categories (Rp−), or (c) were unpracticed members of unpracticed categories (Nrp). Nrp items were further divided into those tested in the first four serial positions of the category (Nrp1st) or in the last four (Nrp2nd), to provide distinct baselines for Rp− and Rp+ items, which were tested in the first and last four positions of their categories, respectively. The percentage of items recalled on a category-plus-stem-cued recall test (e.g., A*nimal* H_____) was measured.

Materials.

Category construction.   Appendix A in the online supplemental materials shows that eight exemplars from each of four unrelated categories (i.e., *Animal*, *Profession*, *Appliance*, *Weapon*) were selected from taxonomic frequency norms (Battig & Montague, 1969; McEvoy & Nelson, 1982; Van Overschelde, Rawson, & Dunlosky, 2004; Yoon et al., 2004). There were no direct associations between members of different categories, according to the University of South Florida free association norms. The word frequency of the category labels was moderate ($M = 30.75$ occurrences per million; Kucera & Francis, 1967). Two additional filler categories (*City*, *River*) were constructed with eight exemplars each (e.g., RENO, MEMPHIS, CONGO, RHINE).

Exemplars were moderate in taxonomic frequency (average position = 20.54 in Battig & Montague's, 1969, norms)[1] and were moderate in word frequency ($M = 42$ occurrences per million; Kucera & Francis, 1967). Because the final test used category-plus-letter stem cues, exemplars were selected so that within each category, exemplars had a unique first letter. Stem difficulty was controlled by selecting items with moderate-high versatility (i.e., number of words in Kucera & Francis, 1967, that can complete the word's two letter stem [$M = 249.78$]; Solso & Juel, 1980).

To examine the effects of semantic integration, four "couples" of exemplars were selected from each category so that couple members were related to each other (e.g., HORSE–PONY) but were not directly related to members from other couples in that category (LION, TIGER, WALRUS, SEAL, etc.). Appendix B in the online supplemental materials shows 11 of the 16 couples were bidirectionally related, with an average strength of 0.24 ($SD = 0.22$) in both directions. Direct strength for the five unidirectionally related couples averaged 0.12 ($SD = 0.19$). The couples also had, on average, 2.69 ($SD = 1.74$) shared associates and an average shared associate strength of 0.03 ($SD = 0.05$).

Two sets of exemplars for each category were created by dividing the categories into two sets of four exemplars each (Set A and Set B). For the inter-set condition, one member from each of the four couples within a category appeared in each set. This ensured strong between-set associations and weak within-set associations. In the intra-set condition, two whole couples appeared in each set, ensuring that association strengths were strong within sets but were weak between the sets. Within each of the intra-set and inter-set conditions, the A and B sets were equated on several variables that affect cued recall including concreteness, set size, connectivity, probability of resonance, and frequency.

Counterbalancing.   Two categories served in the two baseline conditions, and two categories served in the retrieval practiced conditions. Counterbalancing ensured that every exemplar from every category served equally often in the intra-set and inter-set conditions and in each of our retrieval-practice status conditions (baseline conditions vs. retrieval practiced conditions). In addition, to ensure that practiced and baseline categories were equated on serial position in the testing sequence, two testing orders were developed. In Order 1, a practiced category was tested first followed by a baseline category, a practiced category, and finally a baseline category. In Order 2, a baseline category was tested first, followed by a practiced category, a baseline category, and finally a practiced category. As such, there were eight levels of counterbalancing, when all measures are considered (4 retrieval practice × 2 test levels).

Filler tasks.   Participants completed the Shipley Vocabulary Test (Zachary, 1991), the Morningness–Eveningness Questionnaire (Horne & Ostberg, 1976), and the Cognitive Failures Questionnaire (Broadbent, Cooper, FitzGerald, & Parkes, 1982) during the retention interval. In addition, participants performed a series of reasoning problems.

Procedure.   There were four phases: a study phase, a retrieval practice phase, a retention interval phase, and a test phase. In the study phase, participants saw category–exemplars pairs in the center of the computer screen at a rate of 5 s each, and they were told to study each pair by relating the exemplar to its category name. The order for the pairs was block randomized, so that each block contained one exemplar from each category. The resulting eight blocks of six items (four items from the experimental categories and two items from the filler categories) ensured that exemplars from the categories were evenly distributed throughout the list. Within each block, the order was random except that (a) in the first block, filler items were presented first to minimize primacy effects; (b) in the last block, two fillers were presented last to minimize recency effects; (c) no two categories appeared in sequence more than once; and (d) exemplars within the same category were spaced so that there was an average of seven other items in between them. Also, the average distance between the Rp+ and the Rp− items for a given category was kept constant across the inter-set and intra-set conditions, and Rp+ and Rp− items within each category were distributed throughout the study list.

---

[1] The average taxonomic frequency of our exemplars ($M = 20.5$) fell midway between that used in Anderson et al.'s (1994) strong-exemplar ($M = 8$) and weak-exemplar ($M = 33$) conditions. This lower frequency reflects the need to constrain exemplar selection to those items with particular associative relationships. Although this lower frequency puts us at a disadvantage in finding RIF (which is greater for higher taxonomic frequency items), many of the current exemplars are in fact high frequency.

In the retrieval practice phase, participants were told that their memory would be tested. They were told that on each test trial, they would see a category name and the first two letters of one member from that category on the screen, and they were asked to recall the studied category member that fit the letters and say the category and member out loud within 10 s. They were told that some of the words would be tested more than once and that they should recall them as they would any other. The presentation order of the category-plus-two-letter stem cues was pseudorandomized with the following constraints: (a) The first three and last three cues were filler items to acquaint participants with the task and to control for primacy and recency effects; (b) items were tested three times on an expanding schedule with an average of 3.5 trials between the first and second presentations and an average of 6.5 trials between the second and third presentations; (c) no two category members were retrieval practiced in succession; and (d) the use of filler items helped ensure that no two pairs appeared consecutively more than once.

A 20-min retention interval—during which participants completed the Shipley Vocabulary Test, the Morningness–Eveningness Questionnaire, the Cognitive Failures Questionnaire, and reasoning problems—was used to show that RIF is not short-lived.

In the final test phase, participants were given a surprise category-plus-letter stem cued recall test (e.g., *Animal* H___) for all studied exemplars. They were told that on the computer screen, they would see a category name and the first letter of one member from that category, and that they would have 10 s to recall the exemplar from the study phase that fit the letter. The presentation order of the category-plus-letter stem cues was pseudorandomized in blocks as follows: (a) All the members of a category were tested in a block; (b) the two filler categories from the study and retrieval practice phases were tested first; (c) to control for output interference from stronger practiced items (Rp+ items) in practiced categories, Rp− items were tested first within each category block, followed by the Rp+ items; and (d) practiced category blocks were alternated with baseline categories so that their serial positions were similar.

Results

A series of mixed-subjects analyses of variance (ANOVAs) was conducted to examine recall performance in the retrieval practice and final test phases. In the analysis of the final test data, integration pattern (inter-set, intra-set) and counterbalancing were analyzed as between-subjects factors, and retrieval practice status (Rp+, Rp−, Nrp1st, Nrp2nd) was analyzed as a within-subjects factor. In this and in the remaining studies, analyses examining hypothesized effects (e.g., RIF) within one level of a between-subjects manipulation were done in a follow-up ANOVA conducted on the relevant group, using a planned comparison. Unless otherwise noted, none of our counterbalancing measures interacted with RIF effects.

Retrieval practice performance. The retrieval practice success rate did not differ for the inter-set ($M = 0.87$, $SD = 0.10$) and the intra-set ($M = 0.87$, $SD = 0.13$) conditions ($F < 1$).

RIF. Collapsing over the pattern of integration factor revealed significant RIF, replicating prior work: The probability of recalling Rp− items was reliably lower than the probability of

recalling baseline items (-7%), $F(1, 48) = 5.46$, partial $\eta^2 = .102$, $p = .024$.

More important, however, was how RIF varied with the pattern of integration. As predicted, there was significant RIF (-13%) in the intra-set condition, $F(1, 24) = 12.64$, partial $\eta^2 = .345$, $p = .002$, but this effect was reduced (0%) in the inter-set condition ($F < 1$; see the top panel of Figure 2), a difference confirmed by an interaction between RIF (baseline vs. Rp−) and the integration pattern, $F(1, 48) = 5.36$, partial $\eta^2 = .10$, $p = .025$. Thus, when associations between the practiced (Rp+) and nonpracticed (Rp−) sets were maximized, as in the inter-set condition, there was no difference in recall between Rp− and baseline items. These results strongly support the assumption that associations between practiced items and competitors moderate RIF.
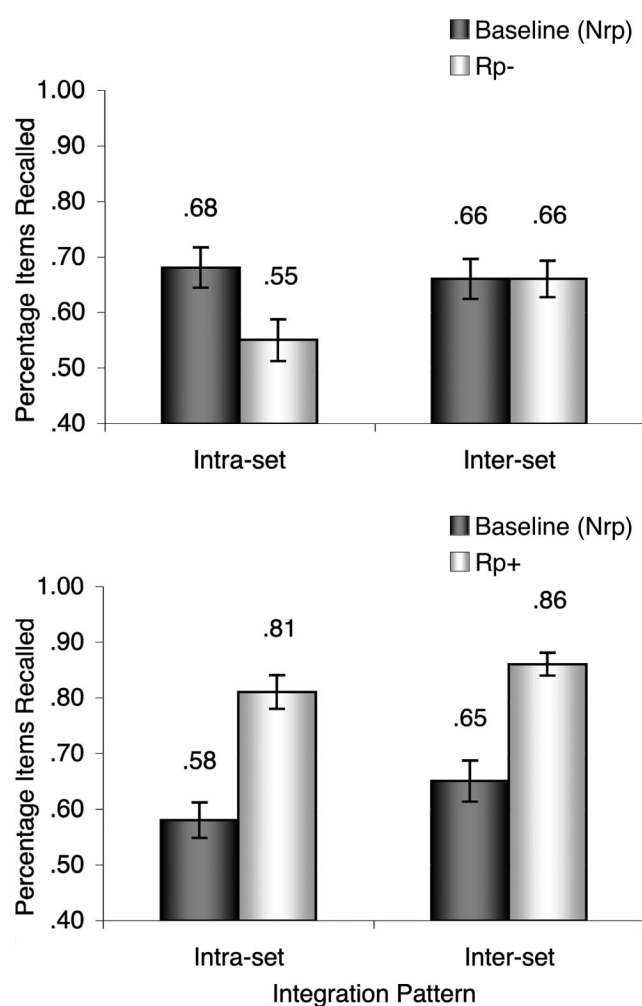

Figure 2. Results of Experiment 1: Significant retrieval-induced forgetting of Rp− items only in the intra-set condition but significant strengthening of Rp+ items in both the intra-set and inter-set conditions. Top panel: Retrieval-induced forgetting effect; mean percentage of baseline and Rp− items recalled as a function of retrieval practice pattern (intra-set, inter-set). Bottom panel: Retrieval practice benefit effect; mean percentage of baseline and Rp+ items recalled as a function of retrieval practice pattern. Nrp = no retrieval practice; Rp+ = practiced items; Rp− = their nonpracticed competitors. Error bars indicate standard error (*SE*).

Benefits of retrieval practice on practiced items. Performing retrieval practice improved the recall of practiced items on the final test, $F(1, 48) = 84.30$, partial $\eta^2 = .637$, $p < .001$, and this benefit did not vary with integration pattern ($F < 1$; see the bottom panel of Figure 2). Indeed, the amount of RIF was independent of the strengthening of practiced items: There was facilitation in the intra-set condition ($+23\%$) where forgetting was present, $F(1, 24) = 48.36$, partial $\eta^2 = .668$, $p < .001$, and facilitation in the inter-set condition ($+21\%$) where forgetting was absent, $F(1, 24) = 36.92$, partial $\eta^2 = .606$, $p < .001$.

Output interference. Output interference within the baseline condition was calculated by comparing recall of items from the last half of the category block (i.e., Nrp2) with recall from the first half of the category block (i.e., Nrp1). Overall there was output interference within the baseline condition, $F(1, 48) = 4.37$, partial $\eta^2 = .083$, $p = .042$, with fewer items recalled in the last half of the category block (see the Nrp conditions in the bottom panel of Figure 2; $M = 0.62$, $SD = 0.20$) than in the first half (see the Nrp conditions in the top panel of Figure 2; $M = 0.67$, $SD = 0.20$). Interestingly, integration moderated output interference just as it moderates RIF. Although the interaction between output interference and integration pattern was not significant, $F(1, 48) = 2.24$, partial $\eta^2 = .045$, $p = .141$, in the inter-set condition, output interference was not significant (-1% effect; $F < 1$). By contrast, in the intra-set condition, output interference (-10%) was significant, $F(1, 24) = 5.15$, partial $\eta^2 = .177$, $p = .033$. These findings support the idea that output interference is a form of RIF and is moderated by similar factors (see, e.g., Bäuml, 1998; Bäuml & Hartinger, 2002).

Discussion

The current findings highlight the importance of controlling semantic integration in studies of RIF. Strong preexisting semantic associations between category exemplars clearly moderated RIF in Experiment 1. Crucially, it was not the mere presence of associations between the items but rather the specific pattern of those associations that mattered: With strong associations *between* the practiced and nonpracticed sets (target–competitor associations), forgetting was reduced, but with associations *within* the practiced and the nonpracticed sets (competitor–competitor associations), robust forgetting was observed. Similar effects occurred for output interference, providing converging evidence. It is especially noteworthy that these entirely different findings were observed with exactly the same exemplars simply by arranging items into different sets according to their patterns of associations.

It remains unclear, however, whether such preexisting associations moderate RIF directly or indirectly by encouraging the explicit formation of episodic associations at the time of study. We address the latter possibility next.

Experiment 2: Are Semantic Integration Benefits Mediated by Explicit Integration Strategies?

Although Experiment 1 suggests that preexisting associations moderate RIF, this effect may depend on intentional integration. Strong a priori associations may have led participants to seek inter-connections between items to facilitate later memory. If semantic integration benefits are entirely dependent on intentional integration, these findings may be another example of episodic integration (Anderson & McCulloch, 1999) and would not present a distinct boundary condition. Moreover, it would suggest that inter-set semantic associations are not sufficient to moderate RIF in the absence of an intentional effort to link items together.

To address this possibility, we sought to minimize episodic integration through experimental manipulations and through analyses based on postexperimental reports. We manipulated whether participants received intentional or incidental encoding instructions. In the intentional study group, participants received the study instructions of Experiment 1. In the incidental encoding group, participants were not told to study the items. Rather, they were asked to judge how well each exemplar fit as a member of its category. We hoped this would focus participants on the relation between each exemplar and its category name and would discourage intentional integration. If this procedure is successful, we should observe lower integration rates on a postexperimental integration rating scale than in the intentional learning condition.

If semantic integration truly modulates RIF, independently from episodic integration, we should observe significantly reduced RIF in the inter-set condition compared with the intra-set condition for our incidental encoding participants. This reduction should occur despite a significant reduction in (or perhaps absence of) intentional episodic integration.

Method

Participants and design. One hundred and forty-seven undergraduate psychology students from the University of South Florida participated for extra credit. Participants ranged from 18 to 52 years of age. The design was identical to Experiment 1 but with the addition of a between-subjects manipulation of encoding method (intentional vs. incidental). Because the incidental encoding instructions were intended to eliminate intentional encoding, participants who reported that they expected a later memory test ($n = 19$) were excluded from the analyses; these participants were replaced so that there were 32 participants in each of the four between-subjects conditions: (a) intentional encoding (intra-set), (b) intentional encoding (inter-set), (c) incidental encoding (intra-set), and (d) incidental encoding (inter-set). This replacement also ensured that the experiment was completely counterbalanced ($n = 128$), with 16 participants in each of the eight counterbalancing conditions. Excluding these participants does not alter the conclusions drawn from this study.

Procedure and materials. The procedures of Experiment 1 were used, except that we manipulated study instructions. Half of the participants received intentional encoding instructions (as in Experiment 1), the other half received incidental encoding instructions. Participants in the incidental condition were told that the experiment concerned judgment and reasoning. They were told that category–exemplar pairs would appear in the center of the screen and that they would have 5 s to judge how well the exemplar "fit" the category on a scale ranging from 1 (*no fit*) to 5 (*best fit*). All other procedures for the encoding phase, retrieval practice phase, retention interval, and test phase were identical to those used in Experiment 1.

A postexperimental questionnaire was included that measured whether participants episodically integrated the items during the study phase. For each category, participants indicated on a 5-point

scale ranging from 1 (*none of the time*) to 5 (*all of the time*) how often during the study phase they intentionally thought back to previous exemplars when they encountered a new exemplar, and rehearsed them together. In addition, a second question measured whether participants "incidentally linked" the exemplars while making their judgments. For each category, participants indicated on a 5-point scale ranging from 1 (*none of the time*) to 5 (*all of the time*) how often during the judgment phase they intentionally thought back to previous category members when they encountered a new exemplar, to judge the current item in relation to previous ones. This question addresses the concern that in the course of judging the relatedness of an exemplar to its category, participants might make comparative judgments, introducing an unintentional form of episodic integration. Another question was added to determine whether participants expected their memory of the judged items to be tested. This enabled us to identify and eliminate participants in the incidental encoding condition who might have used intentional learning strategies that we sought to eliminate with incidental encoding.

Results

A mixed-subjects ANOVA was conducted, with integration pattern (inter-set, intra-set), encoding method (intentional, incidental), and counterbalancing as between-subjects factors, and with retrieval practice status (Rp+, Rp−, Nrp1st, Nrp2nd) as a within-subjects factor.

Retrieval practice success. Retrieval practice success rate did not vary between the inter-set ($M = 0.93$, $SD = 0.06$) and the intra-set ($M = 0.94$, $SD = 0.06$) groups in either encoding condition ($F < 1$ in all cases).

Overall RIF. Collapsing across integration pattern and type of learning, reliable RIF was found: Fewer Rp− than baseline items were recalled (−5% RIF effect), $F(1, 96) = 9.05$, partial $\eta^2 = .086$, $p = .003$. The amount of RIF did not vary with encoding method ($F < 1$). RIF was observed in the incidental group (−6%), $F(1, 48) = 5.75$, partial $\eta^2 = .107$, $p = .020$, and RIF approached significance in the intentional group (−5%), $F(1, 48) = 3.63$, partial $\eta^2 = .070$, $p = .063$.

RIF as a function of integration pattern. Collapsing across encoding groups, there was a highly significant interaction of RIF with integration pattern, $F(1, 96) = 16.90$, partial $\eta^2 = .150$, $p < .001$. This interaction was produced by a significant RIF effect in the intra-set condition, $F(1, 56) = 18.74$, partial $\eta^2 = .251$, $p < .001$, but no such effect in the inter-set condition ($F < 1$). Thus, as

in Experiment 1, when a priori target–competitor associations were absent, substantial RIF (−13%) occurred, but when those associations were present, RIF was eliminated (+2%).

Of critical concern, however, was whether semantic integration benefits would extend to incidental encoding. Crucially, the three-way interaction between RIF, integration pattern, and encoding method was not reliable, $F(1, 96) = 1.50$, partial $\eta^2 = .015$, $p = .223$. Indeed, the effect of integration pattern on RIF was present for both the incidental group, $F(1, 48) = 16.75$, partial $\eta^2 = .259$, $p < .001$, and the intentional group, $F(1, 48) = 3.63$, partial $\eta^2 = .070$, $p = .063$. RIF occurred for incidental participants (−15%) in the intra-set condition, $F(1, 24) = 14.89$, partial $\eta^2 = .383$, $p = .001$, but not in the inter-set condition (+4%), $F(1, 24) = 2.45$, partial $\eta^2 = .093$, $p = .130$ (see Table 2). Similarly, intentional learning participants exhibited RIF (−10%) in the intra-set condition, $F(1, 24) = 5.75$, partial $\eta^2 = .193$, $p = .025$, but not (0%) in the inter-set condition ($F < 1$).

Benefits of retrieval practice as a function of integration pattern and encoding method. The benefit (Rp+ vs. baseline) for practiced items (Rp+) collapsed across integration pattern and encoding method was significant, $F(1, 96) = 97.03$, partial $\eta^2 = .503$, $p < .001$. None of the two- or three-way interactions of this effect with encoding method or integration pattern were reliable. As shown in Table 2, benefits were attained in all conditions. In the intentional learning condition, there was facilitation in the intra-set condition (+21%) where forgetting was present, $F(1, 24) = 24.30$, partial $\eta^2 = .503$, $p < .001$, and in the inter-set condition (+17%) where forgetting was absent, $F(1, 24) = 22.66$, partial $\eta^2 = .486$, $p < .001$. In the incidental encoding condition, there was significant facilitation in the intra-set condition (+13%) where forgetting was present, $F(1, 24) = 12.00$, partial $\eta^2 = .333$, $p = .002$, and in the inter-set condition (+23%) where forgetting was absent, $F(1, 24) = 48.21$, partial $\eta^2 = .668$, $p < .001$.

Output interference. Output interference within the baseline condition was calculated by comparing recall of items from the last half of the category block (i.e., Nrp2) with recall from the first half of the category block (i.e., Nrp1). Within the baseline condition, the overall output interference effect was significant, $F(1, 96) = 5.57$, partial $\eta^2 = .055$, $p = .020$, with baseline items cued in the last half of the category block (i.e., Nrp2) being recalled more poorly ($M = 0.62$, $SD = 0.21$) than those cued in the first half (i.e., Nrp1; $M = 0.67$, $SD = 0.17$). As in Experiment 1, the interaction between output interference and integration pattern was suggestive, $F(1, 96) = 2.36$, partial $\eta^2 = .024$, $p = .128$. However, the

Table 2

Retrieval Practice Costs (RIF) and Benefits as a Function of Encoding Method and Integration Pattern in Experiment 2

Encoding method	Integration pattern	Retrieval practice costs			Retrieval practice benefits		
		Baseline	Rp−	RIF effect	Baseline	Rp+	Benefit effect
Intentional	Intra-set	.70	.60	−.10**	.64	.85	+.21**
	Inter-set	.64	.64	.00	.63	.80	+.17**
Incidental	Intra-set	.70	.55	−.15**	.61	.74	+.13**
	Inter-set	.64	.68	+.04	.62	.85	+.23**

Note. RIF = retrieval-induced forgetting; Rp+ = practiced items; Rp− = their nonpracticed competitors.
** $p < .05$.

pattern of output interference was consistent with a semantic integration effect. In the intra-set condition, there was a reliable –8% decrement in recall for items in the last half ($M = 0.62$) compared with items in the first half ($M = 0.70$) of the baseline, $F(1, 56) = 7.70$, partial $\eta^2 = .121$, $p = .008$. In the inter-set condition, however, there was no reliable difference in recall between first-half ($M = 0.62$) items and last-half ($M = 0.64$) items ($F < 1$). The three-way interaction between output interference, integration pattern, and encoding method was not significant ($F < 1$). In the intra-set condition, incidental encoding participants exhibited a –9% decrement between the first ($M = 0.70$) and last ($M = 0.61$) half of the baseline, $F(1, 24) = 7.26$, partial $\eta^2 = .232$, $p = .013$, whereas intentional learning participants exhibited more modest (−6%) and less reliable effects, $F(1, 24) = 1.86$, partial $\eta^2 = .072$, $p = .185$. Neither intentional nor incidental encoding participants showed output interference in the inter-set condition ($F < 1$ in each case), as might be expected if semantic integration reduced output interference.

Do semantic integration effects rely on episodic integration? As expected, incidental encoding successfully reduced postexperimental questionnaire ratings of intentional episodic integration ($M = 1.53$, $SD = 1.01$), relative to that observed in the intentional learning condition ($M = 3.60$, $SD = 1.07$), $F(1, 96) = 130.78$, partial $\eta^2 = .577$, $p < .001$. Thus, incidental participants did not intentionally integrate frequently, suggesting that the benefits of semantic integration they experienced were not due to their using the preexisting semantic associations to intentionally form new episodic connections between the items. Consistent with this, even those participants with the lowest integration ratings that were possible ($M = 1.00$, $SD = 0.0$, in which 1 = *never*) still failed to exhibit RIF effects in the inter-set condition (+4% facilitation; $F < 1$). This strongly suggests that semantic integration benefits are not dependent on intentional episodic integration strategies.

However, unintentional episodic integration may have produced the semantic integration benefits. Incidental encoding participants could have integrated exemplars during encoding by thinking back to previously judged items to make their category–exemplar relatedness judgments. However, a median split by comparative rating scores revealed no interaction between overall inhibition and the amount of comparative judgment ($F < 1$). Both high ($M = 3.43$, $SD = 0.69$) and low ($M = 1.92$, $SD = 0.64$) comparative integrators failed to exhibit RIF in the inter-set condition (+2% and +5%, respectively; $F < 1$ in both cases). Similarly, there was no interaction between inhibition and the degree of comparative judgment in the intra-set condition ($F < 1$), with both high ($M = 3.86$, $SD = 0.82$), $F(1, 8) = 11.77$, partial $\eta^2 = .595$, $p < .01$, and low ($M = 2.19$, $SD = 0.83$) comparative judgment groups exhibiting RIF (−14% and –16%, respectively), $F(1, 8) = 5.77$, $MSE = .055$, $p = .043$; $F(1, 8) = 19.09$, partial $\eta^2 = .705$, $p = .002$. These results indicate that semantic integration benefits do not depend on this form of unintentional episodic integration.

In a final analysis, we considered both intentional episodic integration and participants' comparative ratings together in a single composite episodic integration score, to ensure that participants who we considered to be low in episodic integration were low on both of these measures simultaneously. We averaged the episodic and comparative judgment scores and performed a median split of participants on that composite measure. Here again, the effects of semantic integration did not depend on episodic integration. The lowest episodic integrators

exhibited a significant interaction between RIF and integration pattern, $F(1, 16) = 9.62$, partial $\eta^2 = .375$, $p = .007$, with low composite integrators ($M = 1.68$, $SD = 0.43$) exhibiting a reliable RIF effect of –13% in the intra-set condition, $F(1, 8) = 12.57$, partial $\eta^2 = .611$, $p = .008$, and low composite integrators in the inter-set condition ($M = 1.58$, $SD = 0.31$) exhibiting no forgetting (+4%), $F(1, 8) = 1.13$, partial $\eta^2 = .124$, $p = .319$. Thus, even when we used the most stringent measure of episodic integration that we have available, the current data clearly indicate that semantic integration effects on RIF are not dependent on explicit integration strategies.

Discussion

Even though Experiment 2 used an incidental encoding task that greatly reduced intentional integration, the semantic integration pattern still strongly modulated RIF: Participants showed little RIF in the inter-set condition but showed robust forgetting in the intra-set condition. Indeed, even when postexperimental questionnaire responses were used to identify participants who reported *never* intentionally integrating items, or even when other possible sources of incidental episodic integration were considered (comparative ratings), semantic integration benefits remained unaltered. A similar pattern was observed for output interference. It seems that this modulation of RIF reflects an influence of preexisting knowledge.

Experiment 3: Increasing Intentional Integration With Study Time

The clear influence of semantic integration on the magnitude of RIF observed in Experiments 1 and 2 indicates that one's prior knowledge about inter-relationships between competitors can dramatically alter the pattern of RIF. This influence takes place even in the absence of any measurable intentional integration, suggesting that the benefit truly reflects the influence of prior knowledge and does not require episodic integration. Thus, the moderating effects of semantic and episodic integration are distinct. The question remains, however, about how these factors combine to influence RIF. Could distinct contributions of episodic and semantic integration be observed in the same study? If so, do they interact?

Experiment 3 addressed these questions by replicating Experiment 1 while manipulating the amount of time participants are given to study. Prior work has shown that a second exposure to a study list increases subject-initiated episodic integration (Anderson & McCulloch, 1999) and correspondingly reduces the amount of RIF. These findings suggest that manipulating study time during intentional encoding might be an effective way to manipulate episodic integration. Experiment 3 therefore varied whether participants received 5 s to study exemplars, as in Experiment 1, or 10 s, as a means of manipulating the degree of episodic integration.

If semantic and episodic integration are distinct influences on the amount of RIF, one should observe significant effects of both semantic integration pattern and study time. It is unclear, however, whether these factors may interact. On one hand, strong preexisting associations between study items might limit the usefulness of episodic integration, particularly in the inter-set condition, in which between-set associations are stronger. If so, perhaps additional study time might only benefit the intra-set condition, in which preexisting semantic associations between sets are weaker. On the other hand, perhaps people benefit independently from both

episodic and semantic integration; thus, even if there is strong inter-set semantic integration, intentionally integrating during the study phase may confer an additional reduction in RIF, or perhaps even a reversal of the effect. If so, one might expect additive effects of semantic and episodic integration. Thus, collapsing over integration pattern, we should see an effect of study time on RIF, with longer study times yielding less RIF. Collapsing over study time, by contrast, should reveal a main effect of integration pattern, as in previous experiments, with no variation in this benefit of semantic integration as a function of study time.

Method

Participants and design. Sixty-four undergraduates from the University of South Florida participated in exchange for extra credit. Participants ranged from 18 to 31 years of age. Study time (5 s, 10 s) was manipulated between subjects ($n = 32$ participants in each), and integration pattern (inter-set, intra-set) and retrieval practice status (Rp+, Rp−, Nrp1, Nrp2) were manipulated within subjects.

Materials and procedure. The materials and procedures were identical to those of Experiment 2, except that integration pattern (inter-set, intra-set) was manipulated within subjects, and there was an added between-subjects study time manipulation. Intentional learning instructions were used, and half of the participants received 5 s to study each pair, and the other half received 10 s.

Results

A mixed-subjects ANOVA was conducted with study time (5 s, 10 s) and counterbalancing (16 counterbalancing conditions, including the new counterbalancing for pattern of integration) as between-subjects factors, and integration pattern (inter-set, intra-set) and retrieval practice status (Rp+, Rp−, Nrp1, Nrp2) as within-subject factors.

Retrieval practice success. Retrieval practice success rate did not vary between the inter-set ($M = 0.89$, $SD = 0.13$) and the intra-set ($M = 0.89$, $SD = 0.13$) groups in either study time condition ($F < 1$), and there was not a difference in retrieval practice success rate between the 5-s ($M = 0.90$, $SD = 0.11$) and the 10-s ($M = 0.89$, $SD = 0.14$) study conditions ($F < 1$).

Overall RIF and RIF as a function of integration pattern. Overall RIF (baseline vs. Rp−) collapsed across integration pattern, and study time approached significance, $F(1, 32) = 3.25$, partial $\eta^2 = .09$, $p = .081$. The –5% RIF effect was similar to effects found in Experiments 1 and 2 (−7% and –5%, respectively). This overall effect, however, averages over two

types of integration that might reduce the effect (semantic and episodic). Replicating Experiments 1 and 2, the interaction between RIF and integration pattern was reliable, $F(1, 32) = 5.14$, partial $\eta^2 = .138$, $p = .030$, generalizing this effect to a within-subjects design. Collapsing across study time conditions, there was RIF (Nrp – Rp− = 75% − 66% = −9%) in the intra-set condition, $F(1, 32) = 7.67$, partial $\eta^2 = .193$, $p = .009$, but no RIF (Nrp – Rp− = 67% − 67% = 0%) in the inter-set condition ($F < 1$). These results support the idea that semantic integration moderates RIF, strengthening the conclusions of Experiments 1 and 2.

The effects of study time on RIF and on semantic integration effects. Of central concern in Experiment 3 was the effect of increasing study time on RIF, and whether this might encourage episodic integration. Consistent with this proposal, self-reported episodic integration was greater in the 10-s study condition ($M = 3.75$, $SD = 1.01$) than in the 5-s study condition ($M = 3.41$, $SD = 1.08$), $F(1, 32) = 4.21$, partial $\eta^2 = .116$, $p = .048$, and these elevated levels of episodic integration coincided with reduced RIF. Crucially, the interaction between RIF (Rp− vs. baseline) and study time (collapsing over integration pattern) was significant, $F(1, 32) = 5.16$, partial $\eta^2 = .139$, $p = .030$, with participants in the 5-s condition exhibiting a 10% decrement in recall of Rp− items ($M = 0.62$) compared with baseline items ($M = 0.72$), $F(1, 16) = 8.67$, partial $\eta^2 = .351$, $p = .010$, and with participants in the 10-s condition exhibiting no impairment of Rp− items ($M = 0.71$) compared with baseline ($M = 0.70$; $F < 1$). Thus, consistent with our hypothesis, when either semantic integration is high (i.e., inter-set) or episodic integration is high (i.e., 10-s study), Rp− items are protected from RIF.

Interestingly, the effects of study time on RIF did not interact with our manipulation of semantic integration pattern ($F < 1$). This indicates that additional study time decreased RIF comparably for both the intra-set and inter-set conditions (see Table 3), and that inter-set integration decreased RIF comparably for both the 5-s and 10-s study conditions. Indeed, although the overall level of RIF was significantly reduced in the 10-s study condition, the pattern of RIF across the intra-set condition (−4% RIF) and inter-set condition (+7% facilitation) was quite similar to that observed within the 5-s study condition for the intra-set condition (−14% RIF) and inter-set condition (−6% RIF). This supports our hypothesis that semantic and episodic integration have distinct and additive influences on RIF.

Benefits of retrieval practice as a function of integration pattern and study time. The overall benefit (baseline vs. Rp+) in recall for practiced items (Rp+) collapsed across integration

Table 3
Retrieval Practice Costs (RIF) and Benefits as a Function of Study Time and Integration Pattern in Experiment 3

Study time (s)	Integration pattern	Retrieval practice costs			Retrieval practice benefits		
		Baseline	Rp−	RIF effect	Baseline	Rp+	Benefit effect
5	Intra-set	.77	.63	−.14**	.61	.88	+.27**
	Inter-set	.68	.62	−.06	.66	.78	+.12*
10	Intra-set	.73	.69	−.04	.70	.88	+.18**
	Inter-set	.66	.73	+.07	.66	.82	+.16**

Note. RIF = retrieval-induced forgetting; Rp+ = practiced items; Rp− = their nonpracticed competitors.
* $p < .10$.   ** $p < .05$.

pattern and study time conditions was significant, $F(1, 32) = 75.58$, partial $\eta^2 = .703$, $p < .001$, with participants recalling 15% more of the practiced items ($M = 0.84$, $SD = 0.21$) than baseline items ($M = 0.65$, $SD = 0.29$). The facilitation of Rp+ items did not vary as a function of integration pattern ($F < 1$) or study time ($F < 1$), and there was not a three-way interaction between facilitation, integration pattern, and study time, $F(1, 32) = 1.58$, partial $\eta^2 = .047$, $p = .218$.

Output interference. Overall output interference (Nrp1 vs. Nrp2 baseline) within the baseline condition was significant, $F(1, 32) = 5.10$, partial $\eta^2 = .137$, $p = .031$, with baseline items cued in the last half of the category block (i.e., Nrp2) being recalled at a lower rate ($M = 0.67$, $SD = 0.29$) than those cued in the first half (i.e., Nrp1; $M = 0.73$, $SD = 0.23$). Although the interaction between output interference and integration pattern was not significant, $F(1, 32) = 1.70$, partial $\eta^2 = .050$, $p = .202$, as in the previous experiments, the pattern of output interference was consistent with the idea that semantic integration modulates this effect. In the intra-set condition, there was a significant (-10%) decrement in recall from the first ($M = 77$) to the last ($M = 0.67$) half, $F(1, 32) = 5.43$, partial $\eta^2 = .145$, $p = .03$, whereas in the inter-set condition, there was no significant decrement from the first ($M = 0.69$) to the last ($M = 0.67$) half ($F < 1$).

Study time also appeared to moderate output interference. Although the interaction between output interference and study time was not significant, $F(1, 32) = 1.36$, partial $\eta^2 = .137$, $p = .251$, the pattern was consistent with the proposal that episodic integration reduces output interference. In the 5-s study condition (collapsing over integration pattern), where episodic integration was lower, the -8% decrement in recall from the first ($M = 0.72$) to the last half of the baseline ($M = 0.64$), approached significance, $F(1, 16) = 4.10$, partial $\eta^2 = .204$, $p = .060$. In the 10-s study condition, however, where episodic integration was higher, there was no significant decrement in recall from the first ($M = 0.70$) to the last half of the baseline ($M = 0.67$), $F(1, 16) = 1.04$, partial $\eta^2 = .061$, $p = .323$. This pattern of results was especially striking within the intra-set condition: Participants given only 5 s to study showed a -16% decrement from the first ($M = 0.77$) to the last ($M = 0.61$) half, $F(1, 16) = 6.25$, partial $\eta^2 = .281$, $p = .02$, whereas participants in the 10-s study condition showed no reliable decrement in recall from the first ($M = 0.73$) to the last ($M = 0.70$) half ($F < 1$). Participants in the inter-set condition showed no such modulation (-2% output interference in both the 5-s and 10-s conditions), perhaps because output interference was already muted because of semantic integration. From these results, it appears that episodic integration may moderate output interference in a manner parallel to that of semantic integration.

Discussion

Experiment 3 crossed manipulations of semantic integration (inter-set, intra-set) and episodic integration (5 s study time, 10 s study time) within a single experiment, allowing us to examine the contributions of semantic and episodic integration to reducing RIF. These contributions were clearly observed. As predicted, increasing study time from 5 to 10 s significantly increased self-reported episodic integration and significantly reduced RIF. Indeed, Experiment 3 suggests that the influence of episodic and semantic integration was largely additive, yielding the potential to turn

significant RIF into retrieval-induced facilitation. The joint influences of these two sources on RIF were considerable: In the case in which neither factor was present (5 s of study time in the intra-set integration condition), -14% RIF was observed, whereas 7% facilitation was found when both factors were present (10 s of study in the inter-set condition), yielding a net change of 21% in the amount of RIF that was found. We also found parallel reductions in RIF in output interference, both as a function of semantic and episodic integration, again confirming the notion that this phenomenon is produced by the same underlying mechanisms as the overall RIF effect. Thus, the moderating effects of integration on RIF can be supported by preexisting knowledge or novel, episodically encoded relationships between competitors, and these influences can have independent effects.

Experiment 4: Semantic Integration Effects Using Butler et al.'s (2001) Stimulus Set

The findings from Experiments 1–3 demonstrate how semantic integration significantly modulates the amount of RIF that occurs, and that this modulatory influence occurs even when intentional episodic integration is well controlled. To the extent that this factor has gone uncontrolled in prior work, it may account for variability in the amount of RIF observed. In the present experiment, we illustrate the usefulness of this point by examining whether semantic integration may account for otherwise unexplained failures to find RIF under circumstances in which it has been previously observed. We focus in particular on the repeated failures of Butler et al. (2001) to observe RIF under conditions of item-specific cuing.

The findings of Butler et al. (2001) provide an especially interesting case to examine for several reasons. First, the failure to find RIF was consistent. Using a variant of the standard RIF paradigm reported by Anderson et al. (1994), they found little RIF across four types of final tests that provided item-specific cues, including category-plus-two-letter-stem cued recall (e.g., *bird* SP_____, cuing for SPARROW), category-plus-fragment cued recall (e.g., *bird* _P_R_OW), fragment cued recall (e.g., _P_R_OW), and implicit fragment completion (e.g., _P_R_OW). These failures were observed despite finding reliable, though small, RIF on a category cued recall test.[2] Second, these consistent failures led the authors to a strong conclusion with theoretical implications: that item-specific cuing poses a limit on RIF. Third, the conclusion that item-specific cuing is a boundary condition on RIF has received little support in previous and subsequent work, which has provided many demonstrations of RIF on item-specific cuing tests including category-plus-letter stem cues (e.g., Anderson, Bjork, & Bjork, 2000; Anderson et al., 1994; Anderson & McCulloch, 1999; Aslan, Bäuml, & Pastotter, 2007; Bäuml, 2002; Bäuml & Hartinger, 2002; Johansson, Aslan, Bäuml, Gabel, & Mecklinger, 2007; Storm, Bjork, & Bjork, 2007, 2008; Storm et al., 2006;), propositional-plus-unique-letter stem cues (Anderson & Bell,

---

[2] Butler et al. (2001) did find reliable RIF on a category cued recall test. However, this effect was 5.2% and was only reliable by a one-tailed t test. Semantic integration in their materials may also explain why their category cued recall effect is unusually small, compared with the normal category cued recall effect (10%–20%).

2001; Gómez-Ariza et al., 2005), associate-plus-stem cued recall (Kuhl, Dudukovic, Khan, & Wagner, 2007), extralist semantic cues plus unique letter stems (Anderson, Green, & McCulloch, 2000; Johnson & Anderson, 2004; Levy et al., 2007), and letter stem cues in isolation (Bajo, Gómez-Ariza, Fernandez, & Marful, 2006). Nevertheless, Butler et al. clearly failed to find RIF, creating a genuine puzzle. If it can be shown that uncontrolled semantic integration underlies these failures, it would resolve this anomaly and would illustrate the potential to be led theoretically astray by failures to control for semantic integration.

From the results of Experiments 1–3, one might argue, however, that semantic integration does not provide a likely account of Butler et al.'s (2001) findings. Although inter-set associations were prevalent in their stimulus set (their list contained 24 direct associations and 332 indirect associations between practiced items and competitors), they arguably were so weak as to seem irrelevant. The average strength of association between Rp+ and Rp− exemplars in their study was only .07 according to the University of South Florida association norms, a value that would understandably be perceived as near 0. In contrast, the list used in Experiments 1–3 was designed with strong associations. The present effects may be limited to these specialized stimuli. If so, inter-set integration may not be the correct account of their failures to find RIF.

To address these questions, Experiment 4, examined whether the (apparently) weak inter-set associations existing in the stimulus set of Butler et al. (2001) contributed to their failure to observe RIF. To study this, we mimicked our manipulation of inter-set and intra-set semantic integration using only Butler et al.'s stimulus set. We simply rearranged the assignment of their exemplars to retrieval practice conditions, so as to either minimize or maximize associations between practiced and nonpracticed exemplars. We predicted that when items were configured to maximize inter-set associations, as in Butler et al.'s original design, no RIF should be observed; however, when the sets were reconfigured to minimize inter-set associations, significant RIF should be found. If such effects are obtained, it would also indicate that the current semantic integration effects are not limited to our specially designed stimulus set and are a source of concern even when they are apparently weak, as measured by free association norms.

Method

Participants and design.   Sixty-four undergraduates from the University of South Florida participated in exchange for extra credit ($n = 32$ in each integration condition). Participants ranged from 18 to 53 years of age. The experimental design was identical to that of Experiment 1, except for the use of materials drawn from Butler et al.'s (2001) stimulus set, which had a larger number of categories, each with a smaller number of exemplars than were used in our earlier study.

Materials.

Category construction.   Eight categories were selected from the 12 used in Butler et al.'s (2001) study: *Bird*, *Drink*, *Fruit*, *Furniture*, *Insect*, *Metal*, *Profession*, and *Sport*. Four of their categories were excluded (i.e., *Fish*, *Flower*, *Ship*, and *Tool*) because many of their exemplars were not included in the University of South Florida free association norms, and so the presence of associations between practiced and nonpracticed sets could not be

determined. Butler et al.'s filler categories (*Clothing* and *Disease*) were used as fillers.

Exemplar selection.   Not all of the exemplars used by Butler et al. (2001) were used. Butler et al. created three triads of three exemplars each per category, and they had each participant study only two of these triads. The exemplars from the third triad were developed to serve as fillers in their implicit memory testing condition. Because the current experiment did not include an implicit memory test, only two of their original triads for each category were therefore included. We decided which particular two triads to use on the basis of several design constraints. First, because our final memory test cued participants with category-plus-letter stem cues, items triads were excluded if the exemplars shared a first letter with another exemplar from the same category. Second, three exemplars were replaced with exemplars from other triads, because their initial two letter stems had versatility values of less than 100 (Solso & Juel, 1980). The average taxonomic frequency of the exemplars was moderate (average position = 13.3 in Battig & Montague's, 1969, norms). The word frequency of the exemplars was low to moderate and averaged 25.27 ($SD = 64.24$) occurrences per million (Kucera & Francis, 1967).

Design of the integration pattern conditions.   To design the inter-set and intra-set conditions, we assigned the six exemplars of each category of Butler et al.'s (2001) list into subsets of three, so as to either maximize associations between the subsets or within the subsets. For the inter-set condition, the association strengths *between* the triads were stronger than the association strengths *within* the triads. As shown in Table 4, there were stronger direct connections and many more indirect connections (shared associates) between the triads than within the triads. Specifically, there were 24 direct associations (on average, three per category) between the triads, with an average strength of .08. Furthermore, there were a total of 112 indirect connections (i.e., shared associates) between the triads (on average, 14), with an average strength of .06. In contrast, there were fewer direct associations (one per category) within the triads. and they were weaker ($M = 0.03$). There were also fewer shared associates (total = 84; on average, 10.4 per category) that were weaker ($M = 0.03$) within the triads. The resulting between-triad and within-triad strength characteristics for the inter-set condition are similar to those of Butler et al.'s original list shown in Table 4. That is, even though we only used eight of their 12 categories and two of their three triads per category, our inter-set condition was extremely similar to theirs in the quantity and strength of inter-set associations.

For the intra-set condition, the triads were rearranged so the association strengths were weak *between* triads but stronger *within* the triads. As shown in Table 4, there were more direct connections and many more indirect connections (shared associates) within the triads than between the triads. Specifically, there were 25 direct associations (on average, 3.12 per category) within the triads, with an average strength of .06. In addition, there were 114 indirect connections (14 per category) within the triads. In contrast, there were fewer direct associations (i.e., three; less than one per category) between the triads, and they were weaker ($M = 0.06$). There were also fewer shared associates (95) that were weaker ($M = 0.04$) between the triads. Thus, we created an intra-set condition similar to those developed for our stimulus sets by simply rearranging Butler et al.'s (2001) exemplars into different triads so as to minimize inter-set associations.

Table 4

Normative Measures Between and Within the Stimulus Sets in the Inter-Set and Intra-Set Conditions of Experiment 4 and Butler et al.'s (2001) Experiments

Variable	Associative characteristics	Inter-set condition, Experiment 4	Butler et al.'s (2001) stimulus sets	Intra-set condition, Experiment 4
Between triads	No. of direct associations	24	39	3
	Mean associative strength	.08 (.11)	.08 (.10)	.06 (.04)
	No. of indirect associations	112	332	95
Within triads	No. of direct associations	8	7	25
	Mean associative strength	.03 (.02)	.04 (.02)	.06 (.06)
	No. of indirect associations	84	109	114

	Target characteristics	Experiment 4	Butler et al.'s (2001) stimulus sets
All items	Concreteness	5.55 (1.73)	5.67 (1.48)
	Associative set size	13.73 (6.03)	13.55 (5.80)
	Connectivity	2.00 (1.15)	1.85 (1.08)
	Resonance	.35 (0.24)	.33 (0.24)
	Frequency	25.27 (64.24)	24.52 (55.07)

Note. Values in parentheses indicate standard deviations.

Procedure. The procedures were the same as those in Experiment 1. The presentation order in the study phase followed the same criteria: Items were pseudorandomized in blocks so that each block contained one exemplar from each category resulting in six blocks of 10 items (each block containing eight items from the experimental categories and two items from the filler categories). All other ordering restrictions and procedures for the study phase, retrieval practice phase, and test phase were identical to those in Experiment 1.

Results

A mixed-subjects ANOVA was conducted, with integration pattern (intra-set, inter-set) as a between-subjects factor, and retrieval practice status (Rp+, Rp−, Nrp1, Nrp2) as a within-subjects factor. Unlike in prior studies, counterbalancing did interact with some of our hypotheses concerning RIF. In particular, the amount of RIF interacted with counterbalancing separately within the intra-set and inter-set conditions. This interaction arises because one counterbalancing condition out of eight (different in each group) anomalously showed a pattern reliably opposite to what was theoretically expected. The reliability of the present findings actually grows if these counterbalancing groups are excluded. We include them, however, because none of our conclusions are affected by doing so.

Retrieval practice success. Retrieval practice success rate did not vary between the inter-set ($M = 0.94$, $SD = 0.06$) and the intra-set ($M = 0.91$, $SD = 0.08$) conditions, $F(1, 48) = 3.09$, $p = .085$.

Overall RIF. The overall RIF effect (baseline vs. Rp−) collapsed across the integration pattern manipulation was significant and revealed that Rp− items were recalled more poorly than were baseline items (−7% RIF effect), $F(1, 48) = 16.17$, partial $\eta^2 = .252$, $p < .001$.

Semantic integration effects. Of greater concern, however, is whether RIF varies with semantic integration pattern, using a simple rearrangement of Butler et al.'s (2001) stimuli. Crucially, the interaction between RIF (baseline vs. Rp−) and integration pattern was significant, $F(1, 48) = 12.66$, partial $\eta^2 = .209$, $p = .001$. There was RIF (−13%) in the intra-set condition, $F(1, 24) = 20.65$, partial $\eta^2 = .463$, $p < .001$, and no reliable RIF (−1%) in the inter-set condition ($F < 1$). These results strongly support the idea that preexisting associations between practiced items and their nonpracticed competitors moderate the RIF effect and generalize the results of Experiments 1–3.

Benefits of retrieval practice on practiced items. The overall recall benefit (baseline vs. Rp+) for practiced items (Rp+; collapsed across the integration pattern) was significant, $F(1, 48) = 133.134$, partial $\eta^2 = .735$, $p < .001$, with 23% more practiced items recalled in the final test ($M = 0.87$, $SD = 0.12$) compared with baseline items ($M = 0.64$, $SD = 0.17$). The interaction between facilitation of Rp+ items and integration pattern was not significant ($F < 1$), with retrieval practice producing a benefit for practiced items relative to baseline in both the intra-set condition (+25%), $F(1, 24) = 47.63$, partial $\eta^2 = .665$, $p < .001$, and the inter-set condition (+21%), $F(1, 24) = 94.23$, partial $\eta^2 = .797$, $p < .001$.

Output interference. Output interference effects within the baseline condition were significant, $F(1, 48) = 11.25$, partial $\eta^2 = .190$, $p = .002$, with baseline items cued in the last half of the category block recalled at a lower rate ($M = 0.64$, $SD = 0.17$) than those cued in the first half ($M = 0.71$, $SD = 0.16$). The interaction between output interference and integration pattern was again not significant ($F < 1$). Unlike in prior experiments, the pattern of output interference was at best weakly consistent with the idea that semantic integration modulates output interference. Although there was a significant decrement in recall between the first half ($M = 0.72$) and last half of the baseline ($M = 0.64$) in the intra-set condition, $F(1, 24) = 7.66$, partial $\eta^2 = .242$, $p = .011$, but no significant decrement between the first half ($M = 0.70$) and the last half of the baseline ($M = 0.64$) in the inter-set condition, $F(1, 24) = 3.95$, partial $\eta^2 = .141$, $p = .058$, the differences in output interference were very small (8% vs. 6%). Thus, unlike Experiments 1–3, the present results only weakly suggest that associations moderate output interference effects.

Discussion

Experiment 4 replicates the core findings of Experiments 1–3 using Butler et al.'s (2001) stimulus set. By simply rearranging the assignments of Butler et al.'s exemplars into practiced and unpracticed sets that either maximized or minimized inter-set integration, we were able to make RIF disappear or appear robustly. Indeed, the condition most like the one used by Butler et al.—the inter-set condition—very clearly replicates their failure to find RIF, even though sizeable RIF was found with precisely the same exemplars, when arranged according to our intra-set configuration (no inter-set associations). The results indicate that Butler et al.'s failure to replicate RIF with item-specific cues was due to their introduction of preexisting associations between the practiced items and their nonpracticed competitors.

The strength of the effects of preexisting associations on RIF in the present study is surprising, given the relatively weak associations that exist in Butler et al.'s (2001) stimulus set. Even though their inter-exemplar associations were not nearly as strong as those in the stimulus set we designed for Experiments 1–3, RIF was still entirely eliminated in the inter-set condition. One possibility is that free association values may underestimate the likelihood that associations will have an influence on performance when participants encode many different exemplars from the same category. Whatever the proper explanation may be, the current findings indicate that even weak associations are sufficient to produce semantic integration that moderates RIF effects.

Combined Analysis

To more completely explore the effects of semantic integration and their causes with greater statistical power, the data from our four very similar experiments were combined, and a combined analysis was performed. A mixed-subjects ANOVA was conducted on the combined data, with integration pattern (intra-set, inter-set), counterbalancing (eight counterbalancing conditions), and experiment (Experiments 1–4) as between-subjects factors, and with retrieval practice status (Rp+, Rp−, Nrp1, Nrp2) as a within-subjects factor.

RIF and Semantic Integration Benefits

The overall RIF effect collapsed across our integration pattern manipulation was significant, $F(1, 320) = 22.78$, partial $\eta^2 = .066$, $p < .001$, with the recall of Rp− items being lower than that of baseline items (−6% RIF effect). This RIF effect was very clearly moderated by integration pattern, $F(1, 320) = 24.97$, partial $\eta^2 = .072$, $p < .001$ (see the top panel of Figure 3), with highly robust RIF (−12%) in the intra-set condition, $F(1, 160) = 43.49$, partial $\eta^2 = .214$, $p < .0001$, and no evidence for RIF (0%) in the inter-set condition ($F < 1$).

The Role of Baseline Recall in Producing Integration Benefits

Although the semantic integration effect on RIF seems clear, one might be concerned that the reduced RIF in the inter-set condition arose more from a decrease in the baseline in that condition rather than from improved Rp− performance. The baseline in the inter-set condition tended to be lower than that in the
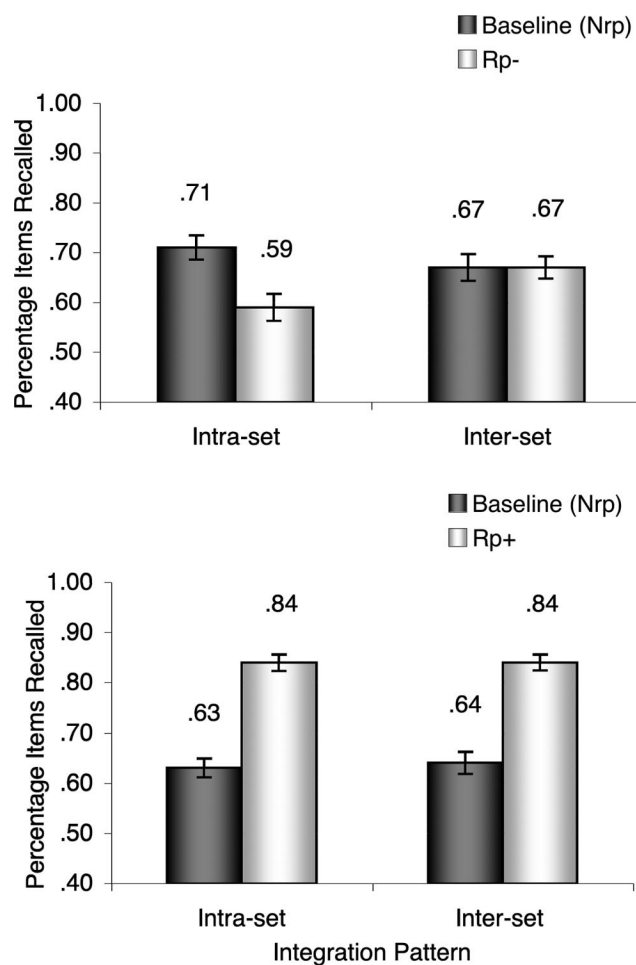

Figure 3. Results of combined analyses: Significant retrieval-induced forgetting effect only in the intra-set condition (top panel) but significant retrieval practice benefit in both the intra-set and inter-set conditions (bottom panel). Nrp = no retrieval practice; Rp+ = practiced items; Rp− = their nonpracticed competitors. Error bars indicate standard error (*SE*).

intra-set condition in Experiments 2 and 4. We examined this difference across all of our experiments, and indeed, there was a small (4%) but significant difference between the baselines in the intra-set ($M = 71\%$) and the inter-set ($M = 67\%$) conditions, $F(1, 320) = 4.11$, partial $\eta^2 = .013$, $p < .044$. This baseline difference is unlikely to explain the reduced RIF, however. First, even if we treated the lower inter-set Nrp baseline as the best estimate of baseline recall, the recall of Rp− items ($M = 59\%$) in the intra-set condition was still lower than recall in that baseline ($M = 67\%$). In fact, Rp− items in the intra-set condition (59%) were recalled more poorly than in the inter-set condition (67%), $F(1, 320) = 10.01$, partial $\eta^2 = .030$, $p = .002$, indicating that inter-set Rp− items benefited from inter-set associations. Second, Experiments 1 and 4 showed no reliable difference in baselines ($F < 1$ in both cases), illustrating that the effects of semantic integration are present even when the baseline difference is absent. It is unlikely that the baseline recall difference caused apparent semantic integration benefits.

The Relation Between Strengthening and RIF

The overall benefit in recall of practiced items was significant, $F(1, 320) = 286.01$, partial $\eta^2 = .820$, $p < .001$, with 21% more practiced items recalled ($M = 0.84$, $SD = 0.17$) compared with baseline items ($M = 0.63$, $SD = 0.22$). More interestingly, as shown in the bottom panel of Figure 3, the facilitation of Rp+ items did not vary as a function of integration pattern ($F < 1$). Retrieval practice produced a benefit of 21% for practiced items in the intra-set condition where forgetting was present, $F(1, 192) = 174.99$, partial $\eta^2 = .477$, $p < .001$, and a benefit of 20% for practiced items in the inter-set condition where forgetting was absent, $F(1, 192) = 132.58$, partial $\eta^2 = .408$, $p < .001$. This establishes that differences in RIF are not attributable to differences in strengthening for practiced items, and is compatible with studies documenting the strength-independence property of RIF (see Anderson, 2003, for a review).

The Modulation of Output Interference by Integration Pattern

Output interference within the baseline condition collapsed across the integration pattern conditions was significant, $F(1, 320) = 19.21$, partial $\eta^2 = .057$, $p < .001$, with items cued in the last half of the category recalled at a -6% lower rate ($M = 0.63$, $SD = 0.24$) compared with items cued in the first half ($M = 0.69$, $SD = 0.20$). Unlike in the individual experiments, the interaction between output interference and integration pattern was significant, $F(1, 320) = 5.29$, partial $\eta^2 = .016$, $p = .022$, with results paralleling the effects of semantic integration on RIF. Output interference was found in the intra-set condition (-8%), $F(1, 192) = 19.38$, partial $\eta^2 = .092$, $p < .001$. However, output interference was marginally significant in the inter-set condition (-3%), in which there were strong associations between the first and last halves, $F(1, 192) = 3.99$, partial $\eta^2 = .020$, $p = .055$. Although the moderating effects of semantic integration are weaker for output interference, the results are consistent with the proposal that RIF and output interference are produced by the same mechanisms and are thus both subject to semantic integration effects.

Reduced output interference in the inter-set condition might arise if participants used items in the first half of a category to help them recall items in the second half. We measured this strategy in our questionnaire by using a 5-point Likert scale. When the analysis was focused on participants who claimed to have never done this (average rating of 1 [*never*] across all categories rated) in all of the experiments we conducted ($n = 20$ and $n = 16$ participants for the inter- and intra-set conditions, respectively), the inter-set condition showed less output interference (Nrp1 – Nrp2 = 63%−73% or 10% facilitation; $F < 1$) than did the intra-set condition (Nrp1 – Nrp2 = 72%−64% or 8% impairment), $F(1, 15) = 4.29$, partial $\eta^2 = .222$, $p = .056$, suggesting that reduced output interference does not depend upon the intentional use of associations (nor did RIF).

Discussion

This combined analysis addresses several important aspects of the case for a role of semantic integration in modulating RIF. One might have been concerned, for example, that the difference in the amount of RIF across the inter-set and intra-set conditions might have been due to either differences in baseline recall across these conditions or to differences in the extent to which Rp+ items are strengthened. By combining our experiments to examine these hypotheses, we were able to address these issues in an analysis with considerable statistical power. These analyses show that there are no reliable differences in strengthening across the intra-set and inter-set integration conditions, despite highly robust differences in RIF. Moreover, although a small difference in baseline performance (4%) was uncovered, it is clear that this difference cannot explain the reduced RIF in the inter-set integration condition. These analyses also revealed a significant interaction between pattern of integration and output interference, paralleling the effects of integration on RIF. Taken together, these findings provide strong converging evidence that semantic integration moderates RIF, as hypothesized.

General Discussion

The four experiments reported here establish several findings concerning how preexisting knowledge moderates the inhibitory effects of retrieval. First, the magnitude of RIF that was observed was strongly affected by the presence of a priori associations between exemplars. In all of the current experiments, associations linking retrieval-practice targets and their competitors reduced RIF. Indeed, RIF was largely eliminated by such relationships. This semantic integration effect occurred in the specialized stimulus set we designed to maximize these associations in Experiments 1–3. Strikingly, however, these findings also generalized to a second stimulus set developed by Butler et al. (2001). The latter finding is noteworthy because the target–competitor associations in Butler et al.'s list were fewer and weaker than the ones in our stimulus set, and one could have justifiably doubted whether the degree of inter-set association present would be sufficient to support semantic integration. Apparently, semantic integration effects occur even with an average associative strength of .07 (University of South Florida association norms), demonstrating the robustness of this effect in the face of weak associations between sets.

Second, semantic integration effects are specific to target–competitor associations. When participants studied the same exemplars with exactly the same inter-associations, RIF occurred, provided that the inter-item associations did not link retrieval-practice targets to their competitors. Thus, the mere presence of associations does not eliminate RIF. These findings converge with research showing that intentional encoding of target–competitor, but not competitor–competitor, relationships eliminates RIF (Anderson, Green, & McCulloch, 2000; also see Bäuml & Hartinger, 2002). This dissociation occurred both with stimuli designed to maximize inter-item associations and with the stimuli used by Butler et al. (2001). Using Butler et al.'s stimuli, we made RIF appear or disappear, simply by rearranging which items were practiced so as to vary the strength of target–competitor associations.

Third, semantic integration effects occur in the absence of any intention to encode relationships between items. In all four experiments, the pattern of semantic integration moderated RIF even though participants were not instructed to focus on encoding inter-exemplar relationships. Thus, at a minimum, overt instruc-

tions to integrate are unnecessary. Of course, participants may have spontaneously sought inter-item associations during study. Although likely, several findings indicate that semantic integration benefits occur even when participants do not engage in this behavior. In Experiment 2, semantic integration effects occurred even when participants incidentally encoded exemplars, eliminating motivation to seek inter-item associations to improve memory. Postexperimental rating scales confirmed that the vast majority of participants did not intentionally integrate items under these conditions, and the semantic integration effect remained robust when the few intentional integrators in the study are excluded. Integration effects even occurred when we controlled for unintentional episodic integration that might have occurred during encoding. We considered whether participants might have rated exemplars comparatively (i.e., by considering the current item in relation to prior, similar items), perhaps associating the exemplars in the process. Even when we considered only those participants who claimed to never have done this type of comparison, and who also claimed to not intentionally integrate, semantic integration effects remained. Taken together, these findings indicate that semantic integration effects occur without intention and likely reflect the influence of preexisting semantic structure on RIF in episodic memory.

Finally, the current findings validate earlier work on episodic integration. Specifically, we found that increasing study time decreases RIF while also increasing the rate of self-reported integration. In Experiment 3, participants received 5 s or 10 s of study time per exemplar. Strikingly, RIF was reduced in the 10-s group in both the inter-set and intra-set conditions. More importantly, Experiment 3 indicates that semantic and episodic integration are two distinct moderators of RIF: Collapsing over our study time manipulation, participants in the intra-set and inter-set conditions study showed –9% and 0% RIF, respectively; similarly, collapsing across our semantic integration conditions revealed –10% RIF and +1.5% facilitation in the 5-s and 10-s study conditions, respectively. Thus, study time and semantic knowledge additively influenced RIF. These findings extend work by Anderson and McCulloch (1999), showing that increasing study time through repeated exposures reduces RIF also while increasing self-reported integration.

The current studies yielded two additional observations that support the notion that semantic integration reduces RIF, and that question the idea that associative interference underlies RIF. First, semantic integration also modulated output interference in the same way that it modulated long-lasting RIF. In a combined analysis conducted across all of the current experiments, there was greater output interference for baseline categories in the intra-set condition than in the inter-set condition. Thus, when the first half of the items tested in the baseline categories had associations to items tested in the second half, RIF was reduced. To the extent that output interference is another manifestation of RIF arising during the final test, these findings provide converging evidence concerning the effects of semantic integration on RIF.

Second, we consistently failed to find a relationship between the degree of strengthening exhibited by practiced items and the amount of RIF. In the combined analysis across 300 participants, the facilitation of practiced items was highly similar for the intra-set condition (21%) in which significant RIF was observed (−12%) and for the intra-set condition (20%) in which no RIF was found (0%). This observation is significant in that it speaks against

any obvious causal role of competitor strengthening in RIF. If strengthening practiced items caused RIF, comparable RIF should have been observed in these conditions. This lack of a relationship between the facilitation and inhibition is another example of the property of strength independence found numerous times in studies of RIF (Anderson, 2003).

Although the influence of semantic structure on RIF is clear in the current studies, several issues remain. One concerns the role of competitor–competitor associations in moderating RIF. In our design, we sought to hold the exemplars constant between our inter-set and intra-set conditions and to ensure that precisely the same preexisting associations were present in each case. This enabled us to demonstrate that we could make RIF come and go by varying how well these associations matched the inter-set association patterns present in Butler et al. (2001). However, these goals required that we vary competitor–competitor association strength along with target–competitor association strength. So, for example, when associations existed between retrieval-practice targets (e.g., Horse, Lion, Walrus, Duck) and their competitors (e.g., Pony, Tiger, Seal, Goose), there were necessarily fewer associations among the retrieval practice targets themselves or among the competitors. In contrast, when there were few associations between the targets (e.g., Horse, Pony, Lion, Tiger) and their competitors (e.g., Walrus, Seal, Duck, Goose), there were necessarily more associations among the targets and among the competitors. It is thus fitting to wonder to what extent the current differences in RIF across the target–competitor and competitor–competitor conditions arose from variations in target–competitor strength or from variations in competitor–competitor strength.

Fortunately, data relevant to this question exist in other published studies. One clearly established finding is that significant RIF occurs when there are no inter-associations amongst competitors or amongst targets. For instance, the studies of Anderson et al. (1994); Anderson, Bjork, and Bjork (2000); and Anderson and McCulloch (1999) all used categorical stimuli like the current ones, except that a priori associations between exemplars were deliberately eliminated. RIF was observed in all cases. Moreover, RIF has been observed with propositional materials in which the competing facts were semantically unrelated to one another and to the retrieval practice targets (e.g., Anderson & Bell, 2001; Gómez-Ariza et al., 2005; MacLeod, 2002). RIF has even been found when there is only a single target and a single competitor, in which there could not possibly be inter-item associations with other studied competitors (e.g., Shivde & Anderson, 2001; Veling & van Knippenberg, 2004). Together, these findings indicate that it should have been possible to observe RIF in our target–competitor condition even with the very low degree of competitor–competitor semantic integration, had it not been for the existence of target–competitor associations. Thus, the elimination of RIF in the inter-set condition is likely to reflect the protective role of target–competitor associations. This does not imply, however, that competitor–competitor associations do not influence RIF as well. Competitor–competitor associations may increase RIF (Anderson, Green, & McCulloch, 2000). As such, it would be profitable to vary a priori competitor–competitor relatedness independent of target–competitor relatedness to examine its influence on RIF, as has been done for episodic similarity (Anderson, Green, & McCulloch, 2000).

One final issue concerns the extent to which semantic integration effects might have been caused by episodic integration during retrieval practice. Although Experiment 2 eliminated episodic integration during encoding, participants might have episodically integrated during retrieval practice trials. For instance, a participant who recalled "Horse" in response to Animal Ho___ during retrieval practice may have used the remaining time before the next trial to recall additional exemplars related to Horse (e.g., Pony) and might have rehearsed them together. In the intra-set condition, this supplementary retrieval practice would not have benefited Rp− items, because the most closely related associate that would be retrieved would also be a Rp+ item. In the inter-set condition, by contrast, Rp+ items would be more likely to cue the associated Rp− items. This would benefit Rp− items both by offsetting increments in inhibition on those items by compensatory facilitation and by episodically integrating the items together. If so, the mere presence of a semantic association between Rp+ and Rp− items may be insufficient to prevent RIF, unless that association causes episodic integration during retrieval practice.

Fortunately, we assessed participants' use of covert retrieval practice in the present experiments using a postexperimental questionnaire. As expected, participants did use the additional time during retrieval practice to covertly rehearse extra items that were not currently being cued. On a 5-point Likert scale (1 = *none of the time*, 2 = *25% of the time*, 3 = *50% of the time*, 4 = *75% of the time*, 5 = *100% of the time*), the average rating was 2.87, suggesting that covert rehearsal did occur. To examine whether the reduction in RIF in the inter-set condition depended on this strategy, we first conducted a median split, across all studies, matching for counterbalancing and for experiment, based on participants' self-reported covert retrieval practice ratings. Importantly, little RIF was found in the inter-set condition for participants reporting lower amounts of covert retrieval practice ($M = 2.06$ or "25% of the time" on our scale; –2% RIF) and for participants who reported higher amounts ($M = 3.70$, nearly 75% of the time on our scale; 1% facilitation). One might be concerned that even the lower amount of covert practice in the low group may have eliminated RIF. To address this, we isolated a group of 30 participants who reported never having used this strategy at all (an average rating of 1, or 0% of the time). Even this group showed no evidence (0%) of RIF in the inter-set condition. These findings suggest that, as far as we are able to measure, covert retrieval practice played little role in reducing RIF in the inter-set condition. Nevertheless, it would clearly be desirable to experimentally control this strategy (perhaps, for instance, through divided attention or reduced retrieval practice time) to provide converging evidence for this conclusion.

Taken together, the current findings indicate that semantic integration moderates the amount of RIF in standard episodic memory designs. Importantly, these effects do not require participants to intentionally integrate items during study, indicating that they reflect an effect of preexisting semantic structure. As such, semantic integration constitutes a novel boundary condition on RIF that is of both empirical and theoretical relevance. On an empirical level, discovery of this boundary condition will likely prove useful in accounting for otherwise unexplained variability in the phenomenon of RIF. In fact, as seen in Experiment 4, lack of control over semantic integration provides a compelling account of repeated failures to find RIF on item-specific cuing tests reported by Butler et al. (2001), helping to align these data with the rest of the literature on RIF. This revisiting of Butler et al.'s findings vividly illustrates how even relatively weak target–competitor relations might influence performance. On a theoretical level, semantic integration provides an important constraint on accounts of RIF that may help us to better understand the phenomenon, a topic to which we turn next.

Theoretical Accounts of Semantic Integration Effects

Semantic integration effects bear similarity to the effects of episodic integration and episodic similarity reported previously (Anderson, Green, & McCulloch, 2000; Anderson & McCulloch, 1999). Many of the mechanisms proposed to explain those findings may explain the current ones. The key difference is that the current dynamics reflect an influence of semantic structure on episodic memory. We discuss semantic generalization as a key mechanism that could account underlie the current findings, and other possible mechanisms.

The semantic generalization hypothesis.   The current findings accord well with the semantic generalization hypothesis developed in Anderson, Green, and McCulloch (2000) to explain the differential effects of target–competitor and competitor–competitor similarity on RIF. This hypothesis builds on the pattern suppression model proposed by Anderson and Spellman (1995). According to the model, a memory item is represented by features in a distributed network. The more semantically or episodically related two items are, the greater their feature overlap, and the more they compete with one another. According to the model, successful retrieval practice activates all of the features in the target pattern and strengthens them, including those features that are shared with other similar, competing items; however, features unique to the competing items are inhibited to the extent that their activation interferes with target retrieval. These outcomes are represented graphically in Figure 4 by darkened features (strengthening) and "X-ed" circles (inhibition).

According to the semantic generalization hypothesis, although feature overlap leads two items to compete, high target–competitor similarity should actually reduce RIF, because strengthening features of the practiced items will enhance the ability to recall competitors. For example, retrieval practice of HORSE (i.e., Rp+ item or target) should strengthen HORSE's features and inhibit those features unique to nonpracticed competitors (i.e., Rp− items). As shown in the top panel of Figure 4, high similarity between targets (e.g., HORSE) and competitors (e.g., PONY) will strengthen a large proportion of the competitor's features because of feature overlap. Thus, the seven features that PONY shares with HORSE will be strengthened when HORSE is retrieved. Because only the three unique features of PONY are inhibited and the rest are strengthened, and because the model assumes that recall probability reflects the summed activity of an item's units, competitors that are similar to targets should show less RIF. Essentially, feature overlap with targets limits the number of distinctive features that can be inhibited while increasing the
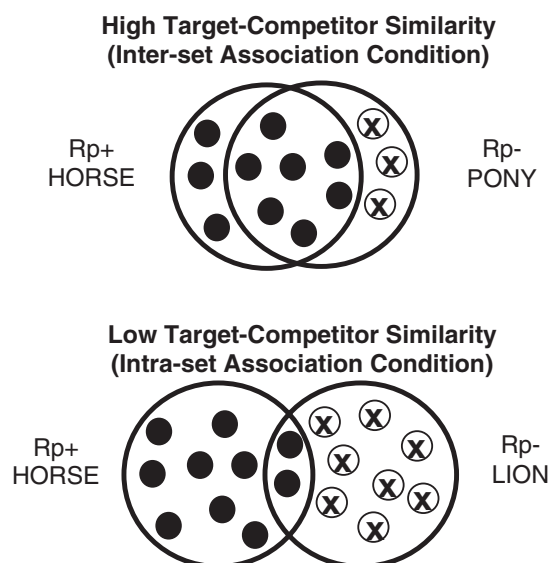
**High Target-Competitor Similarity
(Inter-set Association Condition)**

Rp+
HORSE

Rp-
PONY

**Low Target-Competitor Similarity
(Intra-set Association Condition)**

Rp+
HORSE

Rp-
LION

Figure 4. Illustration of how high target–competitor similarity (top) and low target–competitor similarity (bottom) influence inhibition, according to the semantic generalization hypothesis. When similarity is high (i.e., strong associations between the target [Rp+] and competitor [Rp−]), the strengthening effects of retrieval practice are implicitly generalized because strengthening occurs on the features shared by the target and competitor, resulting in attenuated retrieval-induced forgetting (RIF). However, when similarity is low (i.e., weak associations between the target and competitor), there are more unique features of the competitor to inhibit, resulting in significant RIF. Rp+ = practiced items; Rp− = their nonpracticed competitors.

influence of facilitated features, offsetting whatever inhibition takes place.[3,4]

The model makes very different predictions, however, when targets and competitors are only weakly related, such as with HORSE and LION. As shown in the bottom panel of Figure 4, when target–competitor similarity is lower, strengthening the same number of features (i.e., 10 features) strengthens the competitor less because it shares fewer features with the target (Rp+ item), resulting in inhibition of LION's eight unique features. Assuming there is constant facilitation of practiced features across both high (top panel) and low (bottom panel) similarity cases, the probability of recalling the competitor should be reduced with low target–competitor similarity (bottom panel) because higher proportions of the competitor's features are inhibited. For example, only 30% of the Rp− item's representation is inhibited when target–competitor similarity is high (top), compared with 80% when target–competitor similarity is low (bottom). The model therefore predicts significant RIF when the target and competitor are moderately similar. Entirely dissimilar, nonoverlapping items will not compete and therefore will not be inhibited.[5]

The semantic generalization hypothesis accounts for the current findings quite well.[6] Reduced RIF in the inter-set condition, relative to the intra-set condition, falls naturally out of the model's predictions concerning target–competitor similarity. Moreover, because target–competitor similarity effects emerge from feature overlap, they do not require the participant to explicitly encode relationships or even to be aware of inter-item connections. These

characteristics explain why participants show target–competitor similarity effects even when they encode items incidentally and when they report no integration, either during encoding or retrieval practice. The semantic generalization hypothesis also predicts that target–competitor similarity effects should not depend upon participants' recalling practiced items prior to the competitors on the final test. Because a competitor's overlapping features are strengthened by retrieval practice, these items should remain preserved on the final test, even when tested prior to Rp+ items. Finally, all of the foregoing dynamics should occur during the final test because retrieval practice and test trials engage similar mechanisms, explaining the parallel findings of integration on RIF and on output interference. Thus, the semantic generalization hypothesis provides a promising approach to explaining current and previous effects of integration on RIF (see Footnote 5).

Other accounts. At least two other accounts of target–competitor similarity effects are possible. First, the current findings can be explained by *mediated retrieval* during the final test (Anderson & McCulloch, 1999). According to this hypothesis, an exemplar can be retrieved in two ways: by directly recalling it, using the category name, or by recalling it via an association from another exemplar. So, for example, participants might recall "HORSE" directly by using ANIMAL as a retrieval cue; alternatively, they might recall "HORSE" indirectly by using the

---

[3] As shown in Appendix A in the online supplemental materials, most of the highly related exemplar pairs in Experiments 1–3 are similar (e.g., Horse, Pony), though some are related associatively (e.g., Toaster Oven). This arose because we used the University of South Florida norms to operationally define relatedness. These norms record the frequency with which one idea elicits another and are silent about the nature of the relationship. It would be desirable to establish whether similarity or associative relatedness is essential to producing the semantic integration effect.

[4] The effect of inter-exemplar similarity on RIF must be distinguished from that of taxonomic frequency (Anderson et al., 1994). Taxonomic frequency reflects the associative strength linking a category to an exemplar. In general, when retrieving a target from memory, competitors with higher taxonomic frequency are thought to cause more interference. This dimension of competition is not the same as inter-competitor similarity discussed here, though both ought to trigger elevated inhibitory control.

[5] A special case arises when two entirely dissimilar items are associated to the same cue (as in classical A–B, A–C interference studies). By being encoded with the same cue, two otherwise dissimilar items take on contextual similarity that renders them competitive when their shared cue is provided (Anderson, Green, & McCulloch, 2000).

[6] One might be concerned that the semantic generalization hypothesis is inconsistent with the notion that inhibition helps to distinguish a target from similar competitors. By this view, one might intuitively expect more RIF for similar items, and less for very dissimilar ones, whereas the present argument implies the opposite. This inconsistency is more apparent than real. First, this intuitive analysis neglects the powerful role played by facilitation of overlapping features, which clearly must influence performance. An alternative reading, consistent with the semantic generalization hypothesis, is that the distinctive features of a highly similar competitor are more inhibited than those of a less similar competitor. Given this view, there still will be a point at which facilitation from overlapping features compensates for featural inhibition. If so, there should be a nonmonotonic relationship between similarity and inhibition such that inhibition increases and then decreases with increasing similarity (see Anderson, Green, & McCulloch, 2000, for a thorough discussion of similarity and inhibition in long-term memory).

"ANIMAL" cue to first recall "PONY" and only then recall "HORSE" via its association to "PONY."

To explain why RIF is eliminated in the present inter-set conditions requires special assumptions. Because inter-exemplar associations are present in both the practiced and baseline categories to the same degree, both conditions should benefit similarly, improving overall recall but not eliminating inhibition. However, because of retrieval practice, participants should be able to recall Rp+ items (e.g., Pony) more effectively, allowing them to more easily access the exemplar–exemplar (e.g., Pony–Horse) association than they can for items in the baseline condition. This differential access to useful exemplar–exemplar pathways may offset the decrement in recall that would otherwise be produced by inhibition, causing a null RIF effect. By this view, some Rp− items may well be inhibited, but the behavioral manifestation of this inhibition (RIF) is masked (Anderson, 2003; Anderson & McCulloch, 1999).

Although the mediated retrieval hypothesis can account for main features of the current data, key predictions fail in other related paradigms. According to mediated retrieval, inter-set integration benefits should depend on participants' ability to access the Rp+ item from the final test cue, so that that item can be used to cue recall of Rp− items. Anderson, Green, and McCulloch (2000), however, found target–competitor similarity effects even when the ability to use exemplar–exemplar associations was eliminated by using an independent test cue for Rp− items, unrelated to Rp+ items. Thus, the test cue should not have elicited recall of Rp+ items and should have prevented mediated retrieval, an assumption confirmed by exit questionnaires. Pattern suppression provides a better account of these data. Because the current studies used a similar manipulation, and because the findings of Anderson et al. favor pattern suppression, semantic integration effects may be produced by this process as well. Nevertheless, Anderson et al. studied episodic integration, not semantic integration. It would be useful to conduct the current study using the independent probe method to verify that semantic inter-set integration benefits would still emerge. In addition, mediated-retrieval might be tested by dividing attention during the final test, which should reduce or eliminate the benefits of target–competitor integration by discouraging elaborate indirect retrieval strategies.

A final approach to explaining the current effects rests on the idea that integration reduces retrieval competition between the retrieval practice targets (Rp+ items) and their competitors (Rp− items). Research on fan effects suggests that the competition that ordinarily occurs during the retrieval of propositional information can be eliminated when otherwise competing facts can be integrated into a coherent mental model describing a situation (e.g., Radvansky, Spieler, & Zacks, 1993; Radvansky, Wyer, Curiel, & Lutz, 1997; Radvansky & Zacks, 1991). If integrating the associates of a retrieval cue reduces competition, then perhaps integrating two exemplars within a category reduces their tendency to compete (Anderson & McCulloch, 1999). In fact, if two associates of a cue are reciprocally connected, they may mutually activate one another, functioning as a coherent "object" in memory. If so, perhaps semantic integration reduces RIF because it eliminates the need for inhibitory control. Thus, reduced RIF in the inter-set condition might reflect *diminished suppression* and not compensatory factors, such as the facilitation of overlapping features or the addition of mediated retrieval routes. One weakness of this hy-

pothesis, however, is that whereas a given Rp− item may be integrated with its paired Rp+ item, it is not integrated with the remaining three Rp+ items and so should compete and be subject to inhibition during retrieval practice of those other items. The fact that no RIF at all was observed in the present studies suggests that this mechanism by itself may not be sufficient to accommodate the present data.

Concluding Remarks

Research on the boundary conditions of a phenomenon can place important constraints on theoretical accounts of that phenomenon. Previous research has established boundary conditions that limit the magnitude of RIF, including episodic integration (Anderson & McCulloch, 1999), episodic target–competitor similarity (Anderson, Green, & McCulloch, 2000), and the degree of competition during retrieval (Anderson et al., 1994; Anderson, Bjork, & Bjork, 2000). The present work examined whether semantic integration poses another boundary condition. Specifically, when a semantic association links a target and a competitor, RIF is eliminated. Using exactly the same stimuli with precisely the same semantic associations, however, we showed sizable RIF as long as the associations did not link targets and competitors. Indeed, RIF appeared or disappeared within the same experiment when we varied the pattern of semantic integration, explaining at least one failure to find RIF previously reported (Butler et al., 2001).

The current findings demonstrate the semantic structure can influence episodic retrieval and, in particular, the dynamics of retrieval competition and inhibition. Earlier work has shown that semantic factors, such as taxonomic frequency (e.g., Anderson et al., 1994; Bäuml, 1998) and the relative strength of two meanings of a homograph (Shivde & Anderson, 2001), influence the competition during episodic retrieval and, correspondingly, the amount of inhibition. Moreover, semantic retrieval can impair episodic memory of competitors (Bäuml, 2002), and episodic retrieval impairs semantic competitors (Starns & Hicks, 2004), indicating the inter-dependence of these types of knowledge during retrieval. Complementing that earlier work, the current work shows that semantic structure protects competing items from RIF. Thus, even if people do not intentionally encode inter-relationships between targets and competitors, a prior history of being related is sufficient to attenuate RIF. That semantic integration protects against RIF speaks to the generality of integration as a boundary condition. Whatever the theoretical process underlying these effects, the current findings establish that whether remembering causes forgetting depends on the structure of the memories being retrieved and their semantic relationships with other aspects of our past.

References

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanism of forgetting. *Journal of Memory and Language, 49,* 415–445. doi:10.1016/j.jml.2003.08.006

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130,* 544–570. doi:10.1037/0096-3445.130.3.544

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

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063–1087. doi:10.1037/0278–7393.20.5.1063

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. doi:10.1037/0278-7393.26.5.1141

Anderson, M. C., & Levy, B. J. (2010). On the relation between inhibition and interference in cognition. In A. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork.* Washington, DC: American Psychological Association.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 608–629. doi:10.1037/0278-7393.25.3.608

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (2nd ed., pp. 237–313). San Diego, CA: Academic Press.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68–100. doi:10.1037/0033-295X.102.1.68

Aslan, A., Bäuml, K.-H., & Pastotter, B. (2007). No inhibitory deficit in older adults' episodic memory, *Psychological Science, 18,* 72–78. doi:10.1111/j.1467-9280.2007.01851.x

Bajo, M. T., Gómez-Ariza, C. J., Fernandez, A., & Marful, A. (2006). Retrieval-induced forgetting in perceptually driven memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1185–1194. doi:10.1037/0278-7393.32.5.1185

Baran, B., Wilson, J., & Spencer, R. M. C. (2010). REM-dependent repair of competitive memory suppression. *Experimental Brain Research, 203,* 471–477.

Barnier, A. J., Hung, L., & Conway, M. A. (2004). Retrieval-induced forgetting of emotional and unemotional autobiographical memories. *Cognition & Emotion, 18,* 457–477. doi:10.1080/0269993034000392

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1–46. doi:10.1037/h0027577

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

Bäuml, K.-H. (2002). Semantic generation can cause episodic forgetting. *Psychological Science, 13,* 356–360. doi:10.1111/j.0956-7976.2002.00464.x

Bäuml, K.-H., & Hartinger, A. (2002). On the role of item similarity in retrieval-induced forgetting. *Memory, 10,* 215–224. doi:10.1080/09658210143000362

Blaxton, T. A., & Neely, J. H. (1983). Inhibition from semantically related primes: Evidence of a category-specific inhibition. *Memory & Cognition, 11,* 500–510.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21,* 1–16.

Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1314–1319. doi:10.1037/0278-7393.27.5.1314

Chan, J. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61,* 153–170. doi:10.1016/j.jml.2009.04.004

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1403–1414. doi:10.1037/0278-7393.25.6.1403

Conroy, R., & Salmon, K. (2005). Selective postevent review and children's memory for nonreviewed material. *Journal of Experimental Child Psychology, 90,* 185–207. doi:10.1016/j.jecp.2004.11.004

Conroy, R., & Salmon, K. (2006). Talking about parts of a past experience: The impact of discussion style and event structure on memory for discussed and nondiscussed information. *Journal of Experimental Child Psychology, 95,* 278–297. doi:10.1016/j.jecp.2006.06.001

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

Ford, R. M., Keating, S., & Patel, R. (2004). Retrieval-induced forgetting: A developmental study. *British Journal of Developmental Psychology, 22,* 585–603. doi:10.1348/0261510042378272

Garcia-Bajos, E., Migueles, M., & Anderson, M. C. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory, 17,* 92–103. doi:10.1080/09658210802572454

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.

Harris, C. B., Sharman, S. J., Barnier, A. J., & Moulds, M. L. (2010). Mood and retrieval-induced forgetting of positive and negative autobiographical memories. *Applied Cognitive Psychology, 24,* 399–413. doi:10.1002/acp.1685

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms. *International Journal of Chronobiology, 4,* 97–110.

Horton, D. L., & Kjeldergaard, P. M. (1961). An experimental analysis of associative factors in mediated generalization. *Psychological Monographs: General and Applied, 75,* 1–26.

Jenkins, J. (1963). Mediated associations: Paradigms and situations. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes* (pp. 210–257). New York, NY: McGraw-Hill. doi:10.1037/11178-006

Johansson, M., Aslan, A., Bäuml, K.-H., Gabel, A., & Mecklinger, A. (2007). When remembering causes forgetting: Electrophysiological correlates of retrieval-induced forgetting. *Cerebral Cortex, 17,* 1335–1341. doi:10.1093/cercor/bhl044

Johnson, S. K., & Anderson, M. C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science, 15,* 448–453. doi:10.1111/j.0956-7976.2004.00700.x

Kanungo, R. (1967). Meaning mediation in verbal transfer. *British Journal of Psychology, 58,* 205–212.

Kjeldergaard, P. M. (1968). Transfer and mediation in verbal learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 1071–1075). Englewood Cliffs, NJ: Prentice Hall.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Kuhl, B. A., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nature Neuroscience, 10,* 908–914. doi:10.1038/nn1918

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second language acquisition. *Psychological Science, 18,* 29–34.

MacLeod, M. (2002). Retrieval-induced forgetting in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology, 16,* 135–149. doi:10.1002/acp.782

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science, 12,* 148–152. doi:10.1111/1467-9280.00325

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When

practice makes imperfect. *Journal of Personality and Social Psychology, 77,* 463–473. doi:10.1037/0022-3514.77.3.463

McEvoy, C. L., & Nelson, D. L. (1982). Category name and instance norms for 106 categories of various sizes. *American Journal of Psychology, 95,* 581–634. doi:10.2307/1422189

Migueles, M., & Garcia-Bajos, E. (2006). Influence of the typicality of the actions in a mugging script on retrieval-induced forgetting. *Psicológica International Journal of Methodology and Experimental Psychology, 27,* 119–135.

Migueles, M., & Garcia-Bajos, E. (2007). Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology, 21,* 1157–1172. doi:10.1002/acp.1323

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36,* 408–420.

Osgood, C. E. (1946). Meaningful similarity and interference in learning. *Journal of Experimental Psychology, 36,* 277–301. doi:10.1037/h0063154

Phenix, T. L., & Campbell, J. I. D. (2004). Effects of multiplication practice on product verification: Integrated structures model or retrieval-induced forgetting? *Memory & Cognition, 32,* 324–335.

Postman, L. (1971). Transfer, interference, and forgetting. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth and Schlosberg's experimental psychology* (3rd ed., pp. 1019–1132). New York, NY: Holt, Rinehart & Winston.

Postman, L., & Stark, K. (1964). Proactive inhibition as a function of the conditions of transfer. *Journal of Verbal Learning and Verbal Behavior, 3,* 249–259. doi:10.1016/S0022-5371(64)80050-5

Radvansky, G. A., Spieler, D. H., & Zacks, R. T. (1993). Mental model organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 95–114. doi:10.1037/0278-7393.19.1.95

Radvansky, G. A., Wyer, R. S., Curiel, J. M., & Lutz, M. F. (1997). Situation models and abstract ownership relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1233–1246. doi:10.1037/0278-7393.23.5.1233

Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 940–953. doi:10.1037/0278-7393.17.5.940

Runquist, W. N., & Marshall, M. A. (1963). Transfer, synonymity, and anticipatory interval in paired-associate verbal learning. *American Journal of Psychology, 76,* 281–286. doi:10.2307/1419165

Sahakyan, L., & Goodmon, L. B. (2007). The influence of directional associations on directed forgetting and interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 1035–1049. doi:10.1037/0278-7393.33.6.1035

Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of a memory: The role of retrieval-induced forgetting effects in misinformation effects. *Journal of Experimental Psychology: Applied, 8,* 127–142. doi:10.1037/1076-898X.8.2.127

Shapiro, S. I. (1970). Facilitation and interference effects as a function of the free associative strength of mediators. *Journal of Experimental Psychology, 84,* 69–73. doi:10.1037/h0028944

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forget-

ting in an eyewitness memory paradigm. *Psychonomic Bulletin & Review, 2,* 249–253.

Shivde, G., & Anderson, M. C. (2001). The role of inhibition in meaning selection: Insights from retrieval-induced forgetting. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 175–190). Washington, DC: American Psychological Association. doi:10.1037/10459-010

Solso, R. L., & Juel, C. L. (1980). Positional frequency and versatility of bigrams for two-through nine-letter English words. *Behavior Research Methods & Instrumentation, 12,* 297–343.

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

Storm, B. C. (2010). Retrieval-induced forgetting and the resolution of competition. In A. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork.* Washington, DC: American Psychological Association.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2007). When intended remembering leads to unintended forgetting. *Quarterly Journal of Experimental Psychology, 60,* 909–915. doi:10.1080/17470210701288706

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2008). Accelerated relearning after retrieval-induced forgetting: The benefit of being forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 230–236. doi:10.1037/0278-7393.34.1.230

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

Tandoh, K., & Naka, M. (2007). Durability of retrieval-induced forgetting. *Shinrigaku Kenkyu, 78,* 310–315.

Underwood, B. J. (1951). Associative transfer in verbal learning as a function of response similarity and degree of first-list learning. *Journal of Experimental Psychology, 42,* 44–53. doi:10.1037/h0059484

Van Overshelde, 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.

Veling, H., & van Knippenberg, A. (2004). Remembering can cause inhibition: Retrieval-induced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 315–318. doi:10.1037/0278-7393.30.2.315

Yoon, C., Feinberg, F., Hu, P., Gutchess, A. H., Hedden, T., Chen, H., . . . Park, D. C. (2004). Category norms as a function of culture and age: Comparisons of item responses to 105 categories by American and Chinese adults. *Psychology and Aging, 19,* 379–393. doi:10.1037/0882-7974.19.3.379

Young, R. K. (1955). Retroactive and proactive effects under varying conditions of response similarity. *Journal of Experimental Psychology, 50,* 113–119. doi:10.1037/h0047382

Zachary, R. A. (1991). *The manual of the Shipley Institute of Living Scale.* Los Angeles, CA: Western Psychological Services.