MEMORY, 2002, *10* (3), 215–224

On the role of item similarity in retrieval-induced forgetting

Karl-Heinz Bäuml & Armin Hartinger

Universität Regensburg, Germany

We report on two experiments designed to examine how the similarity of retrieval-practised and not-retrieval-practised items influences the amount of retrieval-induced forgetting. Participants studied categorised item lists with each category consisting of exemplars from two different semantic subcategories. Using both the retrieval practice paradigm (Experiment 1) and the output interference paradigm (Experiment 2) we found that the retrieval of a subset of the studied items impaired the subsequent recall of the nonpractised items if the two types of items were fairly dissimilar to each other (same category but different subcategory) but did not induce impairment if the two types of items were highly similar (same category and subcategory). These results indicate that a high degree of similarity of practised and nonpractised items can eliminate retrieval-induced forgetting. They also suggest that forgetting in the retrieval practice paradigm and forgetting in the output interference paradigm are mediated by the same mechanisms. The relation of the present results to other very recent findings about the role of item similarity in retrieval-induced forgetting in discussed.

The recall of learned material can be impaired through preceding retrieval of related, already stored material. Evidence for such retrieval-induced forgetting was first observed in studies on output interference. These studies examined how the recall of learned items varies as a function of the items' serial position in the testing sequence. The general result of these studies was that an item's recall probability declines with its testing position (Roediger & Schmidt, 1980; A. Smith, 1971; Tulving & Arbuckle, 1966), thus indicating that the act of recall itself can induce forgetting. More recently, Anderson, Bjork, and Bjork (1994) made a similar point when using their retrieval practice paradigm. They demonstrated that repeated retrieval practice of some of the learned items can enhance these items' later recall, but that, at the same time, this repeated retrieval can cause forgetting of the nonpractised items.

The fact that the retrieval of learned material can cause forgetting of related material is interesting for a number of reasons. For instance, there is quite converging evidence from several recent studies that, whereas repeated retrieval of material can cause forgetting of related material, repeated presentation of material does not (Anderson, Bjork, & Bjork, 2000a; Bäuml, 1996b, 1997; Ciranni & Shimamura, 1999; DaPolito, 1966). This finding suggests that retrieval-induced forgetting is a recall-specific effect and is not due to the strengthening of the practised items (see Bäuml, in press, for a related point). Furthermore, there is a surprising role of item strength in retrieval-induced forgetting. Results by Anderson et al. (1994), using the retrieval practice paradigm, and by Bäuml (1998), using the output interference paradigm, indicate that items that are strongly associated to the common cue are subject to retrieval-induced forgetting, whereas items that

Requests for reprints should be sent to Karl-Heinz Bäuml, Institut für Psychologie, Universität Regensburg, 93040 Regensburg, Germany. Email: karl-heinz.baeuml@psychologie.uni-regensburg.de

Experiment 1 was presented at the November 1999 meeting of the Psychonomic Society ini Los Angeles (Bäuml & Hartinger, 1999). This research was supported by the Deutsche Forschungsgemeinschaft (Ba-1382/2).

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210143000362

are weakly associated to the common cue do not show forgetting and may even benefit from the prior recall of related material. This finding was interpreted as evidence for the action of a retrieval suppression mechanism.

Whatever the mechanisms are that mediate retrieval-induced forgetting, in order to specify these mechanisms in some detail we need to know more about which factors influence this type of forgetting, and which factors do not. A particularly interesting factor here is item similarity. Although item similarity is known to have substantial effects on, for instance, retroactive interference (Bäuml, 1996a; Shuell, 1968; Young, 1955), to date the role of item similarity in retrieval-induced forgetting has largely been ignored. Until recently, there was just one study in the literature in which the issue was addressed. Using an output interference paradigm, Roediger and Schmidt (1980) presented participants with categorisable item lists, which consisted of either semantically unrelated or semantically related categories. At test, participants attempted to recall the words within categories when given the category names as retrieval cues. As expected, recall of words declined with the output position of the category. This decline, however, was largely the same whether the categories in the list were related or unrelated.

Very recently, Smith and Hunt (2000) published a second study on the role of item similarity in retrieval-induced forgetting. Unlike Roediger and Schmidt (1980), they addressed the issue by using the retrieval practice paradigm rather than the output interference paradigm. As another difference, they varied item similarity by varying the degree of relational or distinctive processing of the items, as induced by similarity or difference judgements among the studied items. They presented participants with items from different semantic categories and, in a subsequent phase of the experiment, asked participants to generate either similarities or differences among all the items of a category. Relative to a standard encoding task, they found no effect of relational processing on the amount of retrieval-induced forgetting. In the case of distinctive processing, however, they found a reduction, and even elimination, of the forgetting. On the basis of these results Smith and Hunt argued that similarity is a pre-condition for retrieval-induced forgetting to occur. According to this view similarity is associated with competition. Distinctive processing of the items supposedly reduces the competition between the items, and thus also reduces retrieval-induced forgetting.

Both Roediger and Schmidt (1980) and Smith and Hunt (2000) varied item similarity by varing the similarity among all the items of a category, or list, i.e., the similarity of practised and nonpractised items and, simultaneously, the similarity between the practised and nonpractised items themselves. Anderson, Green, and McCulloch (2000b) pointed out that, in principle, these two types of similarity manipulations might have different and even opposing effects on the amount of retrieval-induced forgetting, and examined the two types of similarity manipulations separately. They presented participants with items from different semantic categories and, in a subsequent phase of the experiment, asked participants to generate either similarities or unique properties among pairs of items from the same category. These pairs were composed of either a practised and a nonpractised item—addressing what they call *target–competitor* similarity—or two nonpractised items—addressing what they call *competitor–competitor* similarity. With respect to competitor–competitor similarity, Anderson and colleagues found retrieval-induced forgetting only in the case of a relational processing of the items, but did not find forgetting in the case of the processing of unique properties. With respect to target–competitor similarity, Anderson and colleagues found retrieval-induced forgetting only in the case of the processing of unique properties of items, whereas there was no forgetting, and even a recall improvement, in the case of a relational processing. These results suggest that the similarity of practised and nonpractised items and the similarity between the practised and nonpractised items themselves can affect retrieval-induced forgetting in a different and even opposing way.

The suggestion that there are two opposing effects of item similarity is theoretically interesting because it is in agreement with predictions from Anderson and Spellman's (1995) feature suppression model of retrieval-induced forgetting. According to this model, items are represented as sets of features. When an item is retrieved, the suppression mechanism enhances the retrieval of the target item by inhibiting all the features of competing items that these competing items do not share with the target item. If the probability of recalling a competing item is related to the summed activity of all its features—including those activated due to their

overlap with the target item and those inhibited in the nonoverlapping set—then less impairment should result for competing items that are very similar to the target item than for competing items that are less similar to the target item. On the other hand, if competitors themselves show a major overlap in features, suppressing a feature in one of these items can do ''double duty'' by impairing the competitor sharing the same feature as well. As a result, more impairment should result for competing items that have similar competitors than for competing items that have less similar competitors (Anderson et al., 2000b). The result that a high degree of target–competitor similarity and a low degree of competitor–competitor similarity eliminates retrieval-induced forgetting is consistent with this proposal.

The feature suppression model provides a promising account of retrieval-induced forgetting effects. However, before accepting it as a general model of the effect of item similarity in this type of forgetting, a number of things have yet to be shown. Besides issues that have to do with the question of how exactly effects of target–competitor and competitor–competitor similarity combine (see Discussion), more basic issues have to be addressed as well. Among these issues, of course, is the question of whether the effects reported by Anderson et al. (2000b) generalise to other types of similarity variations as well, particularly to those typically used in previous interference studies (Bäuml, 1996a; Roediger & Schmidt, 1980; Shuell, 1968; Sowder, 1973). More important, it needs to be demonstrated that the same results that hold true in the retrieval practice paradigm do also hold in the output interference paradigm (Roediger & Schmidt, 1980). Indeed, the null effect of item similarity found in the Roediger and Schmidt study may be explained by assuming that the same mechanisms mediate retrieval-induced forgetting in the retrieval practice and the output interference paradigm, and that the null effect was just the result of a perfect cancellation of the two opposing effects of item similarity. However, at this stage this explanation is fairly speculative and may be premature. To bridge the gap between Roediger and Schmidt's (1980) results and those obtained by Anderson et al. (2000b), it needs to be demonstrated that the effects of target–competitor and competitor–competitor similarity found by Anderson et al. hold in the output interference paradigm as well. Only if this generalisation is true, can feature suppression serve as a general account of retrieval-induced forgetting.

The present study addresses exactly these two issues. Because, in the Anderson et al. (2000b) study, the effect of variations in target–competitor similarity was much larger than the effect of variations in competitor–competitor similarity, we focused on target–competitor similarity. We then examined whether the results found by Anderson et al. generalise to more classical variations of item similarity (Experiment 1), and, more importantly, whether they generalise to the more traditional output interference paradigm (Experiment 2). Experiment 1 uses the retrieval practice paradigm. A three-phase experiment is reported. In phase 1 of the experiment participants studied an item list which consisted of different semantic categories. Each category consisted of four exemplars, two from one semantic subcategory and two from another. In phase 2, participants repeatedly retrieved one exemplar from each of the categories, before in the final phase 3 they attempted to recall two nonpractised items from each category—the other item from the same subcategory and one randomly chosen item from the other subcategory—and, subsequently, recalled the category's retrieval-practised item. We expected to find that the retrieval practice of an item improves this item's later recall and, at the same time, induces retrieval-induced forgetting for the category's nonpractised items. The main interest was in the question of whether this retrieval-induced forgetting was different for items that are closely related to the practised item (same category and same subcategory) and items that are less closely related to it (same category but different subcategory).

EXPERIMENT 1

Method

Participants. A total of 54 psychology students at the University of Regensburg participated in Experiment 1. Their average age was 21.2 years (min 19, max 26). They were tested individually.

Materials. Two item lists were constructed, each consisting of eight experimental and three filler categories. The experimental categories

contained four exemplars each, the filler categories contained two exemplars each. The items were drawn from several published norms (Battig & Montague, 1969; Mannhaupt, 1983) and had a rank order between 1 and 27 according to these norms (mean 9.03, median 7, SD 9.43). The four items of an experimental category belonged to two subcategories with two items each. The category *tree*, for instance, contained two exemplars belonging to the subcategory *deciduous tree* and two exemplars belonging to the subcategory *conifer*. The category *four-legged animal* contained two exemplars belonging to the subcategory *predator* and two exemplars belonging to the subcategory *hoofed animal*. A preceding inquiry by questionnaire (44 students) made sure that within each category, exemplars of the same subcategory were in fact considered more similar to each other than exemplars of different subcategories.[1] Moreover, each category's exemplars were chosen to be either strongly or moderately associated to their category cue (rank orders < 30). Exemplars with a weak association to their category cue (rank orders > 30) were omitted. This restriction is important because it guarantees a rough control of possible item strength effects. Indeed, in previous studies strong and moderate items were found to show retrieval-induced forgetting and to show about the same amount of forgetting (Bäuml, 1998), whereas weak items were found to show no retrieval-induced forgetting at all (Anderson et al., 1994; Bäuml, 1998). In addition, for each item list effort was made to minimise intercategory similarity and association, as well as to minimise phonemic similarities among the category labels. All items within a category had different first letters. Thus, the uniqueness of a cue consisting of the category label plus one or more initial letters of the target item was guaranteed.

Design and procedure. There were two experimental conditions (*practice condition* and *control condition*), each consisting of a presentation phase, a retention phase, and a test phase. Presentation and test phases were identical in the two conditions. The retention phases, however, differed: in the practice condition the participants repeatedly retrieved one item from each of the previously presented categories, whereas in the control condition no items were practised and an unrelated distractor task was carried out instead. The categories in the practice condition are called *Rp* (or retrieval practice) *categories*, whereas the categories in the control condition are called *Nrp* (no retrieval practice) *categories*. Those items in the practice condition that were repeatedly retrieved and thus were part of Rp categories are referred to as *Rp+ items* in the following, and the remaining items, which were not practised but were also part of Rp categories, are called *Rp− items*. One of the Rp− items always belonged to the same subcategory as the Rp+ item (*$Rp-_{sim}$ item*), whereas the remaining two Rp− items belonged to a different subcategory (*$Rp-_{dissim}$ items*). Items in the control condition received no retrieval practice. As they belonged to Nrp categories they are referred to as *Nrp items* in the following.

Each participant participated in both the control and the practice condition. One half of the participants started with the practice, the other half with the control condition. There was a break of 5 to 10 minutes between a participant's first and second condition. Each list was used equally often in each experimental condition and in each temporal position. Across all participants, each of a category's four items served equally often as a Rp+ item, a $Rp-_{sim}$ item, and a $Rp-_{dissim}$ item.

Presentation phase. The participants were presented one item list consisting of 32 (8 × 4) experimental items and 6 (3 × 2) filler items. Each item together with its category and subcategory label was displayed on a computer screen for 6 s (e.g., *ANIMAL PREDATOR tiger*). The item was presented within a frame underneath the subcategory label, which in turn was presented underneath the category label. Participants were instructed to spend the whole exposure time relating the exemplar to its category and subcategory label and to rehearse only this very item in order to maximise recall performance. The

[1] We presented item triples from each category to the students, with each triple consisting of one target and two nontarget items. One of the nontarget items was from the same subcategory as the target item, while the other was from the other subcategory. The students had to indicate which nontarget item shared more features with the target item, and on a 7-point scale they rated the feature similarity between the target item and each of the two nontargets (1 = "*no common features*", 7 = "*all features identical*"). In 91.8% of all cases, exemplars of the same subcategory were judged more similar to each other than exemplars of different subcategories. The mean similarity ratings for exemplars from the same and different subcategories were 4.9 and 2.3, respectively, $F(1, 43) = 930.1$, $MS_e = 0.153$, $p < .001$. To save room, the items are not listed in this paper. However, we will be happy to send English translations of the (originally German) item lists employed to anybody who is interested.

experimental items were presented in blocked random order. That is, a random sequence of four blocks with eight items each was presented to the participants. Each block consisted of one randomly selected exemplar from each of the eight categories. The order of the categories within a block was also random, with the only restriction that a block's last item never belonged to the same category as the next block's first item. At the beginning and end of each list, one item from each of the list's three filler categories was presented.

Retention phase. After 60 s of counting backwards by threes from a random three-digit number, a retrieval practice phase, in which one item (Rp+) from each experimental category had to be retrieved, followed in the practice condition. The retrieval of the eight Rp+ items was controlled by presenting the category label plus two of the items' (unique) initial letters as retrieval cues. The presentation of two instead of only one initial letter (see Test Phase) should enhance the successful retrieval of the to-be-practised items. The Rp+ items had to be retrieved in three consecutive blocks, with a 10 s break between the blocks. Each Rp+ item had to be retrieved once within each block. The order of the items within a block was random. Each retrieval cue was presented for 10 s on the computer screen. During this time, the participants wrote down the target item on one page of a retrieval practice booklet consisting of 24 (3 × 8) blank pages. After these 10 s the participants received a signal to turn the page and to proceed with the retrieval practice for the next Rp+ item. In the control condition, the 60 s distractor task was followed by yet another distractor task. This task took just as long as the repeated item retrieval in the practice condition, and required the participants to estimate the age of nonfamiliar human faces (black-and-white portraits) presented on the computer screen. Subsequent to the retrieval practice procedure (practice condition) and the age estimation procedure (control condition), an additional 3 min distractor task followed. During this task, participants had to rate the attractiveness of nonfamiliar human faces which were presented on the computer screen.

Test phase. After the 3 min distractor task participants were presented with the category label plus the (unique) first letter of items learned in the first phase of the experiment. The order in which the items were tested was blocked by category. From each category only one of the two $Rp-_{dissim}$ items was tested. It was selected randomly among the two $Rp-_{dissim}$ items. Thus, for each category one Rp+, one $Rp-_{sim}$, and one $Rp-_{dissim}$ item had to be recalled. For one half of the categories, the $Rp-_{sim}$ item was tested before the $Rp-_{dissim}$ item, for the other half the $Rp-_{sim}$ item was tested after the $Rp-_{dissim}$ item. The Rp+ item was always tested in a category's last (third) testing position. The mean category testing position was the same for all categories, in both experimental conditions. Each retrieval cue was displayed on the computer screen for 10 s. During this time, the participants wrote down the target item on one page of a test booklet consisting of 24 (8 × 3) blank pages. After this 10 s interval, a signal indicated the presentation of the next retrieval cue.

Results

Success of retrieval practice. We verified first whether the repeated retrieval of the Rp+ items during the retrieval practice phase was successful. An average recall rate of 97.8% indicated that in the overwhelming majority of cases retrieval of the Rp+ items was successful.

Effects of retrieval practice. Cued recall of the Nrp items in the control condition was compared to cued recall of the Rp+ items in the practice condition (see Table 1). On average, 89.5% of the control items and 96.4% of the Rp+ items were recalled. The difference of +6.9% between the two conditions was reliable, $F(1, 53) = 9.24$, $MS_e = 0.013$, $p < .01$, thus demonstrating the expected positive effect of retrieval practice on the recall of the Rp+ items. Then cued recall of the Rp− items ($Rp-_{sim}$ and $Rp-_{dissim}$ items) in the practice condition was compared to cued recall of the Nrp items in the control condition. On average, 89.5% of the Nrp items but only 84.9% of the Rp− items were recalled. The difference of 4.6% was significant, $F(1, 53) = 6.92$, $MS_e = 0.008$, $p < .02$, thus demonstrating the expected pattern of retrieval-induced forgetting.

Effects of item similarity. To verify whether the degree of similarity of a Rp− item to its practised Rp+ item influenced the amount of retrieval-induced forgetting, cued recall of the $Rp-_{sim}$ items was compared to cued recall of the $Rp-_{dissim}$ items. On average, 87.7% of the $Rp-_{sim}$ and 82.1% of the $Rp-_{dissim}$ items were recalled, the difference of 5.6% being statistically

TABLE 1
Experiment 1

	Nrp items	Rp−$_{sim}$ items	Rp−$_{dissim}$ items	Rp+ items
Performance	89.5%	87.7%	82.1%	96.4%
	(1.5%)	(2.0%)	(2.6%)	(1.3%)
Forgetting	—	−1.8%	−7.4%	+6.9%

Recall performance and amount of forgetting on a category-plus-first-letter cued recall test as a function of item type (*Experiment 1*): Nrp = items shown in the control condition; Rp+ = items shown in the practice condition, which received retrieval practice; Rp−$_{sim}$ = items shown in the practice condition, which were not practised and similar to the Rp+ items; Rp−$_{dissim}$ = items shown in the practice condition, which were not practised and dissimilar to the Rp+ items. Standard errors are shown in parentheses.

significant, $F(1,53) = 4.35$, $MS_e = 0.019$, $p < .05$. This result indicates that dissimilar items were more impaired through the retrieval of related items than similar items. This indication is also supported by analyses in which the amount of retrieval-induced forgetting for Rp−$_{dissim}$ and Rp−$_{sim}$ items was considered separately. Rp−$_{sim}$ items were recalled only little worse than control items (87.7% vs 89.5%), the difference being not significant, $F(1,53) < 1$. The difference between the Rp−$_{dissim}$ and the control items, however, was much more substantial (82.1% vs 89.5%) and statistically reliable, $F(1,53) = 8.76$, $MS_e = 0.017$, $p < .01$.

Discussion

The retrieval practice of the Rp+ items improved their later recall but impaired the recall of the nonpractised Rp− items, which replicates previous findings by Anderson and colleagues (Anderson et al., 1994; Anderson & Spellman, 1995). The primary question in this experiment was whether the impairment in the recall of Rp− items, which is caused through retrieval practice of the Rp+ items, varied with the similarity between the two types of items. We found the Rp− items to show less retrieval impairment if the items were similar to the practised Rp+ items (Rp−$_{sim}$ items) than if they were dissimilar (Rp−$_{dissim}$ items). Thus, consistent with a very recent result by Anderson et al. (2000b), we found the similarity of practised and nonpractised items to influence the amount of retrieval-induced forgetting.

Experiment 1 is a first demonstration that item similarity can influence retrieval-induced forgetting. We wanted to replicate this effect of item similarity in another experiment, this time, however, using the more traditional output interference paradigm. A two-phase experiment was conducted. In phase 1 of the experiment participants studied an item list which consisted of different semantic categories. Each category contained eight exemplars, four from one semantic subcategory and four from another. In phase 2 participants attempted to recall four items from each category, two target and two nontarget items. The two target items always belonged to the same subcategory, and the same was true for the two nontarget items. In the control condition a category's two target items were tested before the category's two nontarget items; in the experimental conditions testing order was reversed. In one of the two experimental conditions the nontarget items were from the same subcategory as the target items, in the other the two types of items were from a category's two different subcategories. On the basis of the results from previous output interference studies we expected to find that the prior recall of a category's nontarget items impairs the subsequent recall of the category's target items. On the basis of the results from Experiment 1, however, we expected to find this effect mainly if the target items and the previously recalled nontarget items were from the two different subcategories of a category, and less, if at all, if they belonged to the same subcategory.

EXPERIMENT 2

Method

Participants. A total of 48 psychology students at the University of Regensburg participated in Experiment 2. Their average age was 21.4 years (min 19, max 28). They were tested individually.

Materials. One item list was constructed which consisted of six experimental and three filler categories. The six experimental categories were drawn from the eight experimental categories used in Experiment 1 and contained eight exemplars each. The eight items belonged to two subcategories with four items each. Two of the four items of a subcategory were identical to those employed in Experiment 1, the other two items were new. These added items were drawn from the same published norms as the items used in Experiment 1 (Battig & Montague, 1969; Mannhaupt, 1983). Just like the previously used items, the added items were strongly or moderately associated to their category cue (rank orders < 30). All items within a category had different first letters. Thus, the uniqueness of a cue consisting of the category label plus the initial letter of the target item was guaranteed. The filler categories contained two items each, which were identical to those used in Experiment 1.

Design and procedure. Presentation phase. The participants were presented with one item list consisting of 48 (6 × 8) experimental items and 6 (3 × 2) filler items. The procedure was analogous to the one already used in Experiment 1. After 60 s of counting backwards and another 3 min distractor task in which the participants rated the attractiveness of nonfamiliar human faces, a cued recall test was carried out.

Test phase. For each category, two items from one subcategory were defined as the *target items* for a participant, the other two items from the same subcategory as the *similar items*, and two items from the category's other subcategory as the *dissimilar items*. There were three different conditions, one control and two experimental conditions. In the *control condition* the two target items were tested first in their categories followed by the test of the two dissimilar items; in the *similar condition* the two similar items were tested first and the two target items were tested second; in the *dissimilar condition* the two dissimilar items were tested first and the two target items were tested second. For each participant two of the six experimental categories served as the control condition, two other categories as the similar condition, and the remaining two categories as the dissimilar condition. Categories were tested in two successive blocks, each block consisting of one category from each of the three conditions. Across participants mean category testing position was the same for all categories. Each category served equally often as the control, the similar, and the dissimilar condition. Each item of a category served equally often as a target item, a similar item, and a dissimilar item. The testing procedure was analogous to the one used in Experiment 1.

Results

Effects of testing position. The effect of testing position on recall performance was analysed. Cued recall of the target items when they were tested first in their category, i.e., in positions 1 and 2 of a category block, was compared to cued recall when they were tested second, i.e., in positions 3 and 4 of a category block. On average, 83.3% of the target items were recalled if they were tested first and 77.1% if they were tested second (see Table 2). This difference of 6.2% was reliable, $F(1,47) = 4.48$, $MS_e = 0.021$, $p < .05$, thus demonstrating the well-known output interference effect.

Effects of item similarity. As Table 2 shows, testing position had a different effect on the target items depending on whether similar or dissimilar items were tested first. The prior testing of dissimilar items, i.e., items that were from the same category but a different subcategory as the target items, induced a considerable output interference effect: Mean recall of the target items was 73.4% compared to 83.3% in the control condition, the difference of 9.9% being reliable, $F(1,47) = 6.37$, $MS_e = 0.037, p < .02$. The prior testing of the similar items, i.e., items that were from the same category and same subcategory as the target items, induced only a small output interference effect: Mean recall of the target items was 80.7% compared to 83.3% in the control condition, the difference of 2.6% not being reliable, $F(1,47) < 1$. Recall in the dissimilar

TABLE 2
Experiment 2

	Tf items	Ts_{sim} items	Ts_{dissim} items
Performance	83.3% (2.6%)	80.7% (2.8%)	73.4% (3.4%)
Forgetting	—	−2.6%	−9.9%

Recall performance and amount of forgetting on a category-plus-first-letter cued recall test as a function of item type (*Experiment 2*): Tf = target items when tested first; Ts_{sim} = target items when tested second after the recall of similar items; Ts_{dissim} = target items when tested second after the recall of dissimilar items. Standard errors are shown in parentheses.

condition thus was lower than recall in the similar condition, the difference of 7.3% was statistically significant, 73.4% vs 80.7%; $F(1, 47) = 4.58$, $MS_e = 0.028$, $p < .05$. Mean recall of the similar and dissimilar items in testing positions 1 and 2 of a category was about the same, 83.8% vs 81.8%, $F(1, 47) < 1$. The effect of item similarity on recall performance of the target items, therefore, was not caused by the fact that more dissimilar than similar items were recalled previously.

Discussion

The prior recall of a category's nontarget items impaired the subsequent recall of the category's target items, which replicates the typical finding of output interference reported in a number of previous studies (Anderson et al., 1994; Bäuml, 1997, 1998). The primary question in this experiment was whether the amount of output interference, which is caused through the prior recall of the nontarget items, varies with the similarity of the target and nontarget items. We found the target items to show reliable output interference only if the nontarget items were dissimilar to the target items—i.e., if they were from the same category but a different subcategory as the target items—but did not find reliable output interference if the nontarget items were similar to the target items—i.e., if they were from the same category and same subcategory as the target items. This result replicates the basic finding of Experiment 1 in the context of output interference.

GENERAL DISCUSSION

Experiments 1 and 2 show that item similarity can influence retrieval-induced forgetting. Using the retrieval practice paradigm the results from Experiment 1 demonstrate that the repeated retrieval of previously learned material can cause forgetting of items that are relatively dissimilar to the practised items, whereas it does not cause forgetting of items that are highly similar to the practised items. The results of Experiment 2 reveal an analogous pattern when using the output interference paradigm. Whereas the recall of items during test can cause forgetting of items that are relatively dissimilar to the previously recalled items, there is hardly any forgetting of items that are very similar to the previously recalled items. These results indicate that a high degree of similarity of practised and nonpractised items can reduce, or even eliminate, retrieval-induced forgetting.

The present findings generalise results very recently reported by Anderson et al. (2000b). Using the retrieval practice paradigm, Anderson and colleagues varied the similarity of practised and nonpractised items by varying the degree of relational or distinctive processing of the items as induced by similarity or difference judgements among the studied items. They found retrieval-induced forgetting in the case of a distinctive processing of the items, but found no forgetting, and even a recall improvement, in the case of a relational processing. The results from the present study are in agreement with these findings. They thus indicate that the effect of similarity does not depend on how exactly item similarity is varied, whether it is varied through the presentation of common labels which point out the similarities between the items, or through the instruction to generate similarities between items. More important, the present results address the issue of whether the forgetting observed in the retrieval practice paradigm and the forgetting observed in the more traditional output interference paradigm are mediated by the same mechanisms. Together with previous results by Bäuml (1998), who found the same effect of item strength in the output interference paradigm as had previously been shown in the retrieval practice paradigm (Anderson et al., 1994), the present findings show that there is indeed a great deal of concordance between which factors affect forgetting in the two paradigms. On the basis of this concordance it is suggested that similar mechanisms mediate retrieval-induced forgetting in the two paradigms.

In their output interference study Roediger and Schmidt (1980) varied the similarity of practised and nonpractised items and, simultaneously, the similarity between the practised and nonpractised items themselves. Using this procedure they did not find any effect of item similarity on amount of forgetting. The results from the present study indicate that Roediger and Schmidt's result reflects the effect of two opposing effects of item similarity. Because, as the present data show, an increase in the similarity of practised and nonpractised items reduces the amount of output interference, Roediger and Schmidt's results suggest that there was a second, opposing effect of item similarity in their experiment. This opposing effect should have cancelled the forgetting-reducing effect, to result in an effective null effect of item similarity. The results from the present study

together with those from the three previous studies (Anderson et al., 2000b; Roediger & Schmidt, 1980; Smith & Hunt, 2000) thus converge on the view that a common mechanism with two opposing effects of item similarity underlies forgetting in the two paradigms.

As opposed to previous theoretical accounts of retrieval-induced forgetting, which were largely based on the assumption of strength-dependent competition (Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1980, 1981), in several recent studies retrieval-induced forgetting was proposed to be the result of a retrieval suppression mechanism (Anderson et al., 1994; Bäuml, 1998). On the basis of such a proposal, the role of the similarity of practised and nonpractised items on retrieval-induced forgetting can be explained by using Anderson and Spellman's (1995) feature suppression account. The crucial assumption is that, internally, items are represented as sets of features. When an item is retrieved, the suppression mechanism enhances the retrieval of the target item (*lion*) by inhibiting all the features of competing items (*Tiger*, *horse*) that these competing items do not share with the target item. If the recall probability of a competing item is related to the summed activity of all of its features—including those activated due to their overlap with the target item and those inhibited in the nonoverlapping set—then less recall impairment should result for competing items that are very similar to the target item (*tiger*) than for competing items that are less similar to the target item (*horse*), a prediction that is in agreement with our results. Pattern suppression thus provides a reasonable explanation of the present data.

An alternative explanation of the present results might be based on a two-stage recall proposal according to which the successful recall of an item depends on the access to both its category and subcategory label (e.g., Rundus, 1973). Using such a proposal one could argue that practising an item of a category does two things: first, it inhibits all other items that share the same category label independently of whether they are more similar (same subcategory) or less similar (different subcategory) to the practised item; second, it heightens the accessibility of the practised item's subcategory label at a later test. This heightening of the subcategory label's accessibility should improve recall of the high-similarity items at test and thus induce less forgetting for the high- than the low-similarity items, which is consistent with the present data. This account, which basically assumes similarity-independent inhibition, presumably is inappropriate. Competition is generally regarded as a precondition for inhibition (Anderson et al., 1994). The forgoing account thus rests on the assumption that despite their important role as retrieval cues, the subcategory labels do not restrict competition between items. Two-stage recall models typically include such a restriction, assuming that mainly items sharing the same category *and* subcategory cue show competition, whereas items sharing the same category but different subcategory labels show much less competition, if at all (Raaijmakers & Shiffrin, 1980; Rundus, 1973). Including this restriction in the forgoing account changes predictions. High-similarity items should now suffer much more inhibition than low-similarity items, thus largely cancelling the positive effect of an improved accessibility of the subcategory label. Following this line of reasoning, two-stage models therefore should predict roughly the same amount of forgetting for high- and low-similarity items, a prediction not in agreement with the present results.

As outlined in the introduction, not only can feature suppression account for the effect of the similarity of practised and nonpractised items. It can also account for the opposing effect of the similarity between the nonpractised items themselves (Anderson et al., 2000b). A question not yet solved in this account is how exactly the effects of the two types of similarity variations combine in the case of a simultaneous variation to result in a certain amount of forgetting. Indeed, Smith and Hunt (2000) varied both types of similarities simultaneously and found forgetting in the case of a relational processing of items, but no forgetting in the case of a distinctive processing. This pattern of results mimics the one that Anderson et al. (2000b) observed when varying the competitor–competitor similarity. This agreement might indicate that in the Smith and Hunt experiment the effect of variations in competitor–competitor similarity was much stronger than the effect of variations in target–competitor similarity, or, alternatively, that the two effects were about the same but combined in a highly nonadditive manner. Understanding how the two types of similarity variations combine is interesting both practically and theoretically. The answer to this question will establish an important empirical restriction on computational models of retrieval-induced forgetting.

Manuscript received 4 April 2001
Manuscript accepted 10 October 2001

REFERENCES

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

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

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

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

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.

Bäuml, K.-H. (1996a). A Markov model for measuring storage loss and retrieval failure in retroactive inhibition. *Acta Psychologica*, *92*, 231–250.

Bäuml, K.-H. (1996b). Revisiting an old issue: Retroactive interference as a function of the degree of original and interpolated learning. *Psychonomic Bulletin & Review*, *3*, 380–384.

Bäuml, K.-H. (1997). The list-strength effect: Strength-dependent competition or suppression? *Psychonomic Bulletin & Review*, *4*, 260–264.

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. (in press). Semantic generation can cause episodic forgetting. *Psychological Science*.

Bäuml, K.-H., & Hartinger, A. (1999). Retrieval-induced forgetting: The role of item similarity. *Abstracts of the Psychonomic Society, 40th Annual Meeting*, 385.

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

DaPolito, F.J. (1966). *Proactive effects with independent retrieval of competing responses*. Unpublished doctoral dissertation, Indiana University.

Mannhaupt, H.-R. (1983). Produktionsnormen für verbale Reaktionen zu 40 geläufigen Kategorien. *Sprache & Kognition*, *2*, 264–278.

Mensink, J.-G., & Raaijmakers, J.G.W. (1988). A model of interference and forgetting. *Psychological Review*, *95*, 434–455.

Raaijmakers, J.G.W., & Shiffrin, R.M. (1980). SAM: A theory of probabilistic search of associative memory. In G. Bower (Ed.), *The psychology of learning and motivation, Vol. 14*. New York: Academic Press.

Raaijmakers, J.G.W., & Shiffrin, R.M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Roediger, H.L.III, & Schmidt, S.R. (1980). Output interference in the recall of categorized and paired associate lists. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 91–105.

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior*, *12*, 43–50.

Shuell, T.J. (1968). Retroactive inhibition in free-recall learning of categorized lists. *Journal of Verbal Learning and Verbal Behavior*, *7*, 797–805.

Smith, A.D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*, 400–408.

Smith, R.E., & Hunt, R.R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition*, *28*, 503–508.

Sowder, C.D. (1973). Mediated unlearning. *Journal of Experimental Psychology*, *100*, 50–57.

Tulving, E., & Arbuckle, T.Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, *72*, 89–104.

Young, R.K. (1955). Retroactive and proactive effects under varying conditions of response similarity. *Journal of Experimental Psychology*, *50*, 113–117.