

Strong items get suppressed, weak items do not: The role of item strength in output interference

KARL-HEINZ BÄUML

Universität Regensburg, Regensburg, Germany

An experiment is reported that examines the role of item strength in output interference. Subjects studied two types of categorized item lists: lists in which each category consisted of strong and moderate items, and lists in which each category consisted of weak and moderate items. Different degrees of item strength were accomplished by varying the items' taxonomic frequency within a category. The subjects either recalled a category's strong and weak items before its moderate items, or vice versa. The prior recall of the moderate items impaired the later recall of the strong items, but did not impair the later recall of the weak items. This effect of item strength indicates that output interference is caused by a process of retrieval suppression. It additionally suggests that, in order to minimize output-interference effects in recall, a list's strong items should be recalled before its weak items.

The recall of learned material can impair later recall of other learned material. Evidence for such retrieval-induced forgetting has arisen from a number of experimental paradigms. In studies on output interference, for instance, it has been found that an item's recall probability declines as a function of its serial position in a testing sequence (Roediger & Schmidt, 1980; Smith, 1971). Studies on part-list cuing have shown that the recall of items is impaired when a subset of the learned items is presented as a retrieval cue during test (Roediger, 1973; Rundus, 1973). More recently, Anderson and colleagues (Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995) have demonstrated that the repeated retrieval of some of the learned items can cause long-lasting forgetting of the nonpracticed items.

Retrieval-induced forgetting has often been attributed to strength-dependent competition. According to this principle, memories associated to a common cue compete for access to conscious recall when that cue is presented. In particular, the cued recall of an item is supposed to decrease as a function of the increase in strength of its competitors' associations to the cue (Raaijmakers & Shiffrin, 1981; Rundus, 1973). Strength dependence predicts that if the act of retrieval is a learning event in the sense that it enhances subsequent recall of the retrieved item—an assumption typically made in current memory models—then the prior recall of items should reduce recall of related items. This predicted pattern of results has been found in many previous studies.

The amount of output interference an item suffers is often assumed to vary with its associative strength to the retrieval cue: Weak items are proposed to be particularly vulnerable to the effects of a prior recall of related mate-

rial, whereas strong items are proposed to be quite robust against this type of interference. This assumption about how output interference depends on item strength is implied by current memory models (Raaijmakers & Shiffrin, 1981; Rundus, 1973; for a formal demonstration, see Anderson et al., 1994, Appendix A), but it has been proposed in other contexts as well. Brainerd, Reyna, Howe, and Kevershan (1990), for instance, reported evidence that in free recall of item lists, weak items are recalled before strong items much of the time. They argued that this output order reflects a basic retrieval mechanism designed to maximize efficiency of free recall by giving priority to the more vulnerable weak items (for similar arguments, see also Brainerd, Reyna, Harnishfeger, & Howe, 1993). However, whether weak items are really more susceptible to output interference than strong items has never been examined.

As plausible as the assumption of greater interference effects for weak items than for strong items may appear, it is far from clear whether this proposal is empirically sound. In fact, using their retrieval-practice paradigm, Anderson et al. (1994) found that the repeated recall of items impaired only the later recall of items strongly associated to a common cue, whereas it did not impair the recall of items weakly associated to the cue. Moreover, there was even a tendency that the cued recall of weak items was enhanced by this prior recall. Although Anderson et al.'s experiment addressed long-lasting effects of retrieval practice—and thus, in principle, might just reflect a substantially higher rate of spontaneous recovery for the weak items than for the strong ones (Wheeler, 1995)—it casts some doubt on the view that output-interference effects are greater for weak than for strong items. Their finding also challenges the strength-dependence principle as a general explanation of retrieval-induced forgetting because strength dependence predicts that recall of both strong and weak items will be impaired by the prior recall of related items.

I thank M. Anderson, H. L. Roediger III, J. Wixted, and one anonymous referee for their comments on an earlier draft of the manuscript. Correspondence should be addressed to K.-H. Bäuml, Institut für Psychologie, Universität Regensburg, 93040 Regensburg, Germany (e-mail: karl-heinz.bauml@psychologie.uni-regensburg.de).

Finding an answer to the question of how output-interference effects vary with an item's associative strength to the retrieval cue is important for at least two reasons. First, such an answer would have implications regarding the question of what mechanisms mediate output interference. If it is the case that mainly the weak items are subject to output interference, then the interference would be consistent with strength-dependence competition and, in this sense, would support the current memory models. However, the finding that only the strong items are subject to output interference but the weak items are not would challenge the strength-dependence explanation and would point to alternative explanations, such as the assumption of a retrieval-suppression mechanism (Anderson et al., 1994).

Second, such an answer would have implications regarding which output order is best to recall learned material most efficiently. If mainly weak items show interference effects, as argued by Brainerd et al. (1990) and predicted by many memory models, then output could be maximized by recalling the weak items before the strong ones. However, if mainly strong items are subject to interference effects, the reversed order would be more efficient. Evidence for such a reversed order in free recall has arisen from the results of a recent study by Wixted, Ghadisha, and Vera (1997). They found that subjects recalled a list's strong items faster than and prior to its weak items.

Here, an experiment is reported in which output-interference effects were measured for both items strongly associated to a common category cue and items weakly associated to that cue. Subjects studied two types of categorized item lists: those in which each category consisted of strong and moderate items and those in which each category consisted of weak and moderate items. Different degrees of cue-item strength were accomplished by varying the items' taxonomic frequency within a category. The subjects either recalled a category's strong and weak items before its moderate items, or vice versa. The results from this experiment provide evidence on whether and how an item's associative strength to the retrieval cue influences its susceptibility to output interference.

METHOD

Subjects

Eighty psychology students at the University of Regensburg participated in the experiment. They were tested individually.

Material

Two types of item lists were constructed, each consisting of five experimental categories and three filler categories. In the first type of item list, the strong-moderate (SM) list, each of the experimental categories consisted of three strong and three moderate items. In the other type of item list, the weak-moderate (WM) list, each of the experimental categories consisted of three weak and three moderate items. The filler categories consisted of two moderate items each. The items were drawn from several published norms (Battig & Montague, 1969; Mannheim, 1983; Scheith & Bäuml, 1995). The strong items were chosen to have a rank order between 5 and 10 according to these norms ($M = 7.3$), the

moderate items to have a rank order between 15 and 20 ($M = 17.2$), and the weak items to have a rank order between 30 and 40 ($M = 34.8$). Half of the subjects received two SM lists for study, and the other half received two WM lists. Two different pairs of SM lists and two different pairs of WM lists were used throughout the experiment.

For each item list, effort was made to minimize intercategory similarity and association, as well as to minimize phonemic similarities among the category labels. No two exemplars from the same category began with the same first letter, ensuring that each letter cue would be unique at test. In addition, to avoid interference from extraexperimental items, no chosen category exemplar had the same first letter as an unchosen category exemplar that was listed in the norms. Items with strong a priori item-to-item associations were avoided.

Procedure

The two lists that a subject received for study were presented within one experimental session. Between the recall of the first and the presentation of the second list, subjects participated in an unrelated experiment of roughly 10-min duration in which they judged the perceived attractiveness of face stimuli. The items were presented using learning booklets. Each of the 36 pages of a learning booklet contained one item with its category label (e.g., VEGETABLE-tomato). The order of exemplars within a booklet was determined by blocked randomization in which each block contained one exemplar from each category, resulting in six blocks of five items. The ordering of exemplars within each block was determined randomly except that throughout the booklet no two items from the same category appeared in sequence. Four different learning booklets were constructed for each list, in which both the ordering of categories within blocks and the list position of particular category items were varied. One exemplar from each of the three filler categories was presented at the beginning of a list and the other exemplar at the end. Each exemplar, together with its category label, was presented for 5 sec, after which time the subjects received a signal to turn the page. Subjects were instructed to spend the whole exposure time relating the exemplar to its category and to rehearse only the pair presented in order to maximize recall performance.

The presentation of each complete list was promptly followed by 30 sec of backward subtraction by sevens from a random three-digit number. Immediately following this task, a cued recall test was carried out. Subjects received a test booklet. Each of the 10 pages of this booklet contained one category cue together with the first letter of three exemplars below it. On each page, the first letter of all three moderate items or the first letter of all three strong (weak) items of an experimental category were presented. The six items of a category were tested on two successive pages. For one of the two SM lists (WM lists), the subjects were tested on a category's strong (weak) items before they were tested on its moderate items; for the other list, they were tested on its moderate items before they were tested on its strong (weak) items. The subjects were instructed that their task was to retrieve the exemplars from any portion of the experiment, that corresponded to the cues. There was no instruction about guessing, and thus guessing was neither encouraged nor discouraged. Subjects were given 20 sec to recall a page's three items, after which time a signal instructed them to turn the page. Four different test booklets were constructed for each item list, in which both the ordering of categories and the ordering of the first letter cues were varied.

RESULTS

Item-Strength Effects

As expected from the items' rank orders in the category norms, high-frequency members of categories (strong items) showed higher recall performance than low-frequency members (weak items). I compared mean recall of the two types of items when they were tested in the first three positions of a category block. In both cases, the other items associated to the category cue were

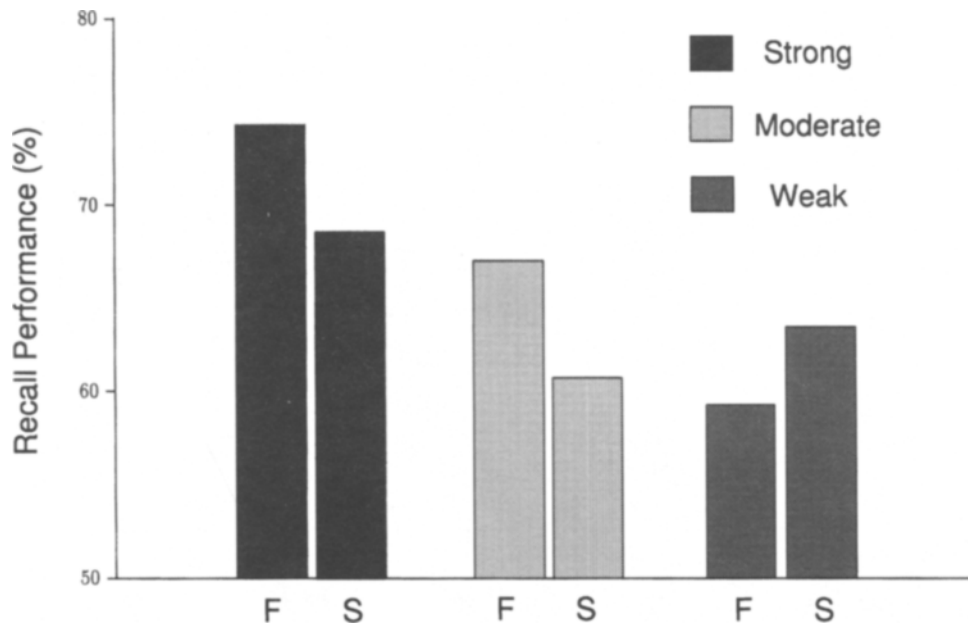


Figure 1: Recall performance on a category-plus-first-letter cued recall test as a function of item strength (strong, moderate, weak) and within-category testing position. Items were tested either in the first three positions (F) or in the second three positions (S) of a category block.

medium-frequency members of categories (moderate items). Mean recall of strong items was 74.4%; that of weak items was 59.3%. This strong-weak difference of 15.1% was statistically reliable [$F(1,78) = 16.0$, $MS_e = .028$, $p < .001$].

Recall performance of moderate items when tested in the first three positions of a category block was examined as well. Mean recall of moderate items was 68.5% when the other category exemplars were strong items and 65.7% when they were weak items. This difference of 2.8% was not reliable [$F(1,78) = 2.1$, $MS_e = .022$, $p = .15$], indicating that the strength of the other items associated to the common cue did not affect recall. As expected from the items' rank orders, recall performance of the moderate items was just between that of the strong and the weak items.

Testing-Position Effects

The effect of an item's testing position within a category on recall performance was analyzed. Performance when items were tested in the first three positions of a category block (tested-first items) was compared for all three types of items with performance when items were tested in the second three positions (tested-second items). On the basis of previous studies on output interference, recall was expected to be lower for tested-second items than for tested-first items.

As Figure 1 shows, testing position had a different effect on the strong and the weak items. Mean recall of strong items was 74.4% when tested first and 68.8% when tested second. This testing-position effect of 5.6% was reliable [$F(1,39) = 4.5$, $MS_e = .014$, $p < .05$], indi-

cating that the prior recall of moderate items impaired the recall of the subsequently tested strong items. Mean recall of weak items was 59.3% when tested first and 62.5% when tested second. The slight increase in recall performance of 3.2% was not reliable [$F(1,39) = 1.1$, $MS_e = .018$, $p = .30$], indicating that the prior recall of the moderate items did not influence the recall of the subsequently tested weak items.

Due to the testing-position effects, the strong-weak difference declined from 15.1% when the items were tested in the first three positions of a category block to 6.3% when they were tested in the second three positions. A two-factor analysis of variance (ANOVA) confirmed that testing position had a different influence on strong and weak items: There was a highly significant main effect of item strength [$F(1,78) = 12.2$, $MS_e = .038$, $p < .001$], no significant main effect of testing position [$F(1,78) < 1$], and a significant interaction between item strength and testing position [$F(1,78) = 4.8$, $MS_e = .016$, $p < .05$].

The moderate items showed testing-position effects as well. Both the moderate items that were learned in the presence of strong items and the moderate items that were learned in the presence of weak items showed impairment in mean recall when tested second (7.3% and 5.5%). In both cases, this impairment was reliable [$F(1,39) = 11.2$, $MS_e = .010$, $p < .01$; $F(1,39) = 5.9$, $MS_e = .010$, $p < .05$]. A two-factor ANOVA revealed that the testing-position effect for the moderate items was roughly independent of whether a category's strong or weak items were recalled before the moderate items: There was a highly significant main effect of testing position [$F(1,78) = 16.6$, $MS_e =$

.010, $p < .001$], no significant main effect of category composition [$F(1,78) = 1.4$, $MS_e = .043$, $p = .24$], and no significant interaction between testing position and category composition [$F(1,78) < 1$].¹

DISCUSSION

Previous studies on output interference have shown that the recall of items during test impairs recall of other items later in the test (Roediger & Schmidt, 1980; Smith, 1971). On the one hand, the present results confirm this finding from previous studies, showing that high-frequency members of categories are impaired by the prior recall of items associated to the same category cue. On the other hand, however, the results demonstrate that such impairment does not always occur. In contrast to high-frequency members, categories' low-frequency members are not impaired by the prior recall of related items and even show a tendency for recall facilitation. These results indicate that the amount of output interference an item may suffer is a function of its associative strength to the retrieval cue, with larger impairment in the case of strong cue-item associations and smaller impairment, if any, in the case of weak cue-item associations.

The output-interference effects observed in this study were not the result of strength-dependent competition. This conclusion arises from the fact that, according to strength dependence, both strong and weak items should suffer from retrieval-induced impairment. Indeed, since the recall of items is assumed to increase these items' associations to the common cue, the likelihood of recall of the still-to-be-remembered items should be relatively decreased. The experimental results, however, indicate that only items with strong associations to the common cue are impaired, whereas items weakly associated to the cue are not impaired. Strength dependence, therefore, cannot explain the influence of prior recall on the subsequent recall of related material.

The observed pattern of output-interference effects is consistent with the assumption of a retrieval-suppression mechanism. This view rests largely on two propositions. The first proposition is that retrieving an item inhibits other items associated to the same cue; the more that items associated to a cue are retrieved, the more impaired those related items will be. Second, the inhibition does not depend on the strength of the retrieved item and, thus, is the same whether a strong or a weak item is successfully recalled. The inhibition, however, depends on the strength of the yet-to-be-remembered item, with a larger impairment if the item is strongly associated to the common cue and a smaller impairment, if any, if the item is weakly associated to the cue. This second proposition rests on the theoretical idea that, because of their greater associative strength, strong items interfere more with the retrieval of their competitors than do weak items and thus are more severely inhibited (Anderson et al., 1994; Roediger & Neely, 1982; see also Postman, Stark, & Fraser, 1968). The present data indeed show larger impairment for the strong items than for the weak items. Using a different experimental paradigm, Anderson et al. recently reported a similar pattern of results. To first approximation, both their results and the present ones are well described by retrieval suppression.

One might argue that the difference in amount of output interference found between the strong and weak items could be due to differences in study conditions or guessing at test. In the present study, subjects learned lists of strong and moderate items and lists of weak and moderate items. This difference between lists could have led to interactions between the strong and moderate items, on the one hand, and the weak and moderate items, on the other. Recall of the moderate items when tested in the first three positions of a category block, however, was the same whether the other learned category exemplars were strong or weak items, which indicates that there were no substantial item interactions during study (see Bäuml, 1997, for similar results). The difference in study conditions, therefore, should not have caused the difference in output interference. The pattern of output-interference effects should also not be due to guessing at test. During test it was possible that sometimes subjects filled in the correct item even if they did not recall the item from the list; that is, retrieval was from semantic rather than episodic memory. However, such guessing is expected to occur mainly for the high-

frequency members of categories and much less for the categories' medium and low-frequency members. Thus, to the extent that it really occurred, guessing should have mainly attenuated the output interference of the high-frequency members.

The present data indicate that the free recall of item lists should be most efficient if the strong items were recalled before the weak items. Indeed, since mainly the strong items suffer from output interference, these items should be output first. The weak items show less susceptibility to interference effects and, therefore, can also be output later in the testing sequence (thus also reducing the chance that their recall leads to suppression of the strong items). This conclusion arises because, according to the suppression hypothesis, the amount of output interference an item suffers depends only on its own cue-item strength but not on the cue-item strength of the previously recalled item (see also Anderson et al., 1994). There is currently a dispute in the literature about which output order subjects actually show when freely recalling strong and weak items from an item list. Brainerd et al. (1990) analyzed subjects' output order and found that many of the weak items were recalled before the strong ones. Wixted et al. (1997), on the other hand, analyzed subjects' output order and recall latencies and found that the strong items were recalled prior to and faster than the weak items. On the basis of the present findings, the results of Brainerd et al. (1990) suggest that our retrieval is not very efficient, whereas the results of Wixted et al. suggest that our retrieval is quite efficient. Obviously, further studies are necessary to draw more firm conclusions regarding which output order subjects actually show. The results from these studies will tell us how efficient our retrieval really is.

A final observation: Several previous studies have investigated whether additional study time or study trials for items induce impairments in the recall of related material (Bäuml, 1996, 1997; DaPolito, 1966). None of these studies yielded any evidence for such an effect. The present experiment provides a further test on this point: Moderate items were learned in the presence of low-frequency (weak) items, and they were learned in the presence of high-frequency (strong) items. Mean recall of the moderate items was largely the same in the two conditions, which is consistent with the results from the previous studies. A finding that the retrieval of learned items impaired recall of related material but the items' strengthening through additional study time or study trials did not would suggest that the act of retrieval is not just a learning event. This point addresses a very fundamental aspect of current memory models, which typically assume that learning and retrieval are largely equivalent. Examining the validity of this proposed equivalence is a high priority for future research.

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, & Cognition*, *20*, 1063-1087.
- ANDERSON, M. C., & SPELLMAN, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review*, *102*, 68-100.
- 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. (1996). 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.
- BRAINERD, C. J., REYNA, V. F., HARNISHFEGGER, K. K., & HOWE, M. L. (1993). Is retrievability grouping good for recall? *Journal of Experimental Psychology: General*, *122*, 249-268.
- BRAINERD, C. J., REYNA, V. F., HOWE, M. L., & KEVERSHAN, J. (1990). The last shall be first: How memory strength affects children's retrieval. *Psychological Science*, *1*, 247-252.
- 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 [German category norms for verbal items in 40 categories]. *Sprache & Kognition*, *4*, 264-278.
- POSTMAN, L., STARK, K., & FRASER, J. (1968). Temporal changes in interference. *Journal of Verbal Learning & Verbal Behavior*, *7*, 672-694.
- RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93-134.
- ROEDIGER, H. L., III (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning & Verbal Behavior*, *12*, 644-657.
- ROEDIGER, H. L., III, & NEELY, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology*, *36*, 213-242.
- 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 & Memory*, *6*, 91-105.
- RUNDUS, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior*, *12*, 43-50.
- SCHEITHE, K., & BÄUML, K.-H. (1995). Deutschsprachige Normen für Vertreter von 48 Kategorien [German normative data for representatives of 48 categories]. *Sprache & Kognition*, *14*, 39-43.
- SMITH, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning & Verbal Behavior*, *10*, 400-408.
- WHEELER, M. A. (1995). Improvement in recall over time without repeated testing: Spontaneous recovery revisited. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 173-184.
- WIXTED, J. T., GHADISHA, H., & VERA, R. (1997). Recall latency following pure- and mixed-strength lists: A direct test of the relative strength model of free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 523-538.

NOTE

The effects of testing position for the strong and the weak items, on the one hand, and the moderate items, on the other, were analyzed separately because the moderate items were not preceded by the same items at test as the strong and weak items. The amount of output interference that the moderate items suffered, therefore, is not directly comparable to the amount of interference found for the strong and weak items.

(Manuscript received September 22, 1997;
revision accepted for publication February 2, 1998.)