Psychonomic Bulletin & Review
1997, 4 (2), 260-264

The list-strength effect: Strength-dependent competition or suppression?

KARL-HEINZ BÄUML
University of Regensburg, Regensburg, Germany

If several items are associated with a common cue, the cued recall of an item is often supposed to decrease as a function of the increase in strength of its competitors' associations with the cue. Evidence for such a list-strength effect has been found in prior research, but this effect could have been caused both by the strength manipulations and by retrieval-based suppression, because the strengthening and the output order of the items were confounded. The experiment reported here employed categorizable item lists; some categories in each list contained strong items only, some contained weak items only, and some contained both strong and weak items. Strengthening was accomplished by varying the exposure time of the items. The testing sequence of the items from each category was controlled by the use of category-plus-first-letter cues. When the typical confounding of strengthening and output order was mimicked, list-strength effects were found, which is consistent with prior research. However, when this confounding was eliminated, the list-strength effects disappeared: The recall of neither strong nor weak items varied with the strengths of the other category exemplars. This pattern of results indicates that the list-strength effect is not the result of strength-dependent competition, but is caused by output-order biases and a process of suppression.

The cued recall of an item decreases as the number of items associated with the same cue increases. Corresponding evidence has been provided by a number of studies in quite different experimental paradigms (see Watkins, 1978, for a review). This finding is generally seen as support for the assumption that memories associated with a common cue compete for access to conscious recall when that cue is presented.

Often, this competition is assumed to be strength dependent; that is, the cued recall of an item is supposed to decrease as a function of the increase in strength of its competitors' associations with the cue. Evidence for this strength dependence arises particularly from studies on retroactive and proactive interference (Barnes & Underwood, 1959; Briggs, 1957), although the results from studies on part–set cuing have been interpreted similarly (Roediger, 1974; Rundus, 1973). This evidence for strength dependence appears to be so compelling that the concept has been incorporated as a fundamental property into many memory models (Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1981; Rundus, 1973).

The concept of strength dependence, however, is far less well empirically established than it would seem. Recently, Anderson, Bjork, and Bjork (1994) emphasized this point by arguing that previous interference studies, which seemingly provide evidence for strength dependence, confounded the manipulation of the strengths of the competitors' associations with the cue with the competitors' retrieval practice. On the basis of this argument, not only the strength manipulations but also the lack of control over the competitors' retrieval practice might have caused the poorer performance in a situation with high-strength competitors (Anderson & Spellman, 1995; Roediger, 1978). Indeed, two studies, which did not confound the strength of associations with retrieval practice, confirmed this point. Using an A–B, A–C proactive interference design, DaPolito (1966) did not find any evidence that strengthening the A–B associations influenced the later cued recall of C items, which were learned after the A–B associations. Similarly, using a list-learning retroactive interference paradigm, Bäuml (1996) did not find any evidence that higher interpolated learning induced larger amounts of forgetting of the original list than did lower interpolated learning.

The most recent illustration of the apparent relationship between strengthening and impairment comes from a series of studies on what Ratcliff, Clark, and Shiffrin (1990) have termed the *list-strength effect*. Using a so-called mixed–pure paradigm, Ratcliff et al. presented subjects with three kinds of item lists: pure strong lists containing strong items only, pure weak lists containing weak items only, and mixed lists containing half weak and half strong items. Strengthening was accomplished by increasing either the exposure time or the number of repetitions of the to-be-strengthened items. In free-recall tests, Ratcliff et al. found that (1) relatively more strong items were recalled from the mixed lists than from the pure strong lists, and (2) relatively more weak items were recalled from the pure weak lists than from the mixed lists. This finding was interpreted as evidence for strength-

Thanks are extended to M. Anderson, A. Glenberg, H. L. Roediger III, and one anonymous referee for their comments on an earlier draft of this manuscript. The author's address is Institut für Psychologie, Universität Regensburg, 93040 Regensburg, Germany (e-mail: heinz@ rpss3.psychologie.uni-regensburg.de).

dependent competition, since the likelihood of recall of the target items seemed to increase (decrease) with decreases (increases) in the competitors' strengths.

The results of Ratcliff et al. (1990), however, can also be explained by retrieval-based suppression. This view rests on three assumptions. First, retrieving an item suppresses other items that are associated with the same cue. Second, the more items associated with a cue that are retrieved, the more impaired those related items will be. Third, the impairment does not depend on the strength of the retrieved item but only on its successful recall (Anderson et al., 1994; Roediger & Neely, 1982; see also Postman, Stark, & Fraser, 1968). If the strong items from a mixed list are retrieved earlier than the weak items (Anderson et al., 1994), suppression predicts that recall performance of strong items should be higher from mixed lists than from pure strong lists; in fact, the retrieval of items in the first testing positions should inhibit retrieval of the still-to-be-remembered items, yielding lower performance on average in pure strong lists than in the mixed lists. Similarly, suppression predicts that recall performance of weak items should be lower from mixed lists than from pure weak lists, because, on average, there is less retrieval inhibition for the weak items in the pure weak lists than in the mixed lists. Thus, the list-strength effect, as found by Ratcliff et al., need not necessarily have been the result of strength-dependent competition but might also have been caused by output-order biases and a process of suppression.

This study reexamines the strength-dependence explanation of the list-strength effect. An experiment similar to that reported by Ratcliff et al. (1990) was conducted; this time, however, possible output-order effects at test were controlled. In detail, a mixed-category paradigm was used in which categorizable item lists were presented to subjects. One third of the categories contained strong items only, another third contained weak items only, and the last third contained both strong and weak items. Strengthening was accomplished by varying the exposure time of the items. The testing sequence within a category was controlled by cuing subjects with the category name and the (unique) first letter of the category exemplar. The first step was to examine whether similar list-strength effects, such as those found by Ratcliff et al., would also appear in the present paradigm. Such effects are expected when output-order biases that mimic the hypothesized conditions of free recall are introduced—that is, when, in mixed categories, the strong items are tested first and the weak items are tested last. The second step was to examine whether these list-strength effects would still be present when the output order was controlled—that is, when comparisons were restricted to only those items tested first within their categories. If strength-dependent competition underlies the list-strength effect, the effect should remain present whether the output order is controlled or not. However, if the list-strength effect is due to suppression, it should disappear as soon as the output-order biases at test are eliminated.

METHOD

Subjects

Forty-eight psychology students at the University of Regensburg participated in the experiment. They were tested individually.

Material

Twenty-two categories, four of which were used as fillers, were drawn from several published norms (Battig & Montague, 1969; Mannhaupt, 1983; Scheithe & Bäuml, 1995). Two lists, with 11 categories per list, were constructed. An effort was made to minimize intercategory similarity and association, as well as to minimize phonemic similarities among the category labels.

Six exemplars were chosen from each of the 22 categories. No items were selected from among the top three items within each category; the average rank of the chosen items was about 14, according to the respective category norms. 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.

Each list included four types of categories: pure strong categories, in which all items served as strong items—that is, were presented at an exposure rate of 6 sec each; pure weak categories, in which all items served as weak items—that is, were presented at an exposure rate of 2 sec each; mixed categories, in which three of the six exemplars served as strong items and the other three served as weak ones; finally, filler categories, in which all items had a medium strength—that is, were presented at an exposure rate of 4 sec each. For each list, there was an initial random assignment of the 11 categories to 3 pure strong, 3 pure weak, 3 mixed, and 2 filler categories. After one third and again after two thirds of the subjects had participated, each original pure strong category became a pure weak (mixed) category, each pure weak category became a mixed (pure strong) one, and each mixed category became a pure strong (pure weak) one. The 2 filler categories remained the same throughout the experiment.

For each of a list's nine experimental categories, there was an assignment of the six exemplars to two subsets. The first, fourth, and sixth items chosen defined the first subset; the second, third, and fifth defined the second subset. When a category served as a mixed category, the two subsets specified which items were strong and which ones were weak. Then, for half of the subjects, the items of the first subset were presented as strong and those of the second subset were presented as weak; for the other half, it was the reverse.

Procedure

The two lists were presented to each subject within one experimental session, with a break of 10 min between recall of the first list and the presentation of the second. The items were presented on a computer screen and were shown with their category label (e.g., VEGETABLE—Tomato). They were presented in blocks of strong (S) and weak (W) items, either in a 4W-9S-9W-9S-9W-9S-5W or in a 4S-9W-9S-9W-9S-9W-5S pattern. Throughout the experiment, there were six random assignments of a list's category exemplars to the single blocks.

It was important for this experiment to minimize rehearsal redistribution—that is, the possibility of the subject's borrowing some rehearsal time from the stronger items and giving it to the weaker ones. The presentation of the items in blocks of equal exposure times, as shown above, was assumed to largely reduce this possibility (see Ratcliff et al., 1990). In addition, one item from each of the two filler categories was inserted between each pair of blocks. Such insertion smooths the relatively abrupt transitions in exposure time between single blocks and thus further reduces the possibility of rehearsal sharing. Finally, 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.

After the presentation of each complete list, each subject was engaged in a 30-sec distractor task. Immediately following this task, a cued recall test was carried out. The subject was given a test booklet, the 54 (9 × 6) pages of which contained one category cue with the first letter of an exemplar next to it. The subject was instructed that his/her task was to retrieve the exemplar, from any portion of the experiment, that corresponded to those cues. The subject was given 12 sec in which to recall each item, after which a signal instructed him/her to turn the page.

Tests of exemplars were blocked by category. Strong and weak items were always tested in sequence, being either the first three or the last three items tested within their categories. The average test-booklet position of category type was controlled by creating the order of category types (S-W-M)–(W-M-S)–(M-S-W) for List 1, and the order of category types (W-M-S)–(M-S-W)–(S-W-M) for List 2, with S, W, and M representing the pure strong, pure weak, and mixed category types. Three concrete orderings were constructed for each list by randomly assigning the three respective categories of each category type to the appropriate positions. The testing order of particular exemplars within a category was counterbalanced by switching the first subset of category items with the second subset. In this way, six different testing orders were constructed for each list.

RESULTS

List-Strength Effects When Strengthening and Output Order Are Confounded

In the first step, the existence of list-strength effects was examined by comparing the recall of items from pure categories with that from mixed categories, when in mixed categories the strong items were tested first and the weak items were tested second. These comparisons included the typical output-order advantage for strong mixed-category items and the typical output-order disadvantage for weak mixed-category items, which were present in free-recall situations like those investigated by Ratcliff et al. (1990).

As expected, the pattern of list-strength effects that showed up was the same as that in Ratcliff et al.'s study: Recall performance of strong items was lower from pure categories than from mixed categories [72.3% vs. 76.2%; $F(1,47) = 4.5$, $MS_e = .008$, $p < .05$], and recall performance of weak items was numerically greater from pure categories than from mixed categories (63.0% vs. 62.1%), although this was not statistically reliable [$F(1,47) < 1$]. The strong − weak difference, therefore, was 9.3% for pure categories and 14.1% for mixed categories. Ratcliff et al., who used much longer lists, reported somewhat larger effects for both strong items (29.6% vs. 37.6%) and weak items (12.6% vs. 9.9%). Their effect for weak items, however, did also not reach significance and their effect for strong items also just passed the significance level. In this sense, the list-strength effect in the present study exactly mirrors theirs.

List-Strength Effects When the Output Order Is Controlled

Next, the existence of list-strength effects was examined when the items were tested in the first three positions of a category block (tested-first items), thus controlling for output-order biases. Figure 1A shows recall performance of tested-first items as a function of item strength and category composition. As expected, item strength affected performance: Mean recall for strong items was 75.8%, that for weak items was 64.9%. Category composition, on the other hand, showed only a small effect on performance: Of the items from pure categories, 70.0% were recalled and of the items from mixed categories, 70.7% were recalled. More importantly, however, neither for strong items nor for weak items did performance vary substantially with category composition: Mean recall for strong items was 75.4% for pure categories—that is, when the competing items were strongly associated with the common cue—and 76.2% for mixed categories—that is, when the competing items were weakly associated; mean recall for weak items was 64.6% for pure categories—that is, when the competing items were weakly associated with the common cue—and 65.3% for mixed categories—that is, when the com-

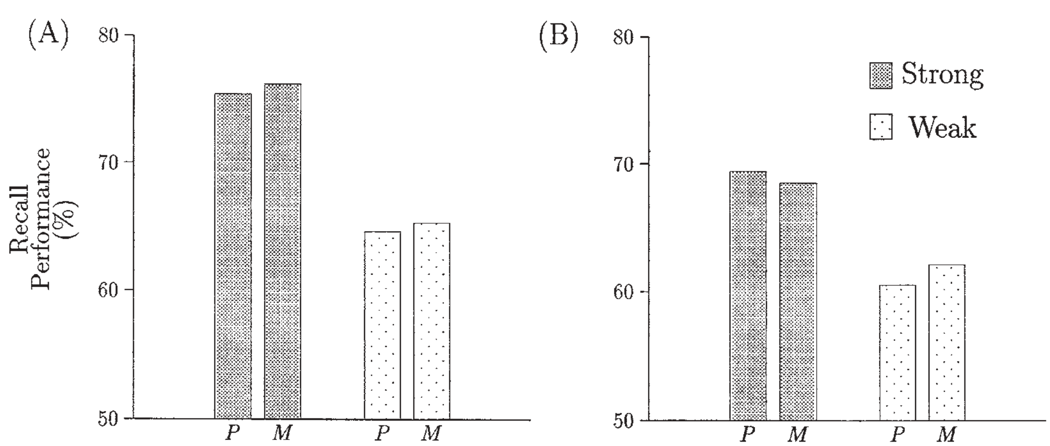

Figure 1. Recall performance on a category-plus-first-letter cued-recall test as a function of category composition and within-category testing position for items tested in the first three positions of a category block (A) and items tested in the second three positions (B). *P* = pure categories, in which all items are strong or all items are weak; *M* = mixed categories, in which half of the items are strong and half are weak.

peting items were strongly associated. Thus, the strong − weak difference was 10.8% for pure categories and 10.9% for mixed categories. These results indicate that it was only item strength and not category composition that affected recall performance. Analysis of variance confirmed this impression: While there was a highly significant main effect of item strength [$F(1,47) = 24.7$, $MS_e = .023$, $p < .001$], there was no significant main effect of category composition [$F(1,47) < 1$] and no significant interaction between item strength and category composition [$F(1,47) < 1$]. The fact that we find a list-strength effect when strengthening and output order are confounded but find no list-strength effect when the output order is controlled indicates the presence of output-order effects.

Testing-Position Effects

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

Indeed, as a comparison of Figures 1A and 1B suggests, testing position mattered. Mean recall for tested-first items was 70.2%, that for tested-second items was 65.1%. Moreover, performance decreased both for strong and for weak items: When tested first, 75.6% of the strong items and 64.8% of the weak items were recalled; when tested second, only 69.1% of the strong items and 61.1% of the weak items were recalled. Thus, testing position decreased performance of strong items by 6.5% and that of weak items by 3.7%; correspondingly, the strong − weak difference declined from 10.8% when the items were tested in the first three positions of a category block to 8.0% when tested in the second three positions. Analysis of variance confirmed that there was a main effect of testing position [$F(1,47) = 15.6$, $MS_e = .008$, $p < .001$]. The small tendency for a stronger position effect for strong items did not reach significance [$F(1,47) = 1.2$, $MS_e = .008$, $p = .27$].

DISCUSSION

Is the list-strength effect the result of strength-dependent competition or is it caused by a process of suppression? In prior research, strengthening and output order of items were confounded, and the effect, therefore, could have been caused by either one. In the present study, single-letter stem cues were used to control subjects' output order within each category, so that the possible effect of strength manipulations could be separated from the possible effect of output order.

When looking at the data in a way that corresponds to the hypothesized conditions of free recall, the pattern of list-strength effects found was the same as that reported by Ratcliff et al. (1990): More strong items and fewer weak items were recalled from mixed categories than from pure categories, which generalizes Ratcliff et al.'s result to the present paradigm. However, when restricting comparisons to only those items tested in the first halves of their categories—a situation no longer contaminated by output-order biases—the pattern of list-strength effects disappeared: Recall performance for both strong and weak items was roughly the same, whether the other items within the category were strong or weak. These results indicate that it was suppression and not strength dependence that caused the list-strength effects in the Ratcliff et al. experiment.

This suggestion is further supported by the testing-position effects found in this study: items tested first within a category showed higher recall performance than did items tested last. This held for both the strong and the weak items. Since, in the free recall of mixed-category items, strong items are recalled before weak items, these position effects predict list-strength effects for both strong and weak items (see above). Moreover, since the position effects were found to be somewhat larger for the strong than for the weak items, one may even predict larger strength effects for strong items than for weak items. It is exactly this pattern of results that was found in the mixed-list paradigm used by Ratcliff et al. and also in the present study when output-order biases were introduced at test.

Ratcliff et al.(1990) demonstrated list-strength effects not only in a free-recall test, but also in a cued-recall test. Comparable to their free-recall test, they found recall performance of strong items to be lower from pure lists than from mixed lists (36.5% vs. 43.5%), and recall performance of weak items to be higher from pure lists than from mixed lists (16.3% vs. 15.6%); the effect for strong items was reliable, the one for weak items was not. Again, these effects should have been caused by suppression. Ratcliff et al. did their cued-recall test by randomly intermixing tests of strong and weak items. On the basis of this procedure, however, the strong items from pure lists should have suffered larger amounts of suppression than those from mixed lists, because the likelihood of cued recall was lower for the weak items than for the strong ones and, hence, should have led to less inhibition in the case of mixed lists than in the case of pure lists. Analogously, the weak items from pure lists should have suffered lower amounts of suppression than those from mixed lists. Thus, both in the free-recall and the cued-recall test, the list-strength effects found by Ratcliff et al. should have been caused by output-order biases and suppression and not by strength dependence.

CONCLUSIONS

Over the years, an increasing number of studies have demonstrated that the retrieval process itself can cause forgetting (Anderson et al., 1994; Bäuml, 1996; Roediger, 1974, 1978; Smith, 1971). These retrieval-induced impairments have mostly been attributed to strength-dependent competition, assuming that the prior recall of items increases these items' associations with the common cue, thus relatively decreasing the likelihood of recall of the still-to-be-remembered items (Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1981; Rundus, 1973). The results of several recent studies, however, challenge the strength-dependence assumption, indicating that such other mechanisms as suppression may more adequately account for retrieval-induced forgetting (Anderson et al., 1994; Anderson & Spellman, 1995; Bäuml, 1996). By indicating that the list-strength effect is also not the result of strength-dependent competition but is caused by suppression, this study adds another example to this list of studies.

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.

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, **58**, 95-105.

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 Monographs*, **80** (3, Pt. 2), 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.

Briggs, G. E. (1957). Retroactive inhibition as a function of the degree of original and interpolated learning. *Journal of Experimental Psychology*, **53**, 60-67.

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*, **4**, 264-278.

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

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.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). The list-strength effect: 1. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 163-178.

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

Roediger, H. L., III (1978). Recall as a self-limiting process. *Memory & Cognition*, **6**, 54-63.

Roediger, H. L., III, & Neely, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology*, **36**, 213-242.

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. *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.

Watkins, M. J. (1978). Engrams as cuegrams and forgetting as cue-overload: A cuing approach to the structure of memory. In C. R. Puff (Ed.), *The structure of memory* (pp. 347-372). New York: Academic Press.

(Manuscript received May 6, 1996;
revision accepted for publication December 16, 1996.)