Journal of Experimental Psychology:
Learning, Memory, and Cognition
2009, Vol. 35, No. 3, 607–617

0278-7393/09/$12.00 DOI: 10.1037/a0015264

The Role of Item Strength in Retrieval-Induced Forgetting

Emőke Jakab and Jeroen G. W. Raaijmakers
University of Amsterdam

In 3 experiments, the role of item strength in the retrieval-induced forgetting paradigm was tested. According to the inhibition theory of forgetting proposed by M. C. Anderson, R. A. Bjork, and E. L. Bjork (1994), retrieval-induced forgetting should be larger for items that are more strongly associated with the category cue. In the present experiments, the authors varied item strength on the study list by manipulating the position of an item within its category (Experiments 1 and 2) and by the number of presentations in the study phase (Experiment 3). Contrary to the predictions from inhibition theory, in all 3 experiments, stronger items did not show more retrieval-induced forgetting than weaker items.

Keywords: retrieval-induced forgetting, strength, serial position, interference dependence

Retrieval-induced forgetting refers to the finding that practicing items associated with a cue impairs the recall of other items associated with the same cue (M. C. Anderson, Bjork, & Bjork, 1994). This retrieval-induced forgetting effect has been demonstrated in a large number of experiments by M. C. Anderson and others (e.g., M. C. Anderson, Bjork, & Bjork, 2000; M. C. Anderson et al., 1994; M. C. Anderson & Spellman, 1995; Bäuml, 1998; Williams & Zacks, 2001).

Retrieval-induced forgetting may be obtained with the retrieval-practice paradigm (M. C. Anderson et al., 1994). In this paradigm, participants learn a list of category–item pairs, presented one pair at a time. Each category in the list is represented by several items. After the initial study phase, half of the items in half of the categories are given additional practice with category plus stem cued recall. After a delay of (usually) 20 min, a test phase follows in which all category names from the study phase are given as cues, and all the items from the study list have to be recalled. Practiced items from the practiced categories (Rp+ items) are, of course, recalled best because of the additional practice. The nonpracticed items from the practiced categories (Rp− items) are often recalled less well compared with the items from the nonpracticed categories (Nrp items). This inferior recall of the Rp− items compared with the Nrp items is the retrieval-induced forgetting effect.

M. C. Anderson and others (e.g., M. C. Anderson, 2003; M. C. Anderson et al., 1994) have argued that the retrieval-induced forgetting effect is due to inhibition. When a category cue is presented in the retrieval-practice phase, other associated items in addition to the target item are activated and compete for recall. To overcome the competition of incorrect responses and be able to recall the target item, the inappropriate items have to be inhibited. This inhibition leads to a temporary unavailability of these items that is reflected in the impaired recall in the test phase. In summary, the decreased recall of the Rp− items is explained by assuming that these items have become inhibited during the retrieval-practice phase of the experiment.

This retrieval-induced forgetting effect can also be explained with the notion of competitive retrieval (J. R. Anderson, 1983; Mensink & Raaijmakers, 1988), in which performance is a function of the relative strength of the target association compared with the associations of other items to the same cues. As a result of the retrieval practice on the Rp+ items, the association between Rp− items and the category cue becomes relatively weaker; therefore, the performance on the Rp− items decreases.

M. C. Anderson et al. (1994) tested the explanation of such strength-based models against the inhibition explanation in a series of experiments in which the strength of the items was manipulated by varying the taxonomic frequency of words within their categories. Strong items were defined as words with a high taxonomic frequency within their category and weak items as words with a low taxonomic frequency within their category. They found more impairment for the strong items than for the weak items compared with the baseline performance on the Nrp items.

This greater impairment of stronger items is in line with the inhibition explanation: Stronger items compete more during the retrieval phase; therefore, these items have to be inhibited to a greater degree. M. C. Anderson et al. (1994) concluded that "highly accessible items are the most vulnerable to retrieval-induced forgetting" (p. 1078). This result demonstrates what M. C. Anderson (2003, 2005) termed *interference dependence*, one of the fundamental properties of the inhibition account that supposedly uniquely supports the inhibition explanation for retrieval-induced forgetting and gives evidence that alternative strength-based models may not be correct. According to this assumption of interference dependence, retrieval-induced forgetting arises only if related memories interfere during the retrieval practice of the target. If the related information does not interfere with the target, there is no need for inhibition. Therefore, interference is necessary for inhi-

Emőke Jakab and Jeroen G. W. Raaijmakers, Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands.

This research was supported by the Netherlands Organisation for Scientific Research (Grant 400-04-471). We thank Diane Pecher for helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Jeroen G. W. Raaijmakers, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, the Netherlands. E-mail: J.G.W.Raaijmakers@uva.nl

bition to occur, because inhibition is the result of the necessity to "override distracting competitors" (M. C. Anderson, 2005, p. 308).

M. C. Anderson et al. (1994) argued that strength-based competition theories could not explain these findings. According to a simple ratio-rule model in which recall probability is directly related to the relative strength of the target item compared with other items associated with the retrieval cue, weaker items should be inhibited to a proportionally greater degree contrary to what was observed. However, the status of this prediction for more complex versions of ratio-rule models is unclear (as was shown by M. C. Anderson et al.'s, 1994, Appendix A), and in particular, it may not hold for more elaborate versions of the relative strength model (e.g., Raaijmakers and Shiffrin's, 1981, search of associative memory [SAM] model) that include a number of additional processes (e.g., a recovery process based on absolute strength, and extralist assocations).[1] Analysis of a SAM-like model shows, however, that such models will usually (i.e., for reasonable parameter values) predict about equal impairment for the weak items, whereas the inhibition account always predicts more impairment for the stronger competitors.

In results similar to those of M. C. Anderson et al. (1994), Bäuml (1998) found more impairment for stronger items. He investigated the effect of item strength on output interference. Participants were presented with lists containing weak and moderate items or strong and moderate items. In the test phase, the order of testing was manipulated. When strong items were tested after moderate items, they showed impairment, whereas weak items tested after moderate items showed no impairment. Bäuml concluded that these findings were consistent with a retrieval-suppression mechanism. When moderate items are recalled first, stronger items compete more than weak items; therefore, stronger items have to be inhibited more. This inhibition, in turn, leads to impaired recall for the strong items later on. According to Bäuml, a strength-dependent competition model cannot explain these results because weakly associated items should suffer more from the output interference, which was not the case. Again, it is not clear whether this prediction does, in fact, hold for more elaborated versions of a strength-dependent competition model.

Storm, Bjork, and Bjork (2007) tested the interference dependence assumption by combining the directed-forgetting procedure with the retrieval-practice paradigm. After the study phase, participants were instructed to forget or to remember the category–item pairs they had just learned. Storm et al. found retrieval-induced forgetting in the remember condition but not in the forget condition. They concluded that when participants intend to remember the study list, more competition occurs; therefore, inhibition is necessary to reduce the competition, resulting in more retrieval-induced forgetting. On the other hand, when the study list has to be forgotten, items are less likely to interfere during retrieval practice and inhibition is not necessary.

Nevertheless, not all findings regarding the effects of item strength are in line with the interference dependence assumption of the inhibition theory. Williams and Zacks (2001) replicated the study of M. C. Anderson et al. (1994) and also found retrieval-induced forgetting for unpracticed items. However, the impairment found for weak items was similar to the impairment found for strong items. They concluded that their pattern was not consistent with the predictions of the inhibition theory and that it could be better explained by strength-dependent competition theories.

A recent experiment by Major, Camp, and MacLeod (2008) manipulated item strength by varying whether items were read or generated during the study phase. Major et al. argued that generating items should lead to stronger cue–item associations, which in turn should lead to greater retrieval-induced forgetting. Contrary to the prediction from inhibition theory, generated exemplars were not inhibited more than read exemplars.

In summary, the empirical evidence for the interference dependence assumption is rather mixed. Some experiments have manipulated strength with taxonomic frequencies and some experiments have manipulated item strength in different ways, but in both cases, there are results that support the interference dependence assumption as well as results that do not support that assumption.

Present Study

In the present study, we manipulated item strength in a way that was independent of the specific items used. In Experiments 1 and 2, we varied item strength by manipulating the position of the items within a category; in Experiment 3, we manipulated item strength by the number of presentations during the study phase. Manipulating item strength experimentally has the obvious advantage of allowing strength to be factorially varied across items.

In the study phase of the standard retrieval-practice paradigm, all items are presented in a sequence with a block design. In each block, one item of each category is presented. The placement of the items within a category is randomly determined in most experiments. The order of the presentation within a category can influence the recall of an item later on. Wood and Underwood (1967) found that items presented earlier in a category are recalled better than later items. This superior recall of early items is not due to a general serial position effect for the whole list of words but is specific to the category.

Such an effect of the position of an item within a category (or subset) was also found in one of our pilot experiments. The collapsed data based on the position within a category showed a strong primacy effect within the categories. As shown in Figure 1, the items presented first within a category were recalled better than the items presented in the middle and at the end of the category. Such a primacy effect suggests better storage for items presented as the initial items of a category (Rundus, 1971). Indeed, Wood and Underwood (1967) demonstrated in their second experiment that the locus of the improved recall of the early items was in the learning phase and not in the recall phase.

This serial position effect can be particularly important for the practiced categories, because here the items are divided into practiced (Rp+) and nonpracticed (Rp−) items. Dodd, Castel, and Roberts (2006) showed that the placement of items within a category has an effect on the occurrence of retrieval-induced

[1] An analysis of a version of the SAM model that includes extralist associates showed that such a model can give an almost perfect fit to the data of M. C. Anderson et al.'s (1994) Experiment 1, with 14% retrieval-induced forgetting for the strong categories and 8% for the weak categories, close to the observed 16% and 6%, respectively. This was the case, however, only for parameter values that were somewhat improbable in view of previous SAM simulations. For more standard values, the model predicted an absolute decrement that was more or less equal for the weak and strong categories.

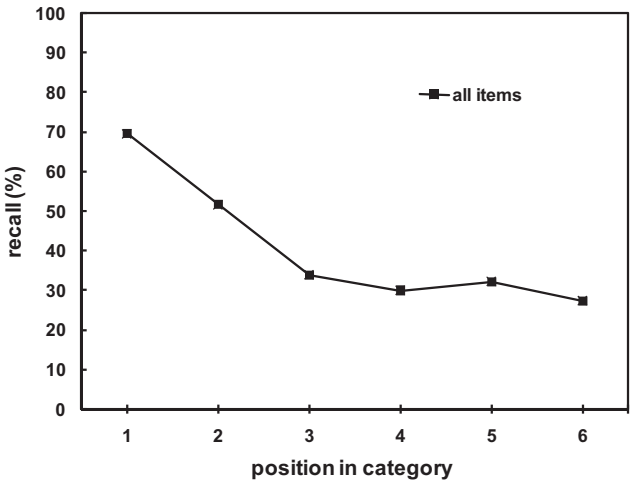

Figure 1. Mean recall percentage as a function of position in a category. Data are from an unpublished pilot experiment.

forgetting. They presented sets of words connected by a common cue and manipulated the position of practiced items within their subset in the study phase and found a retrieval-induced forgetting effect only when the practiced items were randomly chosen from a subset. When the practiced items all came from the final part of the list, no retrieval-induced forgetting effect was obtained. Because with such an arrangement the Rp− items are all from the initial part of the list, they have an advantage compared with the mean performance on the Nrp items (coming from all positions).

According to the interference dependence property of the inhibitory account, the amount of inhibition is determined by competition of the nontarget items when a cue is presented. Inhibition is necessary to reduce this competition to retrieve the target items. Early Rp− items within a category are stronger and recalled better when a cue is given and, therefore, are more likely to interfere during retrieval; as a result, they are also more prone to inhibition. Later Rp− items, however, are weaker and less likely to interfere during retrieval; hence, inhibition is not necessary. However, our pilot study was not specifically designed to test this hypothesis. We therefore designed a new experiment with a simple version of the retrieval-practice paradigm, with more adequate controls.

In the first experiment, every even or every uneven item within an Rp category was practiced; therefore, Rp− items were presented in all possible positions. In all other respects, the experiment was designed to mimic as closely as possible the procedure used by M. C. Anderson et al. (1994, Experiment 1). If, indeed, inhibition is necessary to control competing items during retrieval, the initial Rp− items should be inhibited to a greater degree than later Rp− items, leading to more impairment of the initial items in the final test phase.

Experiment 1

Method

Participants

Fifty-one students from the University of Amsterdam participated in the experiment in exchange for course credits or payment. All participants had Dutch as their mother tongue. The average age of the participants (8 male, 43 female) was 20.2 years, varying between 18 and 27 years. All had normal or corrected-to-normal vision.

Design

Retrieval-practice status was manipulated within subjects. As in previous experiments, half of the categories were practiced, and within these categories, half of the items were practiced during the retrieval-practice phase (Rp+), and the other half were not practiced (Rp−). Nrp items belonged to one of the unpracticed categories; none of the items in these categories received practice in the retrieval-practice phase. The counterbalancing of words in the study phase and within the categories resulted in eight study lists. For the test phase, two types of test lists were constructed: In half of the lists, the practiced categories were tested first; in the other half, the unpracticed categories were tested first.

Materials

Category and exemplar selection. Categories were selected from the Hudson (1982) and Storms (2001) category norms. We chose eight experimental (flowers, drinks, insects, metals, herbs, sports, birds, weapons) and two filler (fabrics, occupations) categories. All category names were unambiguous, had a length of one word, and were no more than three syllables. We chose categories with similar taxonomic frequency distributions.

Six exemplars were chosen from each category. All exemplars in the categories had medium taxonomic frequencies. Exemplars in the categories of drinks, metals, and fabrics were drawn fromHudson's (1982) category norms; the exemplars in the other seven categories were selected if their taxonomic frequencies were similar in the two sets of norms. The average exemplar taxonomic frequency was 18.3 (range = 5–42, median = 16), according to Hudson's (1982) category norms, and 16.8 (range = 5–38, median = 17), according to Storms's (2001) category norms. No two items began with the same two letters to ensure that each target in the retrieval-practice phase was uniquely specified. Items were chosen with a length between four and eight letters and between one and three syllables. The average length was 5.98 letters and 1.97 syllables.

Study lists. Study lists were constructed from 60 category–item pairs: 48 experimental and 12 filler pairs were used. Retrieval-practice status of the exemplars was taken into account in constructing the study lists. On the basis of retrieval-practice status of the exemplars, eight lists were constructed; therefore, all category–item pairs could take the Rp+, Rp−, or Nrp position. All items were presented in all possible positions (Positions 1 through 6) within the categories. We used balanced Latin squares to arrange the order of the items (Wagenaar, 1969). Similar to M. C. Anderson et al.'s (1994) study, eight category–item pair blocks were created. Each block consisted of one of the items from the different categories. Within a block, two Nrp pairs followed two Rp pairs (one Rp+ and one Rp− pair) or vice versa. Half of the lists began with two Nrp pairs and the other half with the two Rp pairs. This arrangement allowed comparisons to be made between the different conditions as a function of serial position at study. At the beginning and at the end of the list, two filler items were

placed. The rest of the filler items were used to avoid having the same two categories appear in the same order more than once.

Retrieval lists. The retrieval-practice list contained 12 category–item stem pairs from the experimental categories and 4 category–item stem pairs from the filler categories. Each exemplar was presented three times. As in the study of M. C. Anderson et al. (1994), exemplars were arranged in an expanding schedule. Between the first and the second presentation of an exemplar, 3.7 exemplars appeared, and between the second and third presentation, 6.7 exemplars were shown. No 2 category members were presented adjacently. In total, 48 category–item stem pairs were presented in the retrieval-practice phase. Counterbalancing the order of the categories, four retrieval-practice lists were constructed.

Test lists. The eight experimental categories were presented in the test phase. In half of the lists, the four practiced categories were presented first, followed by the four nonpracticed categories; in the other half of the lists, the opposite order was used. All categories appeared in all eight positions, so that the average test position of the categories was the same. Here again, Latin squares were used. Eight lists were constructed for the test phase.

Procedure

The experiment was controlled by two Pentium G3 computers. E-Prime software (Schneider, Eschman, & Zuccolotto, 2002) was used to run the experiment. Participants were tested individually or in groups of 2.

The procedure followed the retrieval-practice paradigm used by M. C. Anderson et al. (1994). The experiment consisted of four phases: the study phase, the retrieval-practice phase, the distractor, and the final test phase. Participants were told that they were participating in a memory experiment. All instructions were presented on the computer screen. In the study phase, participants were instructed to learn the category–word pairs. A plus sign was first presented for 500 ms, and then the category–item pairs were presented in the middle of the screen for 5 s, followed again by the plus sign. In the retrieval-practice phase, after the plus sign was presented for 500 ms, a category and the first two letters of an item were presented in the middle of the screen for 7 s. Participants were instructed to fill in the word stem with the items they had learned in the study phase. The retrieval-practice phase was followed by a 20-min distractor task. Two unrelated visual tasks were given as distractor tasks. In the final test phase, participants were presented with a category name in the top half of the screen, and underneath it a square text box was presented. Participants were told to type in all the words they could still remember from the category given at the top of the screen. Each category name was presented for 45 s. When the time was up, the text *Next category* appeared on the screen for 500 ms, and the following category was presented. After the experiment, participants were asked to complete an exit questionnaire.

Results and Discussion

Retrieval Practice

In the retrieval-practice phase, 84% of the exemplars were correctly completed. This rate is similar to the average rate of the strong and weak practiced items ($M = 82\%$) found by M. C. Anderson et al. (1994).

Final Memory Test

For each participant, we computed the number of Rp+, Rp−, and Nrp items that were recalled for each serial position. Recall percentages were analyzed with repeated measures analyses of variance (ANOVAs) in which retrieval-practice status was a within-subject variable and study list and test list were between-subjects variables. An alpha level of .05 was used for all statistical tests.

A significant main effect was found for retrieval-practice status, $F(2, 86) = 119.53$, $p < .001$. A planned comparison showed improved recall of the Rp+ items ($M = 71\%$) compared with the Nrp items ($M = 41\%$), $F(1, 43) = 164.57$, $p < .001$, and showed impaired recall for the Rp− items ($M = 37\%$) compared with the Nrp items ($M = 41\%$), $F(1, 43) = 4.801$, $p = .034$. These results replicate the basic retrieval-based learning and retrieval-induced forgetting effects.

Serial Position Effect in the Final Memory Test

The main question of interest in the present experiment was the effect of serial position of the items within a category. Figure 2 shows the mean recall percentage for each retrieval-practice condition as a function of the within-category serial position. A repeated measures ANOVA was carried out in which retrieval-practice status and serial position were within-subject variables, and study list and test list were between-subjects variables for the Rp− and Nrp items.

The main effect of the serial position was significant, $F(5, 215) = 68.67$, $p < .001$. The overall recall of items decreased as a function of position. The items presented at the first position were recalled best ($M = 78\%$), followed by those in the second position ($M = 52\%$) and the third position ($M = 27\%$). From the

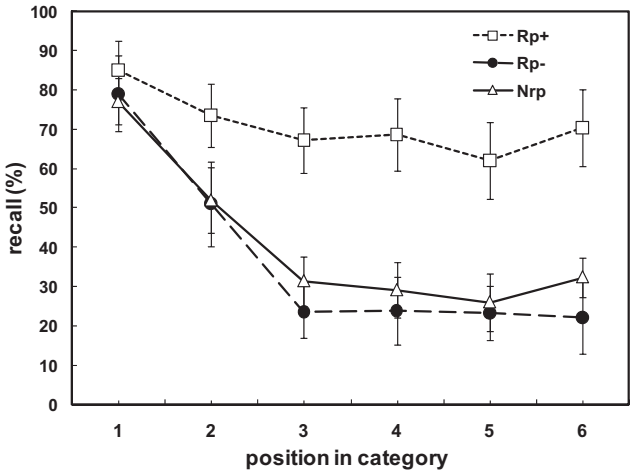

Figure 2. Mean recall percentage (with 95% confidence interval) for the retrieval-practice conditions as a function of serial position in a category in Experiment 1. Rp+ = practiced items from the practiced categories; Rp− = nonpracticed items from the practiced categories; Nrp = items from the nonpracticed categories.

third position on, the recall percentage was about the same (M = 27%, 25%, and 29%, for the fourth, fifth, and sixth positions, respectively). A planned comparison revealed a significant difference in recall between the first and the second positions, $F(1, 43) = 45.57$, $p < .001$, and between the second and the third positions, $F(1, 43) = 49.88$, $p < .001$. From the third position on, no significant differences were found.

Next, we analyzed the extent to which our within-category serial position effects were confounded by list serial position effects given that items early in the category are, of course, also early in the list as a whole. To separate list position from category position, we collapsed the study list into six blocks consisting of one item from each category. List position was defined as the position within a block. A between-subjects ANOVA with category position (six levels) and list position (eight levels) was carried out on the data from the Nrp condition. A significant main effect was found for category position, $F(5, 144) = 23.53$, $p < .001$, but not for list position, $F(7, 144) < 1$. Also, we found no evidence for an interaction effect, $F(35, 144) = 1.04$, $p > .20$. Similar results were obtained when the analysis was restricted to the first two category positions. These results demonstrate that the serial position effects are, indeed, specific to the category and do not seem to be confounded by list serial position effects.

The most important finding of this experiment was that the Retrieval-Practice Status × Serial Position interaction was not significant, $F(5, 215) < 1$. The recall of the Rp− items was similar to that of the Nrp items at all serial positions. Rp− items benefited as much from an earlier position within the category as the baseline items. Taken together, these results do not agree with the predictions of the inhibition theory, according to which there should have been more impairment for the stronger Rp− items (i.e., for the earlier serial positions). We did not find such a pattern.

It might be objected that perhaps our manipulation of strength was not relevant to the conditions leading to inhibition. That is, for inhibition to occur, the competitor item must be activated when the category cue is presented during the retrieval-practice phase of the experiment. It is, in principle, conceivable that our strength manipulation has nothing to do with the associative strength of the category cue but is based on some other dimension of trace strength. However, it can be easily shown that such is not the case by examining the output order of the Nrp items on the final test with the category name as the only cue given. The probability that the first item recalled was the item that was presented in the first serial position was 52%, and the conditional probability that an Nrp item that was initially presented in the first serial position was recalled as the first item on the final test (given that it was recalled) was 64%. Thus, the items that were considered strong were indeed the items that were very likely to be recalled first in response to the category cue and, hence, almost by definition are the ones that are most strongly associated with the category cue. If this holds for the Nrp items at the final recall, it must also hold for the Rp− items during the retrieval-practice phase. Early Rp− items should compete more during the retrieval-practice phase and, therefore, should be inhibited more. Hence, the present strength manipulation was clearly relevant for category-cued recall.

Another objection to the present findings could be that the obtained pattern might possibly be reconciled with the inhibition account by invoking the concept of integration. M. C. Anderson and McCulloch (1999) found that when participants interrelated or integrated the Rp+ and Rp− items within a category, retrieval inhibition was reduced or even eliminated. In experiments of this kind, it might be assumed that the initial items from a category are reactivated when a later item from that category is presented and that these items are then associated or integrated together (rather than associated only with the common category cue). Given that an Rp+ item was always followed by an Rp− item and vice versa, Rp− items at the beginning of a category would be rehearsed with and integrated with the Rp+ items. When the Rp+ items are subsequently practiced in the retrieval-practice phase, the integrated Rp− items might also benefit from this practice, leading to a reduction of the inhibition effect for these items. Alternatively, integration might lead to a strengthening of the interitem associations between the Rp+ and Rp− items, and this might lead to an increased recall at the final test for the Rp− items following retrieval of the Rp+ items.

According to this explanation, Rp− items presented early in a category would benefit more from integration with Rp+ items than later Rp− items because the items in the initial category positions are the ones that are reactivated with the presentation of a later item. According to this line of reasoning, the effects of integration for the early Rp− items might counteract the otherwise expected greater inhibition for these items.

According to M. C. Anderson, Green, and McCulloch (2000), the effects of integration are dependent on whether the integration takes place between the Rp+ and Rp− items or within each of the two sets of items. If integration occurs within the set of Rp+ items or within the set of Rp− items, rather than between, it should lead to an increase in the retrieval-induced forgetting effect. Because the benefit from integration should be present only at the beginning of a category, presenting all Rp− or all Rp+ items at the beginning of a category should lead to integration within the set and hence to an increase rather than a decrease in inhibition effects.

In Experiment 2, we therefore presented the Rp+ and Rp− items in a blocked fashion within their category to avoid integration between targets and competitors. In half of the practiced categories, we presented the three Rp− items first, followed by the three Rp+ items. In the other half of the practiced categories, the three Rp+ items were presented before the three Rp− items. If the failure to find stronger inhibition effects for the initial items was indeed caused by integration, then presenting the Rp− items at the beginning of a category should lead to especially strong retrieval-induced forgetting for the initial items because these not only are stronger (and hence more competitive) than the Rp− items from later serial positions, but are also less integrated with the Rp+ items, compared with the standard randomized order. Moreover, when the Rp− items are strong, the Rp+ items are weak, and this should increase the competition of the Rp− items even more. Hence, if integration is indeed a major factor, there should be an especially large difference between the amount of inhibition for the strong versus the weak Rp− items.

Experiment 2

Method

Participants

Fifty-one students from the University of Amsterdam participated in the experiment in exchange for course credits or payment.

All participants had Dutch as their mother tongue. The average age of the participants (22 male, 29 female) was 22 years (range = 18–32 years). All had normal or corrected-to-normal vision. None of the participants took part in the first experiment.

Design

In Experiment 2, two variables were manipulated within subjects: the retrieval-practice status and the order of presentation in the study phase. As in Experiment 1, the retrieval-practice status had three levels: Rp+, Rp−, and Nrp items. The order of presentations in the study phase had two levels: The Rp+ items were presented first followed by the Rp− items within a category, or the Rp− items were presented first followed by the Rp+ items during the study phase. Again, the counterbalancing resulted in eight study lists and four test lists that were used as between-subjects variables.

Materials and Procedure

The stimuli were the same as those used in Experiment 1. Except for the study phase, the same constraints as in Experiment 1 were used to construct retrieval practice and test lists. In the study phase, in two of the practiced categories, the three Rp− items were presented first followed by the three Rp+ items; in the other two practiced categories, the three Rp+ items were presented first followed by the three Rp− items. The procedure used was identical to that used in Experiment 1.

Results and Discussion

Retrieval Practice

In the retrieval-practice phase, 82% of the exemplars were correctly completed. This rate is similar to the average rate of practiced items found in Experiment 1 ($M = 84\%$) and that found by M. C. Anderson et al. (1994; $M = 82\%$).

Final Memory Test

For each participant, we computed the number of Rp+, Rp−, and Nrp items that were recalled. Recall percentages were analyzed with a repeated measures ANOVA in which retrieval-practice status was used as a within-subject variable, and study list and test list were used as between-subjects variables. A significant main effect was found for retrieval-practice status, $F(2, 86) = 109.76$, $p < .001$. A planned comparison revealed better recall of the Rp+ items ($M = 73\%$) compared with the Nrp items ($M = 45\%$), $F(1, 43) = 127.49$, $p < .001$, and revealed impaired recall for Rp− items ($M = 37\%$) compared with Nrp items ($M = 45\%$), $F(1, 43) = 16.91$, $p < .001$. As in Experiment 1, the retrieval-based learning and retrieval-induced forgetting effects reported by M. C. Anderson et al. (1994) were replicated.

Serial Position Effects in the Final Memory Test

The main question of Experiment 2 was whether presenting the Rp− items in a blocked fashion before the Rp+ items would lead to greater retrieval-induced forgetting effects, compared with presenting the Rp− items after the Rp+ items. Figure 3 shows the mean recall percentage for each retrieval-practice condition as a function of the within-category serial position. As in Experiment 1, a repeated measures ANOVA was carried out with retrieval-practice status and serial position as within-subject variables, and study list and test list were used as between-subjects variables for the Rp− and Nrp items.

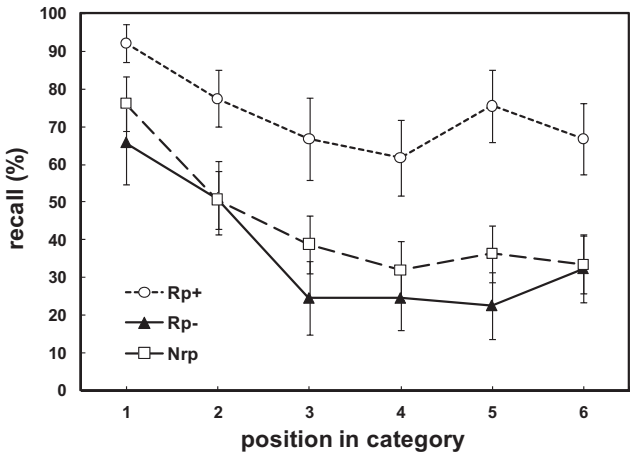

Figure 3. Mean recall percentage (with 95% confidence interval) for the retrieval-practice conditions as a function of serial position in a category in Experiment 2. Rp+ = practiced items from the practiced categories; Rp− = nonpracticed items from the practiced categories; Nrp = items from the nonpracticed categories.

The main effect of serial position was significant, $F(5, 215) = 30.36$, $p < .001$. Recall of the items decreased as a function of position. The items presented in the first position were recalled best ($M = 71\%$), followed by the second position ($M = 41\%$) and the third position ($M = 32\%$). From the third position on, the recall percentage was about the same ($M = 28\%$, 29%, and 33% for the fourth, fifth, and sixth positions, respectively). A planned comparison revealed a significant difference in recall between the first and the second positions, $F(1, 43) = 27.77$, $p < .001$, and between the second and the third positions, $F(1, 43) = 22.5$, $p < .001$. From the third position on, no significant differences were found. These findings replicate the results found in Experiment 1. Items presented at the beginning of a category were recalled better than items in the later positions. The Retrieval-Practice Status × Serial Position interaction effect was not significant, $F(5, 215) = 1.63$, $p > .10$. The recall of the Rp− items was similar to that of the Nrp items at all serial positions. Rp− items benefited as much from an earlier presentation within a category as the baseline items.

A further comparison was carried out on the two groups of Rp− items, given that in half of the categories, the Rp− items were presented in the first three positions (Rp− first), and in the other half of the categories, they were presented in the last three positions (Rp− second). The Nrp items were also split in half depending on their position within their category, resulting in Nrp first and Nrp second. The results are shown in Figure 4. A repeated measures ANOVA was carried out with retrieval-practice status (Rp− and Nrp) and item position (first or second) as within-subject variables.

The main effect of item position was significant, $F(1, 43) = 54.1$, $p < .001$. Items presented in the first three positions were

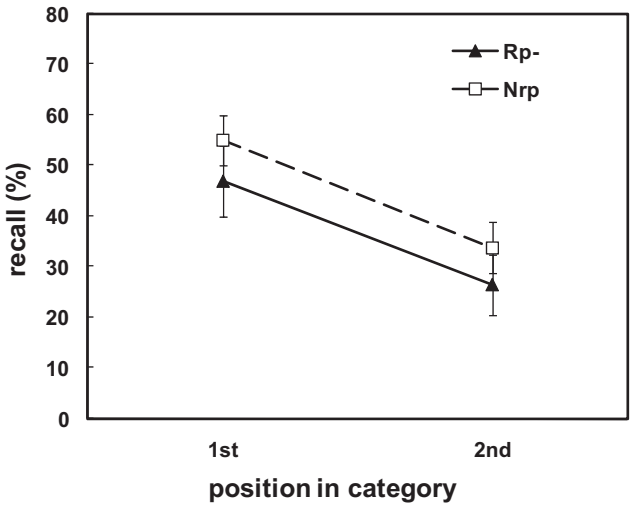

Figure 4. Mean recall percentage (with 95% confidence interval) for the Rp− and Nrp items as a function of whether they were presented in the first or second half of the list in Experiment 2. Rp− = nonpracticed items from the practiced categories; Nrp = items from the nonpracticed categories.

recalled better (M = 51%) than items presented in the last three positions (M = 30%). Retrieval-practice status was also significant, $F(1, 43) = 16.83$, $p < .001$. Nrp items (M = 44%) were recalled better than Rp− items (M = 37%). The Retrieval-Practice Status × Item Position interaction was not significant, $F(1, 43) < 1$. Retrieval-induced forgetting in the first three positions (M = 8%) was similar to that in the last three positions (M = 7%). Hence, the Rp− items were not impaired to a greater extent when presented in the first three positions.

Finally, we compared the data of the two experiments to determine the extent to which integration had an effect on the amount of retrieval-induced forgetting that was observed. In Experiment 1, integration would have led to a decrease in the retrieval-induced forgetting effect (because here Rp− and Rp+ items should have been integrated together), whereas in Experiment 2, integration would have led to an increase in retrieval-induced forgetting (because Rp− items should have been integrated with other Rp− items rather than with the Rp+ items, leading to greater interference). The magnitude of the retrieval-induced forgetting effect was, indeed, somewhat smaller in Experiment 1 (4%) than in Experiment 2 (8%). However, this difference was not statistically significant. A repeated measures analysis with retrieval-practice status as a within-subject variable and experiment as a between-subjects variable showed no significant Experiment × Retrieval-Practice Status interaction, $F(2, 186) < 1$. A planned comparison also showed that the difference between the Rp− and Nrp items was not significantly different in the two experiments, $F(1, 93) = 1.57$, $p > .20$. These results indicate that integration is unlikely to have been a major factor in the results of these experiments.

Experiment 3

In the first two experiments, we manipulated item strength by position of the item within its category, defining early items as strong and later items as weak. In the next experiment, a more direct manipulation of item strength was used: Some of the items were presented once and some twice during the initial study phase. As in the previous experiments, this enabled us to manipulate item strength independently from the particular exemplar because all items could take on the role of strong or weak exemplar.

In this experiment, some of the Rp− items in the categories were presented twice in the original study list and were therefore studied better. If items in a category are studied better, they should become stronger exemplars of the category, at least in the context of the present experiment. If stronger items compete more during the retrieval phase and, therefore, have to be inhibited to a larger degree, items presented twice should be more impaired on the final recall test.

Method

Participants

Forty-one students from the University of Amsterdam participated in the experiment in exchange for course credits or payment. All participants had Dutch as their mother tongue. The average age of the participants (4 male, 37 female) was 19.95 years (range = 17–28 years). All had normal or corrected-to-normal vision.

Design

Two factors were manipulated within subjects: the retrieval-practice status and the number of presentations in the study phase. As in the previous experiments, half of the categories were practiced, and within these categories, half of the items were practiced during the retrieval-practice phase (Rp+) and the other half were not practiced (Rp−). From the remaining categories, none of the items were practiced in the retrieval-practice phase (Nrp). The Rp− items and the corresponding Nrp items were presented either once or twice on the original study list (denoted as Rp−1, Rp−2, Nrp1, and Nrp2, respectively). The counterbalancing of words in the study phase and within the categories resulted in eight study lists. For the test phase, two types of test lists were constructed: In half of the lists, the practiced categories were tested first; in the other half, the unpracticed categories were tested first.

Materials and Procedure

The categories were the same as those used in Experiment 1. In each category, two additional exemplars of medium strength were chosen, resulting in eight exemplars per category. The average taxonomic frequency in the categories was 18.48 (range = 5–42, median = 16), according to Hudson's (1982) category norms, and 16.45 (range = 5–38, median = 16), according to Storms's (2001) category norms. As in Experiment 1, no two items in the categories began with the same two letters. Item length varied between four and eight letters and between one and three syllables. The average length was 6.02 letters and 1.97 syllables.

The same constraints as in Experiment 1 were used to construct the study, retrieval-practice, and test lists. In the study phase, 2 items in each category were presented twice, resulting in 10 items per category. The items presented twice were half of the Rp− items and 2 matching Nrp items from the baseline category. Rp+ items were always presented once during the study phase. In total, 80 experimental and 16 filler category–item pairs were presented on the study lists. In the retrieval phase, 60 category–item stem

pairs were presented. The procedure used was identical to that used in Experiment 1.

Results and Discussion

Retrieval Practice

In the retrieval-practice phase, 78% of the exemplars were correctly completed. This rate is somewhat lower than the rates in Experiments 1 and 2, probably because of the increase in list length.

Final Memory Test

For each participant, we computed the number of Rp+, Rp−, and Nrp items that were recalled. Recall percentages for all items presented once were analyzed with repeated measures ANOVAs in which retrieval-practice status was a within-subject variable, and study list and test list were between-subjects variables. A significant main effect was found for retrieval-practice status, $F(2, 66) = 92.12$, $p < .001$. A planned comparison revealed better recall of Rp+ items ($M = 59\%$) compared with Nrp items ($M = 41\%$), $F(1, 33) = 69.59$, $p < .001$, and revealed impaired recall for Rp− items ($M = 30\%$) compared with Nrp items ($M = 41\%$), $F(1, 33) = 26.62$, $p < .001$. The recall percentages for the Rp− and Nrp items were similar to the percentage for the weak items in M. C. Anderson et al.'s (1994) study (34.7% and 41.0%, respectively).

Effect of the Explicit Item Strength Manipulation

The main question of Experiment 3 was whether explicit manipulation of item strength of the nonpracticed items would affect the retrieval-induced forgetting effect. Instead of choosing items with different association strength, we presented some of the nonpracticed items twice. A two-way within-subject ANOVA was carried out to test the effect of item strength. A significant main effect was found for retrieval-practice status, $F(1, 33) = 22.26$, $p < .001$. The Nrp items were recalled better ($M = 50\%$) than the Rp− items ($M = 41\%$). The main effect of the number of presentations in the study phase was also significant, $F(1, 33) = 50.62$, $p < .001$. Not surprisingly, items presented twice were recalled better ($M = 55\%$) than items presented only once during the study phase ($M = 36\%$). The Retrieval-Practice Status × Number of Presentations interaction, however, was not significant, $F(1, 33) = 1.25$, $p > .20$. The results are shown in Figure 5.

Hence, the benefit from the additional presentation was the same for the Rp− items as for the Nrp items. These findings again contradict the inhibition account: Items presented twice are stored better and, therefore, should be more available and compete more during the retrieval-practice phase. If items that are more available are indeed suppressed more, these items should have shown more inhibition compared with the items presented once.

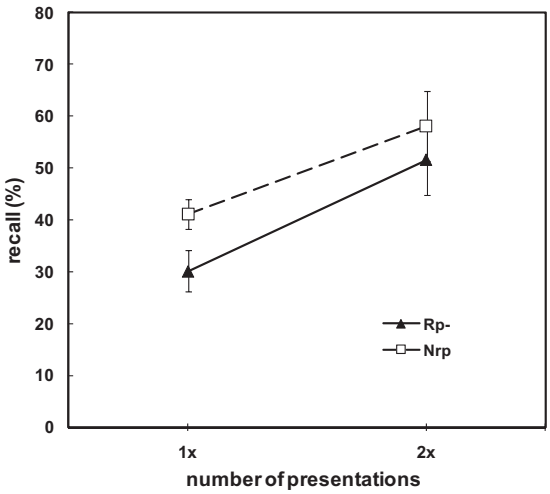

Figure 5. Mean recall percentage (with 95% confidence interval) for the Rp− and Nrp items as a function of the number of presentations in Experiment 3. Rp− = nonpracticed items from the practiced categories; Nrp = items from the nonpracticed categories.

General Discussion

The purpose of the present study was to investigate the effect of item strength in retrieval-induced forgetting. More specifically, we wanted to determine whether explicit manipulation of strength leads to larger inhibition effects, as predicted from current accounts of the inhibition theory of forgetting.

To summarize the data, in all three experiments, we found retrieval-based learning: Practiced items were recalled better than nonpracticed items. We also found retrieval-induced forgetting: Recall of unpracticed items in the practiced category was impaired compared with recall of the baseline items. These findings replicate the standard effects found in studies with the retrieval-practice paradigm. In the first two experiments, an effect of item position within a category was found. The recall of items strongly depended on their position within a category. Items presented at study in the first two positions within a category were recalled better than items presented in later positions. However, no interaction between item strength and amount of inhibition was found. In the second experiment, we grouped the Rp− items to avoid integration between the Rp+ and Rp− items. This grouping of the items should have led to an even larger effect of item strength; however, it did not change the pattern that was found in Experiment 1. In the third experiment, we also found that an additional presentation of the Rp− items had no effect on the magnitude of the retrieval-induced forgetting effect.

M. C. Anderson et al. (1994) argued that stronger items should be more inhibited than weaker items, because they compete more for recall during the retrieval-practice phase. To overcome this competition, they have to be inhibited to a greater extent than weaker items. Contrary to the expectations based on inhibition theory, we did not observe more inhibition for stronger items than for weaker items in any of the three experiments. In summary, the present findings do not agree with the expectations based on the inhibition account of forgetting proposed by M. C. Anderson (2003) and others.

One factor that is often put forward to explain why an expected inhibition effect does not occur, is the notion of integration. M. C. Anderson and McCulloch (1999) argued that integration can eliminate retrieval-induced forgetting. The lack of impaired recall in the first two positions observed in the first experiment might be explained by the integration of the early Rp− and later Rp+ items.

Under the assumption that the increased recall of the items in the initial serial positions is due to reactivation of those items by the presentation of later items (including later Rp+ items), the early Rp− items would benefit more from integration with Rp+ items, and this might counteract the predicted greater inhibition for the early Rp− items.

However the integration explanation holds only for Experiment 1. Here Rp+ and Rp− items were presented in an alternating order within a category. In Experiment 2, however, we grouped the Rp− items, presenting these either in the first half of the category or in the second half. If integration was the main reason for the failure to obtain the predicted effects in Experiment 1, then grouping the Rp− items at the beginning of a category should have led to especially strong inhibition effects. Integration would then have occurred primarily between the Rp− items; therefore, they could not have benefitted any more from the practice of the Rp+ items. Moreover, if integration did occur, the inhibition for the early Rp− items should have been even stronger in Experiment 2, because integration took place between items belonging to a similar condition: Rp− items should have been integrated with Rp− items and Rp+ items with Rp+ items. To put it differently, integration in this case should have led to competitor–competitor similarity. According to M. C. Anderson, Green, and McCulloch (2000), increasing competitor–competitor similarity should have led to larger retrieval-induced forgetting. Because the integrated Rp− items at the beginning of a category were stronger competitors, they should have shown more inhibition than integrated Rp− items in the later positions of a category. However, we found that retrieval-induced forgetting was similar for Rp− items at the beginning and at the later positions within a category.

A similar argument based on integration could be advanced to explain the results of our third experiment. M. C. Anderson and McCulloch (1999) demonstrated that presenting the study list twice increases spontaneous integration, hence reducing retrieval-induced forgetting. M. C. Anderson and McCulloch, however, repeated the whole list twice, whereas we presented only two Rp− items twice out of the eight items in a category. Because a tendency to integrate the two repeated Rp− items would increase rather than decrease the inhibition effect, integration could be a valid explanation for our results only if all items in a category were integrated together. Only in this case could the Rp− item benefit form the recall of the Rp+ items.

We could demonstrate whether Rp+ and Rp−2 items were integrated by examining the output order of the Rp+ and Rp− items in the two conditions. If Rp+ items were more frequently followed by an Rp−2 item than by an Rp−1 item, then Rp−2 items could have benefited from integration with the Rp+ items. In contrast to this, we found that the conditional probability that an Rp−2 item was immediately preceded by a Rp+ item (65%) was lower than the conditional probability that an Rp−1 item was preceded by a Rp+ item (74%); this was conditional on the item being recalled and there being at least one other item in the output sequence before the Rp−2 or Rp−1 item, respectively. Another reason to believe that our pattern was not due to integration comes from an experiment by Jakab and Raaijmakers (2009). In this experiment, we manipulated only the strength of the Rp+ items. We found more retrieval-induced forgetting when the Rp+ items were presented twice than when they were presented once. If, indeed, only partially repeating items during the study phase leads to integration of the whole category, then the opposite pattern should have been found (i.e., less inhibition for Rp− items when the Rp+ items were repeated in the study phase). Finally, it should be noted that in the present experiment with some items repeated, there was a clear overall retrieval-induced forgetting effect, again making it unlikely that there was a major overall effect of integration. In summary, it is questionable whether the integration argument could provide an alternative explanation for the results of our third experiment, making it compatible with the inhibition account of forgetting.

A further objection to the present experiments might be that our strength manipulation was not a valid test for the effect of item strength on retrieval-induced forgetting. M. C. Anderson et al. (1994) varied item strength by taxonomic frequency, whereas the present study manipulated serial position of the items and the number of presentations during the study phase. It would not be correct, however, to claim that the only valid manipulation of item strength is through the variation of taxonomic frequency. The purpose of the item strength manipulation in these experiments is to increase (or decrease) the probability of nontarget item activation during the retrieval of the target item given a common cue. To check that our strength manipulation did lead to an increase in the probability of activating the stronger items when the category cue was presented, we analyzed the output order of the Nrp items on the test phase in the first experiment. As mentioned earlier, we found that the conditional probability that an item presented in the first position during the study phase was recalled as the first item in the test phase (given that it was recalled) was 64%. There is no reason to expect a different pattern for the Rp− items during the retrieval-practice phase (i.e., more and earlier activation of initial items and hence more interference).

There is probably less reason to question the validity of the strength manipulation in our third experiment, because in other experiments by inhibition proponents, such as those testing the list strength effect, similar manipulations have been used to vary item strength. Even if there is no doubt that repeatedly studying an item leads to stronger cue–item associations, it might be argued that studying an item once does not really lead to a weak level of strength but rather leads to a moderate level of strength. Bäuml (1998), for example, demonstrated in his experiment that strong and moderate items show the same amount of forgetting. Consequently, the lack of difference in retrieval-induced forgetting in our two conditions might have been due to the fact that our weak items were not weak enough. This argument, however, can be rejected on the basis of our baseline data. The recall of once-presented Nrp items was 41%, which is the same as the recall that M. C. Anderson et al. (1994) obtained for weak items; the recall of the twice-presented Nrp items was 58%, which is 2% higher than that of their strong items (56%). Consequently, such an argument cannot be used to explain the lack of a strength effect in the present experiments. In summary, the argument that our strength manipulations would not be appropriate for testing the inhibition account of forgetting does not appear to be a valid one.

Finally, one might argue that our experiments failed to demonstrate the inadequacy of the inhibition account because we obtained only a null effect, showing no effect of strength, rather than a reversed effect. Even disregarding the obvious problem that an alternative theory, one not based on inhibition, would not necessarily predict a reversed effect, the strength of such an argument is

also reduced by taking a broader (some might argue more correct) perspective on null hypothesis testing. For example, within a Bayesian approach to hypothesis testing, a comparison would be made between the relative likelihoods of the null and alternative hypotheses given the present data. A Bayesian analysis using the procedures described by Glover and Dixon (2004; see also Wagenmakers, 2007) leads to the conclusion that, in Experiment 3, the null hypothesis is 6.4 times more likely than the alternative hypothesis (a similar value was obtained with the method described by Rouder, Speckman, Sun, Morey, & Iverson, in press).[2] Note that this is even a conservative estimate because it is based on a two-sided *t* test. Even if one is not convinced of the validity of such a Bayesian approach, it is still informative to note that if one assumes that the inhibition account should predict at least a 5% difference in the retrieval-induced forgetting effect for strong versus weak items (a conservative estimate based on the work of M. C. Anderson et al., 1994), the inhibition hypothesis significantly deviates from the observed data in Experiment 3, $t(40) = -1.84$, $p = .036$. Note that in this analysis, the inhibition prediction of a 5% difference becomes the null hypothesis, eliminating the problem of accepting the null hypothesis.

M. C. Anderson (2003, 2005) used the item strength manipulation to support a basic property of the inhibition theory: the assumption of interference dependence. M. C. Anderson claimed that retrieval-induced forgetting arises only if related memories interfere during the retrieval of the target. Because strong items are more likely to compete for recall, they are more likely to be inhibited; therefore, in the present experiments, more impairment should have been found for the stronger items. Thus, the present experiments add to a growing number of results (Major et al., 2008; Williams & Zacks, 2001) that do not support the interference dependence property of the inhibitory account.

As we argued in the introductory section, a reasonable prediction of competitive retrieval explanations of forgetting is that the retrieval-induced forgetting effect should be approximately independent of the strength of the Rp− items. The present results are clearly in line with this prediction. The conclusion reached by proponents of the inhibition account—that retrieval success in recall tasks is not affected by competition from other associations (e.g., M. C. Anderson, 2003)—therefore seems premature. Such a conclusion is also somewhat paradoxical given that competition by other items provides the main rationale for inhibition to occur during retrieval practice: Without competition there would be no need for suppression of those other items.

Although the present results are in line with the competitive retrieval account of forgetting, such an account would still have to provide an explanation for those findings that do show strength-dependent retrieval-induced forgetting effects (M. C. Anderson et al., 1994; Bäuml, 1998; Storm et al., 2007). Of these experiments, the one by M. C. Anderson et al. (1994) comes closest to the present experiments. We do not have a ready explanation for why M. C. Anderson et al. did not find a retrieval-induced forgetting effect for their weak category exemplars. Perfect et al. (2004) noted that in a subsequent experiment, M. C. Anderson and Spellman (1995) did find a substantial retrieval-induced forgetting effect with category members that also would have to be classified as weak. Given that Williams and Zacks (2001), in their replication of M. C. Anderson et al.'s experiment, also obtained a significant retrieval-induced forgetting effect for the weak category members, we are inclined to conclude that M. C. Anderson et al.'s result may have been a chance result.

The present finding of strong within-category serial position effects may also provide an explanation for why Dodd et al. (2006) found a retrieval-induced forgetting effect only when the items in the different conditions were presented in a random order but found no retrieval-induced forgetting effect when the items were grouped within the study list in such a way that the Rp− items were presented earlier in the list. In such a design, the Rp− items benefit more from the primacy effects and, therefore, show better recall than the Nrp items (because performance on these items is the average over all serial positions). Dodd et al. proposed strategy disruption (Basden & Basden, 1995; Basden, Basden, & Galloway, 1977) as a possible account for these findings. If, indeed, Rp− items benefited from encoding strategies, then subgrouped Rp− items presented later in a list should also show a similar pattern. The arrangement of our data was similar to that of Dodd et al. (in their experiments, they did not match the serial positions of the Nrp items and the Rp− items). Comparing the recall of all Nrp items ($M = 45\%$) with the first half of the Rp− items ($M = 47\%$) provides no evidence for retrieval-induced forgetting ($M = -2\%$). However, comparing the later presented Rp− items ($M = 27\%$) with the recall of all Nrp items ($M = 45\%$), a large retrieval-induced forgetting effect is observed ($M = 18\%$). We may therefore conclude that the serial position effect is a better explanation of Dodd et al.'s findings than their strategy disruption account.

The effect of serial position is, of course, not unique to the experiment of Dodd et al. (2006) and the first two experiments of this article. In all experiments using the retrieval-practice paradigm, similar effects are likely to be present. The only difference between our first two experiments and other experiments using this paradigm is that we explicitly controlled the positions of items within a category. In the retrieval-practice paradigm, the practiced category always consists of items from two conditions: practiced and nonpracticed items. In the study list, the positions of the Rp+ and Rp− items within a category are chosen randomly. Therefore, it is possible that more or fewer Rp− items are presented in the earlier positions within a category. If serial position has such a large impact on recall, the exact positioning of the items could strongly influence the retrieval-induced forgetting effect found in these experiments. This may be demonstrated by rearranging the data of our first experiment. If the Rp− items take the first, third, and fifth positions in a category, no retrieval-induced forgetting would be found ($M = -0.3\%$); however, if the Rp− items take the second, third, and fifth positions in a category, it would result in a relatively large retrieval-induced forgetting effect ($M = 9\%$). Therefore, our finding of strong within-category serial position effects implies that researchers using the retrieval-practice paradigm should carefully control the serial positions of Rp+ and Rp− items, or they should match these items with the corresponding Nrp items in terms of position within a category. Even slight imbalances due to random or quasi-random placement of the items without exact counterbalancing may have a noticeable effect on the difference between the Rp− and the Nrp items and hence on the retrieval-induced forgetting effect.

[2] This computation was based on the Bayes factor calculator, which may be found at http://pcl.missouri.edu/bayesfactor

Conclusion

The present study investigated the role of item strength in retrieval-induced forgetting. The pattern of results does not agree with the expectations of an inhibitory account. Given that interference dependence is a fundamental assumption of the inhibition theory, the present results pose a serious problem for the inhibition account of forgetting.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

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

Anderson, M. C. (2005). The role of inhibitory control in forgetting unwanted memories: A consideration of three methods. In N. Otha, C. M. MacLeod, & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 301–331). Tokyo, Japan: Springer-Verlag.

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

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

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

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

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.

Basden, D. R., & Basden, B. H. (1995). Some tests of the strategy disruption interpretation of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1656–1669.

Basden, D. R., Basden, B. H., & Galloway, B. C. (1977). Inhibition with part-list cuing: Some tests of the item strength hypothesis. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 100–108.

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

Dodd, M. D., Castel, A. D., & Roberts, K. E. (2006). A strategy disruption component to retrieval-induced forgetting. *Memory & Cognition, 34,* 102–111.

Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review, 11,* 791–806.

Hudson, P. T. W. (1982). *Preliminary category norms for verbal items in 51 categories in Dutch* (Internal report/Vakgroep Psychologische Funktieleer). The Netherlands: Katholieke Universiteit Nijmegen.

Jakab, E., & Raaijmakers, J. G. W. (2009). *The modulating effect of the target item strength in retrieval-induced forgetting.* Manuscript in preparation.

Major, J. C., Camp, G., & MacLeod, C. M. (2008). *Encoding strength does not modulate retrieval-induced forgetting: Evidence inconsistent with an inhibition account.* Manuscript submitted for publication.

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

Perfect, T. J., Stark, L., Tree, J. J., Moulin, C. J. A., Ahmed, L., & Hutter, R. (2004). Transfer appropriate forgetting: The cue-dependent nature of retrieval-induced forgetting. *Journal of Memory and Language, 51,* 399–417.

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

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (in press). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review.*

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63–77.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

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

Storms, G. (2001). Flemish category norms for exemplars of 39 categories: A replication of the Battig and Montague (1969) category norms. *Psychologica Belgica, 41,* 145–168.

Wagenaar, W. A. (1969). Note on the construction of digram-balanced Latin squares. *Psychological Bulletin, 72,* 384–386.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14,* 779–804.

Williams, C. C., & Zacks, R. T. (2001). Is retrieval-induced forgetting an inhibitory process? *American Journal of Psychology, 114,* 329–354.

Wood, G., & Underwood, B. J. (1967). Implicit responses and conceptual similarity. *Journal of Verbal Learning and Verbal Behavior, 6,* 1–10.

Received April 7, 2008
Revision received December 19, 2008
Accepted December 21, 2008 ■