Journal of Experimental Psychology:
Human Learning and Memory
1980, Vol. 6, No. 1, 91-105

Output Interference in the Recall of Categorized and Paired-Associate Lists

Henry L. Roediger III and Stephen R. Schmidt
Purdue University

Three experiments are reported in which, following presentation of a categorized list, subjects attempted recall of words within categories when given the category names as cues. Recall of words declined with the test position of the category. This output interference was not increased (a) when some items from each category were given as cues in addition to the category name (Experiment 1), (b) when the first categories recalled contained nine rather than three words (Experiment 2), and (c) when the categories in the list were semantically related rather than unrelated (Experiment 3). Interference in recall of categories tested later was not affected by the number of words recalled from prior categories. Output interference in paired-associate recall was obtained in a fourth experiment. These experiments indicate that the act of recall can itself serve as a source of forgetting. Implications of the results for several theories of recall are discussed.

Several different lines of evidence have been used to show that one source of forgetting is the act of recall itself. (See Roediger, 1974, for a review.) The first systematic study of output interference was that of Tulving and Arbuckle (1963). They presented subjects with lists of 10 paired associates in which the digits 1–10 were the stimuli and common words were the responses. Subjects' recall was cued by the stimuli, with output order counterbalanced so that each stimulus occurred equally often at each output position. The primary finding was that recall of responses that had been presented in the initial serial positions of the input list did not vary with output position, but recall of responses presented late in the input sequence became increasingly poor across the output positions. Output interference was most dramatic for the last two or three items, but there was some evidence that it extended to items in the fifth and sixth input positions. The conclusion that may be drawn from this experiment is that output interference is due only to a loss of information from primary memory (Waugh & Norman, 1965), and that there is no output interference in recall of information from secondary memory. Evidence from other paired associate experiments in which short lists were employed does not alter this conclusion (Arbuckle, 1967; Tulving & Arbuckle, 1966). The results of Experiment 4 in the present series constitute the first evidence of output interference in paired-associate recall in secondary memory, though in this instance the lists were relatively long (20 pairs) and were composed entirely of words.

Evidence from other cued recall tasks also implicates the operation of output interference in secondary memory (Dong, 1972; Roediger, 1973; Smith, 1971, 1973; Smith, D'Agostino, & Reid, 1970). In these experiments subjects were presented words from common conceptual categories for later recall (e.g., types of fish, fruits, furniture, etc.). The words were typically grouped together by category membership

The authors thank Beth Adelson and M. Ellen McIntosh for their help in conducting Experiment 4.

Requests for reprints should be sent to Henry L. Roediger III, Department of Psychology, Purdue University, West Lafayette, Indiana 47907.

Copyright 1980 by the American Psychological Association, Inc. 0096-1515/80/0601-0091$00.75

at presentation and were often preceded by the appropriate category name. Subjects received category names as recall cues in a counterbalanced order, and subjects typically received a fixed amount of time per category to recall the relevant words. The general finding of these studies was that recall decreased in a linear manner with output position of the category in the recall sequence (Roediger, 1973; Smith, 1971, 1973; Smith et al., 1970).

Several findings rule out the possibility that this decrease in recall with test position can be attributed to loss of information from primary memory. Smith (1971) showed that the output interference function was not affected when recall of the last category presented was not included in the data, and that a 30-sec delay between list presentation and recall did not affect the output interference function. This latter finding would also seem to indicate that the observed decline in recall with output position is not simply due to the "natural" forgetting resulting from interpolation of a demanding distracting task between presentation and recall of material. The deficit in recall of items from the categories seems specific to the act of recalling previous categories, not to general interference of the sort supplied by mental arithmetic. (See Roediger, 1978, Experiment 3, for related evidence.) A final bit of evidence that the decline of recall with output position is not due to a loss of information from primary memory is that output interference is found for categories in all portions of the input list (Smith, 1973).

The question that initiated the present experiments is whether the critical factor producing poorer recall of items from categories tested later in the output sequence is the recall of prior categories, as has been assumed, or the number of items recalled within the recalled categories. There are several pieces of evidence that provide indirect support for the supposition that it is the sheer number of words recalled, rather than the number of prior categories from which recall is attempted, that is important. Smith (1971, Experiment 4) showed that when the size of categories was varied between lists, output interference became greater as category size was increased. Since there were more items recalled in the case of larger than in the case of the smaller categories prior to recall of items from later categories, this result is consistent with the notion that it is the number of items recalled rather than the number of categories from which recall is attempted that is important in determining output interference. On the other hand, the length of the to-be-remembered lists was necessarily confounded with category size in Smith's (1971) experiment, so it may be this factor and not number of prior items recalled that produced these results.

Other evidence reported by Roediger (1973) agrees with the notion that it is the number of prior words recalled rather than the number of categories accessed that determines output interference in successively recalled categories. In this study subjects were tested by providing them with category names following presentation of a categorized word list, and in a between-subjects design, also providing them with either some items from the category as cues or no such items from the category. The presentation of some items from a list as cues for recall of the remainder is referred to as a part-list cuing manipulation, and it is found to inhibit recall of other items in a category when compared to recall of these same words with only the category name as a cue (Roediger, 1973; Rundus, 1973; Slamecka, 1972; Watkins, 1975). One theory of this effect is that the presentation of part-list cues has an effect similar to their actual recall in that their representation in memory is strengthened and they compete at retrieval with other to-be-recalled words, thus providing the observed inhibition in recall of these other words (Roediger, 1974; Rundus, 1973). If presentation of part-list cues functions in a manner similar to the actual recall of the cue words, and if recall of words produced the output interference effect across re-

called categories, then one might have expected to observe greater output interference for subjects in Roediger's (1973) experiment who received part-list cues in addition to category names as compared to those who received only category names. Besides the usual inhibiting effect of part-list cues, Roediger (1973) did also report a marginally reliable interaction of the part-list cuing manipulation with output position of categories. The proportion of words recalled from successively tested categories declined more rapidly for subjects who received item cues in addition to category names than for those subjects who received only category names.

The first two experiments of the present series were designed to gain further evidence about the issue of whether it is the number of categories accessed or the sheer number of words recalled that is the important variable producing output interference across successively tested categories. Experiment 1 is an attempt to replicate the marginally reliable result reported by Roediger (1973). Subjects were presented categorized lists and tested with successively presented category names. For some lists only category names were presented as cues, for others some items from the category were also presented. It was expected, in line with the earlier results, that presentation of item cues in addition to category names would cause greater output interference across tested categories than presentation of category names only.

A second question examined in Experiment 1 was which items within a category are subject to output interference. Recall drops across successively recalled categories, but does it do so equally for all serial positions of words within a presented category? Tulving and Arbuckle (1963) reported output interference only for the recency portion of a paired associate list; it may similarly be the case that output interference across categories is due to increasingly poor recall of the last couple of items presented in each category with the other words showing no such effect.

Experiment 1

Method

Subjects. The subjects were 80 Purdue University undergraduates who participated in the experiment to fulfill part of a course requirement. They were tested in groups of from 5 to 10.

Materials. Twenty categories were selected from the Battig and Montague (1969) category norms and were randomly placed into two groups of 10. Two lists were made from each set of 10 categories, with seven items selected from each category for each list. For one list in each set of categories, the words selected were the even-numbered words of the first 14 listed in the norms, whereas the other list contained the odd-numbered words. A total of four 70-item lists were constructed, with each list containing 10 categories with seven items per category.

Design. Subjects recalled two lists, one in each of two conditions. In one condition, the category name only (CNO) condition, subjects received only category names as recall cues with the instruction to recall as many words as possible from each category. In the other condition, the category names plus item cues (CN + IC) condition, subjects received category names and four of the seven items from each category as recall cues.

Procedure. The procedure followed closely that used in Roediger's (1973) experiment. In the initial instructions subjects were told about the categorized nature of the lists, including the facts that there would be 10 categories per list and seven items per category. The words were grouped by category at presentation, and the category name preceded the items in each category. Subjects were told that they were not responsible for remembering the category names, but should concentrate only on remembering the words in each category. The lists were tape recorded, and items were presented at the rate of 2 sec per item. Category names were presented in cadence with category instances but were distinguished by intonation. Subjects were presented with and recalled two lists, one under CNO conditions and the other under CN + IC conditions, in a counterbalanced order. The four lists were used equally often across subjects, though any particular subject received two lists containing different categories.

After presentation of each list, subjects were engaged in a digit cancellation task for 30 sec before turning over their recall sheets. The recall sheets consisted of two pages with five columns on a page. At the top of each column was a category name and, for CN + IC subjects, four items randomly selected from the category. Subjects were told that the category names and items from the categories were intended to aid their memories. With regard to the item cues, subjects were told that for one list "some of the words from each category are also presented to serve as hints in remembering the other words. Study all the clues from each category carefully and try to recall the

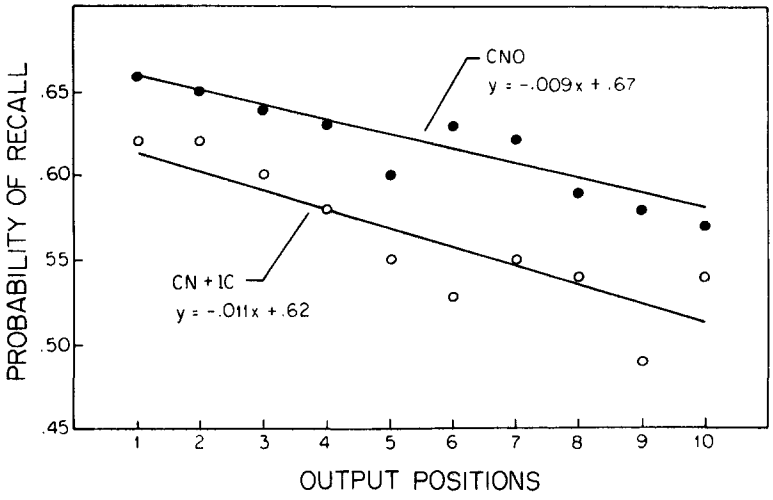

Figure 1. Probability of recall of items within categories across the 10 output positions in the category name only (CNO) and category name + item cue (CN + IC) conditions of Experiment 1. (Lines represent the best-fitting linear functions.)

other items in that category." Subjects were instructed to cover their recall sheets with a blank sheet and to expose successive columns only on instructions from the experimenter. The blank sheet contained a window that allowed only one column to be exposed at a time. Subjects were instructed to try to recall words from the particular category that was in view. They were instructed not to go back to previous categories nor forward to new categories except on instructions from the experimenter. Subjects worked through the categories on the first recall sheet from left to right and then turned to the second sheet on instruction and proceeded in the same fashion. Output order of the categories was governed by a balanced Latin square. Subjects were allowed 45 sec to recall the words in each category. After subjects had attempted to recall words from all 10 categories, they were given an additional 2.5 min to attempt recall of any other words from all the categories. These words were to be written below each category name in a special section at the bottom of the recall sheets.

After subjects finished recalling the first list, there was a brief rest before they were presented the second list and the recall procedure was repeated. One list was tested in the CNO condition for each subject and one in the CN + IC condition.

Results and Discussion

Very few additional words were recalled during the final 2.5 min of the recall period, and thus only words recalled during the initial recall attempt for each category are included in the results reported here. Inclusion of the additional words would affect none of the conclusions drawn. The output interference functions for the CNO and CN + IC conditions are presented in Figure 1, in which the proportion of words recalled for the two conditions is plotted against output position. The lines represent the best-fitting linear functions. These functions account for 56% and 42% of the variance in the CNO and CN + IC conditions, respectively. The superiority of recall in the CNO condition to that in the CN + IC condition was statistically reliable, $F(1, 79) = 7.99$, $p < .01$, $MS_e = 1.85$, as has been found in a number of other experiments (Roediger, 1973; Rundus, 1973; Watkins, 1975). Presenting items from a category as cues in addition to the category name impairs recall relative to that with only the category name. The decline in recall across output positions was also reliable, $F(9, 711) = 2.83$, $p < .005$, $MS_e = .78$, replicating the findings of Smith (1971) and others. However, the interaction between conditions and output positions was not reliable, $F(9, 711) < 1$, $MS_e = .77$. The slopes of the best-fitting lines were practically identical, −.009 in the CNO condition and −.011 in the CN + IC condition. Thus there was a failure to replicate the marginally reliable result of Roediger (1973). Since there was

greater statistical power in the present experiment and the results in Figure 1 do not even hint of an interaction, we conclude that presenting item cues for each category does not produce an increase in output interference in categorized list recall. If presentation of items is functionally equivalent to their recall, then the failure to find the interaction argues against the position that it is the number of prior items recalled that determines output interference. However, this proposition receives a more direct test in Experiment 2.

The output interference results in the CNO condition were also examined to see if recall of particular items within a category contributed disproportionately to the output interference effect. Tulving and Arbuckle (1963) found that only recency items in the paired-associate situation were subject to output interference. It is possible that only recall of the last item or two in a category declines across output positions in the categorized list situation. Although items were always presented in one order within a category, all categories in the four lists were tested in each output position in a counterbalanced manner. Thus no confounding could exist between output position and particular categories or words within categories. The serial position functions are presented in Figure 2 for the first four, middle three, and last three categories recalled in the output sequence. It is apparent from Figure 2 that there is no special part of the within-category serial position curve that is responsible for output interference in categorized list recall. There is roughly the same drop in the first few serial positions as in the last few; at any rate, there is no support for the idea that only the last couple of serial positions within the category contribute to the decline in recall across output positions. Overall there seems to be a large primacy effect but no recency effect in the within-category serial position function. Across all output positions the proportion of words recalled at serial positions 1–7 was .72, .62, .62, .61, .56, .59, and .59, respectively. Each proportion is based on 800 observations. Carey

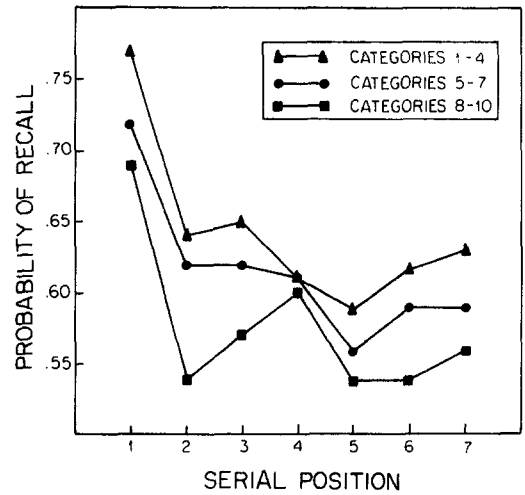

Figure 2. Probability of recall of items within categories plotted against input position of the items with the (grouped) test position of the categories as the parameter.

and Lockhart (1973) also reported only primacy in the within-category serial position function when subjects expected a recall test, but they found little effect of serial position when a recognition test was expected.

Two ancillary analyses rule out two possible artifacts for the reliable findings of Experiment 1, the inhibition produced by part-list cues and the output interference across categories. The results presented in Figure 1 are the proportion of words recalled based on the possible recall of seven words for the CNO condition but only three words for the CN + IC condition, since the other four words were presented as cues in this condition. If the subjects, when tested in the CNO condition, are only scored on the same three words that other subjects had as targets in the CN + IC condition, there is still a difference favoring the CNO condition. The proportion of these critical words recalled per category was .63 in the CNO condition and .56 in the CN + IC condition, when collapsed across output position. (The proportion of all words recalled out of seven in the CNO condition was .62.) Thus the inhibition produced by presenting item cues was not due to some item selection artifact in the particular targets re-

called. This failure to find facilitation from presenting item cues is taken as evidence contrary to the proposition that there are direct interitem associations between words in a category (Slamecka, 1972). It should also be noted that when subjects in the CNO condition are scored only on critical items, there is no change in the conclusion that the slope of the output interference function does not differ from that of the CN + IC condition.

A similar uninteresting interpretation of the output interference function can be ruled out on the basis of intrusion data. It might be imagined that output interference across categories merely reflects the tendency that subjects become less willing to respond during the recall test due to fatigue, boredom, or other factors. Evidence concerning this changing criterion interpretation of output interference can be found in the intrusion rates. Subjects recalling each list were scored on a common set of seven intrusions, the first seven in the Battig and Montague (1969) norms that were not used in making up the list. (It should be remembered that two lists were composed from the same categories by selecting either the first seven odd or even items from the norms.) Thus subjects in the CNO and CN + IC conditions could be scored on the same set of intruding words, and in fact these words comprised almost all extralist intrusions. There was a slight tendency for the guessing rate to be higher in the CNO than the CN + IC condition, but in neither case did intrusions decrease across the 10 output positions. If anything, there was an increase in intrusions across output positions. Combining across the item cuing dimension, the mean number of intrusions for the 10 output positions was .33, .38, .41, .40, .37, .39, .35, .41, .39, and .50. Thus there is no evidence that output interference is due to subjects becoming less willing to respond during the cued recall test, although the interpretation of these data may be hampered by a floor effect. However, Smith (1971, Table 3) presented similar evidence, and further evidence contrary to this "tired subject" hypothesis is found in the results of Experiment 3.

Experiment 2

The results of Experiment 1 do not support the interpretation that it is the number of items recalled rather than the number of categories accessed that is responsible for output interference in the recall of categorized lists. However, the test in the previous experiment was based on the assumption that subjects' study of list items presented as cues mimics processes involved in recall of words. In Experiment 2 the number of items recalled before recall of three target categories was manipulated while the total list length and the number of prior categories recalled was kept constant. Subjects studied lists containing 12 categories. Four categories were represented by nine words, 4 by six words, and 4 others by three words. In a first recall test subjects were cued by the category names for categories containing either three or nine items. The number of items recalled was expected to be greater in the case in which the names were of nine-item categories rather than only three-item categories. In a second test subjects were then given the names of the categories in which there were six items. If the number of prior items recalled contributes to output interference, one would expect recall of the six-item categories to be worse following recall of the nine-item categories than following recall of the three-item categories. The number of prior categories accessed was held constant at four in both cases. In a control condition subjects studied the same lists but rested during the initial recall period before being tested on the six-item categories. In two other conditions subjects were given the category names of the Size 3 or Size 9 categories, but were not required to recall words from the categories. In fact, they were told that they could forget words from those categories. In these conditions subjects were exposed to the category names and presumably accessed the cate-

gories, but they did not recall words from the categories.

Method

Subjects. The subjects were 150 Purdue University undergraduates who served in partial fulfillment of a course requirement.

Materials. Thirty-six categories were selected from the Battig and Montague (1969) norms and assigned to three groups of 12, with the only restriction being that similar categories (e.g., fruits and vegetables) not be assigned to the same group. Nine words were selected from each category, though the first five listed in the norms were always excluded to reduce chance hits from guessing. Three lists were made up from each group of 12 categories to counterbalance particular categories against category size. Thus in one list 4 of the 12 categories of a group contained three words, in another list the same 4 categories were represented by six words, and in a third list they were represented by nine words. When fewer than nine words were used in a category, they were selected randomly from the set of nine. Thus nine lists were constructed from the three sets of categories. Subjects in the five conditions received lists based on the three sets in the same order, but the size of each category varied for a different subset of subjects so that each category was represented equally often in the different category size conditions.

Design. Five different conditions were varied between subjects. Thirty subjects were presented with and tested on three lists in the same condition. In all conditions there was an initial 5-min period after list presentation during which the activity of the subjects varied. Then there was a second 5-min period during which subjects in all conditions were given the category names from the four categories in which there were six items per category in the presentation list. Conditions differed only with respect to the activity that occurred during the initial 5-min period. In a control condition subjects simply rested during the initial 5-min period before they were given the cued recall test for the Size 6 categories. Two other groups of subjects were given cued recall tests during the first 5 min for the four categories of Size 3 or Size 9 (R-3 and R-9 conditions, respectively). Finally, two other groups of subjects were given the four category names of the Size 3 or Size 9 categories, just as were subjects in the R-3 and R-9 conditions, but they were instructed to try to forget the words in these categories (F-3 and F-9 conditions, respectively).

Procedure. Subjects were instructed at the outset that they would be given lists of words on which their memories were to be tested. The categorized nature of the lists was explained, but no mention was made of the varying size of the categories. Subjects were told after the lists were presented that they would be asked to turn to the first page of their recall booklets, read carefully the instruction at the top of the page, and follow it to the best of their abilities. The list presentation was the same as in Experiment 1. After the list was presented, subjects were instructed to turn to the first page of their recall booklets and follow the directions at the top of the page. In the control condition the instruction was simply to rest for 5 min. In the other four conditions, four category names were listed at the top of the page. In the R-3 and R-9 conditions, subjects were instructed to "recall as many items as you possibly can from the categories listed below, and only those categories, in any order you want." In the F-3 and F-9 conditions, subjects were told that they would never be asked to recall words from those categories. They were further told to "try to *forget* the information in these categories, so you can better concentrate on recalling the remaining information . . . try to erase these categories from your mind."

After the initial 5-min period, subjects were instructed to turn to the next page of their booklets and follow the instruction at the top of the page. The instruction was to recall as many words as possible in any order from the four categories whose names were listed below. The category names during the second period were always those of Size 6 categories.

Following presentation and recall of the first list, subjects were presented and tested with two more lists in the same condition. Subjects were tested in groups of from 5 to 15, and the experimental session lasted just under 1 hr.

Table 1
Mean Recall During the Two 5-Min Periods of Experiment 2

	First 5 min			Second 5 min		
Condition	R_w	R_c	$R_{w/c}$	R_w	R_c	$R_{w/c}$
Control				14.12	3.94	3.58
R-3	8.87	3.83	2.31	13.30	3.92	3.39
R-9	19.69	3.99	4.94	13.38	3.88	3.44
F-3				13.93	3.93	3.54
F-9				14.06	3.96	3.55

Note. R_w = recall of words; R_c = recall of categories; $R_{w/c}$ = recall of words per category. During the second 5-min period, all subjects recalled the four categories with six items.

Results and Discussion

The results for the five conditions of Experiment 2 are presented in Table 1 for the first and second recall periods. Besides the number of words recalled in each condition (R_w), this quantity was divided into the components that have been identified as important in other studies of cate-

gorized list recall (e.g., Cohen, 1966; Tulving & Pearlstone, 1966). These components are the recall of categories (R_c), where category recall is assumed to have occurred when a subject recalls at least one word from a category, and within-category recall ($R_{w/c}$), or the number of words recalled per recalled category. It has been assumed that when category names are given to subjects as retrieval cues, category accessibility is provided. Thus the main quantity of interest in the present results is the effect of the activities during the first 5 min on recall of words (R_w) and words within categories ($R_{w/c}$) in recall of the Size 6 categories.

Before considering recall of the Size 6 categories, let us first turn to recall during the initial test for the R-3 and R-9 conditions. By having subjects recall categories of either Size 3 or 9 during this test, it was hoped that the number of words recalled would be quite different, whereas the number of categories accessed would be the same. That this was so can be seen in the left-hand columns in Table 1. Subjects recalling the categories in which nine words were presented recalled slightly over twice as many words as subjects who recalled categories of Size 3. The recall of categories (indexed by recall of at least one word from the category) was quite high in both cases. Category recall was similarly high during the second 5-min period, justifying the assumption that presentation of the category name provides access to the category.

Of primary interest is recall of words within categories and all words during the second 5-min period as a function of the activities that occurred during the first period. Recall of words within categories ($R_{w/c}$) was somewhat greater in the control condition in which subjects had recalled no material prior to recall of the Size 6 categories than in the R-3 and R-9 conditions (3.58 words per category vs. 3.39 and 3.44), which is the usual finding of slight output interference in recall of categorized lists. However, there was clearly no difference between the R-3 and R-9 conditions favoring the R-3 condition in recall of the Size 6 categories; in fact, the slight difference was in the opposite direction. These same trends were observed in recall of words (R_w): Control subjects recalled an average of 14.12 words; R-3 subjects, 13.30; and R-9 subjects, 13.38. Since the number of prior words recalled was twice as great in the R-9 condition as in the R-3 condition, but the recall of words from the Size 6 categories did not differ, it would seem that it is not the number of words recalled that determines output interference.

The results of the F-3 and F-9 conditions indicate that the manipulation of telling subjects to forget categories had no effect on cued recall of the Size 6 categories relative to the control condition. Mean recall of words within categories was 3.58 in the control condition and 3.54 and 3.55 in the F-3 and F-9 conditions, respectively. There was also little difference among the conditions in recall of all words from the Size 6 categories. This null result is unlike that reported by Epstein (1972), who found that telling subjects to exclude part of an input list produced facilitation in recall of the remainder of the list relative to recall in a control condition. However, the present result is in line with the findings of Roediger and Tulving (1979). In three experiments they found that telling subjects to forget or exclude some categories from recall did not enhance noncued recall of items from the remaining categories relative to free recall; in fact, recall of these items was slightly but reliably impaired. The impairment was in recall of categories rather than items within categories in the Roediger and Tulving experiments. Presumably a similar inhibiting effect in the F-3 and F-9 conditions of Experiment 2 did not occur because subjects were provided with category names of the to-be-remembered words.

The evidence from Experiments 1 and 2 is consistent in showing that output interference in categorized list recall is not increased with increases in the number of words subjects have recalled or have been given as cues prior to recall of other

categories. Apparently, the number of prior categories recalled is the important factor determining output interference.

Experiment 3

The between-category output interference of the kind examined in the previous experiments has thus far eluded satisfactory theoretical analysis (Roediger, 1973, 1974). One approach to the issue of a proper account of such interference is to relate it to other types of interference (proactive and retroactive interference) that are better understood, at least with regard to which independent variables are important in controlling such interference. One of the most powerful variables affecting interference phenomena generally is similarity of the learned material. One typically finds greater proactive and retroactive interference with similar material than dissimilar material, though this is, of course, a complicated issue even in interference paradigms in paired-associate learning (e.g., Martin, 1965; Osgood, 1949). In Experiment 3 we attempted to vary the similarity of categorized lists to see if output interference in this situation is affected by the semantic similarity of the categories in the list. We were expecting that lists containing words from categories that were similar in that they could all be subsumed under a superordinate category heading would produce greater output interference than lists of words from relatively unrelated categories.

Method

Subjects. The subjects were 120 Purdue undergraduates enrolled in an introductory psychology course who participated in the experiment as partial fulfillment of a course requirement.

Materials. Twenty-four categories were chosen from the Battig and Montague (1969) and Shapiro and Palermo (1970) category norms to fit the constraints imposed by four different list structures. Two related lists were constructed by choosing categories that could be subsumed under a superordinate. The superordinate and their subordinate category headings for the two related lists follow. Animals: jungle animals, insects, birds, rodents, reptiles, and fish; names: female first names, male first names, novelists, poets, composers, and biblical names. Two unrelated lists were constructed by choosing categories that could not be grouped under a superordinate. One unrelated list contained words from the following categories: musical instruments, diseases, earth formations, kitchen utensils, sports, and metals. The other unrelated list included words from occupations, beverages, vehicles, weapons, kinds of cloth, and fruits. For each category in each list, eight exemplars were chosen with the constraint that they were unambiguous category members and were not of the four most frequent items given in response to the category name.

Design. Each subject was tested on all four lists. Thirty subjects heard and recalled each list under one of four list orders, which allowed lists to be completely counterbalanced against the order of test. Output position of a particular category during cued recall was counterbalanced within each list to ensure that each category appeared equally often in each test position. Five subjects served in each recall order within each list order. Order of presentation of categories and items within categories in each list was the same for all subjects.

Procedure. Subjects were tested in four groups of 30, with each group receiving a particular list order. Conditions within a list order were filled by random assignment of subjects.

Each subject was provided with a recall booklet with a category name printed at the top of each page. Blank pages separated the test pages of successive lists. After they were given their recall booklets, the subjects were fully instructed as to the nature of the task and the procedure of the experiment. They were told that the experiment involved the learning of common English nouns, and that the nouns were grouped at presentation into categories to which they belonged. The subjects were told that they did not have to remember the category names, since they would be given during the test, but only the items belonging to the categories. They were also told that after each list was presented a category name not on the list would be read to them. They were instructed to produce items that were members of the category until they were told to turn the page. When they turned the page to the next category heading, however, the subjects were told to produce only those items that had been presented in the list.

The instructions were followed by the taped presentation of a list. The manner and rate of presentation of the list were similar to those in the previous experiments. After the list was presented, the subjects were instructed, "Now, on the top page of your booklet, write down as many names of_____as you can." They were given 30 sec to perform this distractor task, which was intended to minimize recall from primary memory. A different category was used for the distractor task after each list presentation. The categories were chosen to be similar to one another, but to be unrelated to the categories in the list they followed. The four distractor categories were rivers, states, countries, and cities.

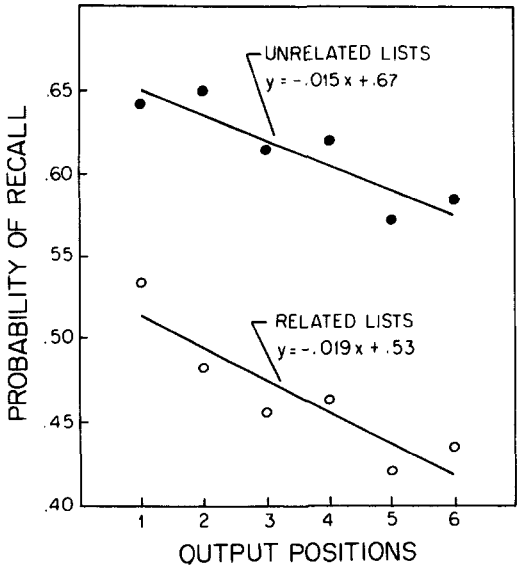

Figure 3. Probability of recall of items within categories for the related and unrelated lists in Experiment 3. (Lines represent the best-fitting linear functions.)

After the distractor task, the recording instructed the subjects to "turn to the next page and begin category recall." A category name appeared at the top of this page, and subjects were given 45 sec to recall the items that were presented as belonging to the category. They were then instructed to turn to the next page on which another category name was presented. After a list was recalled in this category-by-category fashion, a 2-min rest was allowed. The list presentation, distractor task, and recall phase for the next list then ensued. The third and fourth lists were presented and recalled in a similar fashion. Presentation and recall of the four lists lasted approximately 45 min.

Results and Discussion

The mean proportion of items recalled within categories across the six output positions for each list type is presented in Figure 3. The lines represent the best-fitting linear functions as determined by the least squares method. These functions account for 78% of the variance in both cases. There was a reliable effect of list type, $F(1, 116) = 439.57$, $p < .001$, $MS_e = .036$, with the unrelated lists better recalled than the related lists, and a reliable effect of output position, $F(5, 580) = 14.85$, $p < .001$, $MS_e = .037$. However, there was little evidence for the predicted interaction between list relatedness and output position, $F(5, 580) = 1.60$, $.10 < p < .25$, $MS_e = .035$. For the unrelated lists the slope of the best-fitting linear function was $-.015$, whereas the comparable value for the related lists was $-.019$. There is little evidence that the semantic similarity of the categories greatly affects output interference between categories, at least within the limits of the present experiment. It may well be, of course, that other manipulations of similarity could affect output interference.

It is not possible to interpret the superiority in recall of unrelated to related word lists, since the categories and words may have differed on many dimensions in addition to relatedness. The words of the related category lists may have been more difficult to recall for reasons unrelated to their appearance in a list with similar words. However, the differing levels of recall permit another examination of whether the number of words recalled affects output interference across categories. If this were the case, we would expect to find more output interference for unrelated than related lists, since more items were recalled from the unrelated lists. There was no evidence for greater output interference in the unrelated lists, in agreement with the conclusion from the prior experiments that the number of words recalled does not affect output interference.

Output interference was also examined across the four list positions in Experiment 3. (Lists were counterbalanced against the order of their test across subjects.) There was learning to learn across the four lists, with the proportion of words recalled within a category increasing across the four successive lists. The values were .49, .55, .55, and .57. If output interference were merely due to subjects becoming tired during the experimental session, then one would expect greater output interference for lists tested later in the session than earlier. There was no support for such a tired subject hypothesis in the present results, as output interference was about as great in the first two lists tested

as in the last two. Combining the results from Lists 1 and 2 and those from Lists 3 and 4, the slopes of the best-fitting lines were −.017 in both cases.

Experiment 4

The purpose of Experiment 4 was to examine output interference in the recall of paired-associate lists. The first experiments to examine output interference employed paired-associate lists with 10 or fewer digit–word pairs (Tulving & Arbuckle, 1963, 1966). The conclusion from these experiments, as mentioned in the introduction, was that output interference only occurred for the last few pairs in the list. It seems puzzling that output interference is such a robust phenomenon in categorized list recall at all input positions (Smith, 1973) and yet only occurs for the last few serial positions in paired-associate recall. However, one difference between the two situations is that the categorized lists employed were usually long, whereas the paired-associate lists that have been used were quite short. In Experiment 4 output interference was examined in recall of relatively long (20-pair) paired-associate lists in which subjects were presented word–word pairs.

Method

Experiment 4 can be considered as two experiments, 4a and 4b, that were identical in all respects except for the subject populations. In Experiment 4a the subjects were 120 introductory psychology students at Purdue who served for partial credit toward a course requirement. The subjects in Experiment 4b were 100 University of Toronto students who were paid for their service. Subjects were tested in groups in both experiments. The results reported here came from a series of experiments that had an entirely different purpose and that is reported elsewhere (Roediger & Adelson, in press, Experiments 1 and 2). The results are based on two lists regarded as practice with respect to the main purpose of the experiments. However, the subjects did not know that the lists were considered practice, and they thought they were simply the first two of three lists they would see and be tested on. Only the parts of the experiments relevant to present concerns will be discussed.

The lists contained 20 pairs of words. One word in each pair, presented in capital letters, was the to-be-remembered or target word. The other word in each pair, the context word, was presented in lowercase above the target word. Most context words were related to the target words in some way (many were synonyms), but some were pairs of relatively unrelated words. Subjects were told that they would see word pairs with one word in uppercase and the other in lowercase letters. They were told that they would be tested for recall of the capitalized words. However, they were also told to pay attention to the words in lowercase and the relation between the two words, since this would later help them to recall the target words.

The word pairs in all three lists were presented visually by means of a slide projector at the rate of 3 sec per pair. After presentation of the first list, subjects were instructed to turn to the first page of their recall booklets and to cover it with a blank sheet that was attached to it. They were then informed that this was a paced recall task in which they would be given 10 sec to recall the target word corresponding to each recall cue. The recall cues provided were the context words presented with each target word. Subjects were instructed to work only on the appropriate recall cue for a given 10-sec period and not to go forward or back to other cues. Each successive cue on the recall sheet was then uncovered by the subject at a signal from the experimenter. At the end of the paced recall period, subjects were given an extra 3 min to try to recall additional words to any of the cues. They were instructed to write any such additional responses to the right of the blank provided for their paced recall responses. This procedure was repeated for the second list. The order of presentation of word pairs was the same for all subjects for both lists. However, the order of the recall cues for each list was governed by 20 different random sequences distributed with equal frequency across subjects. After the second

Table 2
Proportion of Target Words Recalled in the Two Lists in Experiments 4a and 4b as a Function of Output Position

	Output position (blocks of four)				
List	1	2	3	4	5
	Experiment 4a				
1	.87	.82	.79	.75	.74
2	.86	.84	.82	.77	.77
	Experiment 4b				
1	.83	.85	.80	.76	.74
2	.85	.82	.80	.77	.70
M	.85	.83	.80	.76	.73

Note. Blocks of four output positions are grouped together. *M* is for Experiments 4a and 4b.

list was recalled, subjects saw and recalled a third list that is not of concern in the present context.

Results and Discussion

The results of the two parts of Experiment 4 are presented in Table 2. A strict criterion was used in the results reported. Subjects were given credit for recalling a target word only if it was recalled next to the appropriate cue during the initial cued recall test. However, the conclusions would not be changed by using a more lenient criterion, since subjects rarely recalled words to a cue that was not the context word at input or during the extra 3 min provided after the initial recall test. The results in Table 2 are proportions of words recalled in the different output positions, with the output positions being blocked in groups of four. In Experiment 4a each proportion is based on 960 observations; in Experiment 4b the proportions are based on 800 observations.

For both experiments and both lists within experiments, the same trend is apparent. There is a clear decline in recall across successive blocks of output positions. Observation of this output interference in two experiments and with both lists within the two experiments indicates that the phenomenon is reliable.

The output interference observed in these experiments cannot be attributed simply to an elimination of recency for two reasons. First, there was an instruction of approximately 15-sec duration between presentation of each list and its recall. This should have minimized any recency component. More importantly, even if the initial block of four tests is excluded, there is a mean decline of 10% across Blocks 2–5 when it can be safely asserted that short-term store or recency factors were absent. (The first four tests would have taken some 40 sec in addition to the instructions, pushing the total delay before recall in the second block of tests up to about 1 min.) Thus the results of Experiment 4 may be taken as the first evidence for output interference in recall of paired associates from secondary memory. Output interference may have been found here, but not in the experiments of Tulving and Arbuckle (1963, 1966) because the amount of material to be learned was greater in the present case. However, other differences between the two experiments, such as the use of word–word pairs in Experiment 4 rather than the digit–word pairs used by Tulving and Arbuckle (1963), cannot be ruled out.

General Discussion

The primary findings of the present experiments were that output interference between categories did not increase (a) when some of the items from the categories were presented as cues along with the category names, (b) when subjects recalled large rather than small categories prior to recall of categories of the same size, and (c) when the categories in the list were semantically related rather than unrelated. It was also shown that the output interference was not attributable to a loss of information from particular serial positions within categories, but rather that there was a drop in recall at all within-category serial positions for later output positions of the categories. Finally, output interference was observed in Experiment 4 in recall of a paired-associate list.

The results of Experiments 1, 2, and 3 are consistent in showing that the number of items recalled from a category has little effect on output interference. It seems likely that the number of prior categories recalled is the critical variable. However, two reservations must be made with regard to this statement. First, in Experiment 2 there was no evidence that simply presenting category names (with an instruction to forget items in those categories) produced output interference. Assuming that reading the category names provided access to the categories, it would then seem necessary that at least some items from the categories be recalled for the output interference to occur. But it seems that the number of items recalled from the categories is irrelevant, at least within the limits of the present experiments.

A second difficulty is that there are reports of failures to find output interference between categories even when the number of categories recalled has been varied. Tulving and Pearlstone (1966) compared subjects' recall of categorized lists under conditions of either free or cued recall. In the latter case recall was cued by the category names. Since subjects recalled words from more categories under cued than under free recall conditions, the number of words recalled per category should have been lower in the cued recall condition due to between-category output interference. However, this was not found; although subjects recalled words from more different categories under cued recall conditions, there was no effect on recall of words within categories (Tulving & Pearlstone, 1966, Table 3). A similar difficulty appears in an experiment reported by Roediger (1978, Experiment 2). It is unclear as to why recall of words within categories did not decrease in these cases with increases in the number of categories recalled, but it suggests that the factors producing output interference between categories may be more complicated than now suspected.

There are now three different types of evidence that the act of recall serves as a source of forgetting. Two of these types of evidence are direct, one is indirect. All have been studied in recall of categorized lists, but presumably each will generalize to other situations. One type of recall interference occurs for recall of categories or higher order units (Parker & Warren, 1974; Roediger, 1978). Following presentation of a categorized list, subjects in Roediger's (1978) experiments were given some of the category names from the list and were told to recall words both from the categories whose names were given, as well as from the other categories in the list. These subjects recalled fewer words from the noncued categories than did free recall subjects when they were scored on the same words. Recall of the cued categories interfered with recall on the noncued categories. This recall interference increased with increases in the number of category names presented and was found in the number of categories represented in recall (R_c) rather than the number of words recalled per category ($R_{w/c}$). Further, the inhibition in recall of the noncued categories did not seem due to the simple delay in recall from recalling the cued categories first.

A second source of indirect evidence for recall interference comes from research in the part-list cuing paradigm invented by Slamecka (1968, 1969) and used by many others. The general finding, as in Experiment 1 of the present series, is that presentation of part-list cues in addition to a category name inhibits recall of other words from the category as compared to recall with only a category name as the cue. It has been argued that when subjects study part-list cues, the representation of the cue words is strengthened, much as if they had been recalled, and that the cue words compete at retrieval with the traces of words not recalled (Roediger, 1974). It has also been shown that the amount of inhibition increases with the number of cues presented (Roediger, 1973; Rundus, 1973; Watkins, 1975), and that the phenomenon occurs with unrelated as well as categorically related lists (e.g., Slamecka, 1969; Roediger, Stellon, & Tulving, 1977) as well as in other situations (Mueller & Watkins, 1977).

Both types of recall interference just described can be accounted for by a hypothesis first put forward by Brown (1968) and elaborated by others, most notably Rundus's (1973) modification of Shiffrin's (1970) theory. Briefly, it is assumed that information about categorized lists is stored in a hierarchical arrangement with individual elements (words) stored below a higher level control element (Estes, 1972) that can be represented by a category name. The category representations in turn may be thought to be controlled by a yet higher level "list" representation that indicates their membership in the same list. The nested elements vary in strength of relation to their control elements, which determines their order of recall. However, the act of recall is assumed to increase

the strength of the recalled element to its control element. The final crucial assumption is that the retrieval process underlying recall involves sampling with replacement such that when a word or category has been recalled, it is not eliminated from consideration on future retrieval attempts. In fact, the more items that have been recalled, the more they will compete with traces of items not yet retrieved and the less likely will be retrieval of this material. This mechanism can thus account for the inhibition in recall of noncued material produced when subjects are given either some category name cues or list items to serve as cues. In both cases the representations of the relevant elements are strengthened and compete successfully at retrieval with the unrecalled information.

On the other hand, the Rundus (1973) model makes no provision for the third sort of recall interference, the sort examined in the present experiments. Why should recall of items from earlier categories depress recall of items within categories recalled later when subjects receive the names of the categories as recall cues? There is no answer to this question within the Rundus model. The present experiments and those of Smith (1971) seem to rule out two rather unexciting interpretations. One is that subjects simply change their criterion as the experiment or recall test proceeds. However, there was no evidence from Experiment 3 that the output interference increased across the testing of four lists, and the number of intrusions across the tested categories within a list did not decrease in Experiment 1 (see also Smith, 1971). Thus the decline in recall does not seem due to fatigue or some other type of criterion change. A second interpretation that seems unlikely is that the act of recalling early categories simply serves as a demanding intervening task and produces forgetting much as would mental arithmetic or any other verbal interpolated activity. Smith's (1971, Experiment 3) results seem to indicate that this is not so. The interference seems produced by recall and is not mimicked when mental arithmetic is used as a distractor activity.

What theoretical commodity might be responsible for the decline in within-category recall across the tested categories? Smith (1971) distinguished between core and extracore information within higher order units. It is assumed that some of the words in a category, the core words, become strongly integrated within the higher order units and that presentation of the category name allows recall of these words. Other words are more loosely linked to the category control element, and their recall is believed to depend on other sorts of spatial or temporal retrieval cues. From within a generation/recognition framework, Smith (1971) argued that subjects become increasingly unable to recognize the words across the successive test positions, although they can be generated. Unfortunately, there is little or no direct evidence supporting these assumptions and some of the evidence from Smith's (1971) experiments seems inconsistent with this account. For example, in his Experiment 2 in which recognition memory was tested for seven categories, recognition did decline across test positions, but only for the first three categories tested. However, in the experiments in which recall was tested with the same lists, there was a continuing decline across all seven tested categories. Roediger (1973) proposed that access to the extracore items was diminished due to recall interfering with the fragile spatial/temporal retrieval cues, but there is a similar lack of evidence for this proposal or even for the distinction between core and extracore processes. Developing a satisfactory theory of between-category output interference remains a challenge.

References

Arbuckle, T. Y. Differential retention of individual paired associates within an RTT "learning" trial. *Journal of Experimental Psychology*, 1967, *74*, 443–451.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, *80*(3, Pt. 2).

Brown, J. Reciprocal facilitation and impairment in free recall. *Psychonomic Science*, 1968, *10*, 41–42.

Carey, S. T., & Lockhart, R. S. Encoding differences in recognition and recall. *Memory & Cognition*, 1973, *1*, 297–300.

Cohen, B. H. Some-or-none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 182–187.

Dong, T. Cued partial recall of categorized words. *Journal of Experimental Psychology*, 1972, *93*, 123–129.

Epstein, W. Mechanisms of directed forgetting. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6). New York: Academic Press, 1972.

Estes, W. K. An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: V. H. Winston, 1972.

Martin, E. Transfer of verbal paired associates. *Psychological Review*, 1965, *72*, 327–343.

Mueller, C. W., & Watkins, M. J. Inhibition from part-set cueing: A cue-overload interpretation. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 699–710.

Osgood, C. E. The similarity paradox in human learning: A resolution. *Psychological Review*, 1949, *56*, 132–143.

Parker, R. E., & Warren, L. Partial category cuing: The accessibility of categories. *Journal of Experimental Psychology*, 1974, *102*, 1123–1125.

Roediger, H. L. Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 644–657.

Roediger, H. L. Inhibiting effects of recall. *Memory & Cognition*, 1974, *2*, 261–269.

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

Roediger, H. L., & Adelson, B. Semantic specificity in cued recall. *Memory & Cognition*, in press.

Roediger, H. L., Stellon, C., & Tulving, E. Inhibition from part-list cues and rate of recall. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, *3*, 174–188.

Roediger, H. L., & Tulving, E. Exclusion of material from recall as a postretrieval operation. *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 601–615.

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

Shapiro, S. I., & Palermo, D. S. Conceptual organization and class membership: Normative data for representatives of 100 categories. *Psychonomic Monograph Supplement*, 1970, *3*, 107–127.

Shiffrin, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Slamecka, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology*, 1968, *76*, 504–513.

Slamecka, N. J. Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology*, 1969, *81*, 557–560.

Slamecka, N. J. The question of associative growth in the learning of categorized materials. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 324–332.

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

Smith, A. D. Input order and output interference in organized recall. *Journal of Experimental Psychology*, 1973, *100*, 147–150.

Smith, A. D., D'Agostino, P. R., & Reid, L. Output interference in long-term memory. *Canadian Journal of Psychology*, 1970, *24*, 85–87.

Tulving, E., & Arbuckle, T. Y. Sources of intratrial interference in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1963, *1*, 321–334.

Tulving, E., & Arbuckle, T. Y. Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, 1966, *72*, 145–150.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 381–391.

Watkins, M. J. Inhibition in recall with extralist "cues." *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 294–303.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, *72*, 89–104.

Received June 27, 1979
Revision received August 21, 1979 ■