JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR **18**, 601–615 (1979)

Exclusion of Learned Material from Recall as a Postretrieval Operation

HENRY L. ROEDIGER III

Purdue University

AND

ENDEL TULVING

University of Toronto

Three experiments are reported in which, following presentation of a categorized list, subjects either recalled the whole list or a part of the list other than words in specified categories or those beginning with certain letters. Instructions to exclude part of the list from recall did not improve recall of the remainder, regardless of whether the exclusion instructions were based on the semantic dimension (category names) or the graphemic dimension (initial letters). The results suggest that it is difficult to retrieve selectively parts of a studied list from memory when recall instructions specify only what not to recall. This fact can be interpreted by assuming that retrieval is a consequence of matching information from two sources, the memory trace and cues in the retrieval environment.

Some 10 years ago Slamecka (1968, 1969) published experimental results that continue to puzzle researchers in memory. In his experiments subjects were presented with a list of words and were then tested for recall under either free recall or part-list cued recall conditions. In the latter, some of the list words were given to subjects as retrieval cues and their task was to recall the remaining target items. The number of target words recalled in the part-list cueing condition was compared with the number of these same items recalled under noncued conditions. Slamecka's finding, corroborated by others (e.g., Roediger, 1973; Rundus, 1973), was that recall of the remainder of the list is unaffected or even impaired by part-list cues in comparison with free recall of control subjects.

These results contradict certain theories of free recall learning that assume that the study of list words results in the creation of associations among list items, either within groups of subjectively organized words (Tulving, 1964) or in the form of a rich and complex network (e.g., Anderson, 1972). If retrieval were mediated by these associations, then presenting some words from the list would facilitate recall of the remaining words. The results of Slamecka and others using the part-list cuing paradigm are especially puzzling since it has been shown in other situations that words from a list can be useful cues for aiding recall (e.g., Hudson & Austin, 1970; Tulving & Pearlstone, 1966). Why are list words effective cues in some situations, yet not in the typical part-list cuing situation?

An important characteristic of the part-list cuing paradigm is that the cues specify two sets of words: one to be and the other not to be recalled. The subjects' task is to recall only the complement of the set of cue words (the target set) and to refrain from recalling the

This article was written while the first author was on leave at the University of Toronto and the second author was visiting Oxford University. This research was supported by Grant A8632 from the National Research Council of Canada to Endel Tulving. We thank Olga Watkins, Nora Keane, and Beth Adelson for their experimental assistance. Requests for reprints should be sent to Henry L. Roediger III, Department of Psychology, Purdue University, West Lafayette, Ind. 47907.

022-5371/79/050601$02.00/0

words in the cue set. If the necessity to suppress recall of the cue words has an inhibiting influence, this may cancel out any positive effect of the cues, so that the net effect may be no different from or slightly worse than free recall of the target words. Anderson (1972, pp. 359–361) and Roediger, Stellon, and Tulving (1977) have evaluated the role of having to suppress recall of the cue set in making part-list cues ineffective and have concluded that this is not the only source of difficulty in utilizing such cues. The research described here began as yet another attempt to determine the effects of having to suppress recall of some words on recall of other words presented at about the same time, although the results seem to have implications that go beyond the part-list cuing phenomenon.

The same general procedure characterized the first two experiments reported here. Subjects studied a list of words which could be categorized by both meaning and initial letter. At recall, subjects were instructed not to recall words from certain semantic categories or beginning with certain letters. These exclusion instructions would be expected to lead to different recall strategies depending on the nature of the categories to be omitted from recall. When words are presented blocked by semantic category it may be assumed that conceptual categories are the salient encoding dimensions subjects use, while the classification in terms of initial letters would not be noticed by subjects. Lauer and Battig (1972) and Lauer (1974) have shown that following presentation of words grouped by semantic category, category names are effective retrieval cues while initial letters are not, relative to free recall. Since category names serve as effective cues while initial letters do not in this situation, we refer to semantic categories as encoded features (or an encoded dimension) and initial letters as unencoded features (or an unencoded dimension). The distinction between encoded and unencoded features corresponds roughly to Underwood's (1969) distinction between retrieval and discriminative attributes. Of course, by referring to initial letters as unencoded features we do not mean to imply that they are not perceived when a word is studied; they are referred to as unencoded since there is no evidence under the conditions of our experiment that their representation in the memory trace is of sufficient salience to aid retrieval.

When subjects are instructed to exclude words beginning with certain letters, an unencoded dimension, they should retrieve much as do free recall subjects, but they should also have to check each retrieved word against the exclusion criteria before recalling it. Thus such subjects may be forced to use a strategy that subjects in the part-list cuing experiments of Slamecka (1968, 1969) and others may also use. In both cases subjects may implicitly retrieve every accessible word, but then only recall it on the basis of a post-retrieval classification. On the other hand, when subjects are instructed to exclude information on an encoded dimension (category names), they should not have to retrieve each word and check it against the exclusion criteria, but instead may exclude whole classes of responses that would never be retrieved. Thus exclusion instructions on encoded dimensions should speed and aid recall of target words relative to similar instructions to exclude on the basis of an unencoded dimension, or relative to performance of free recall subjects on the same target words.

The basic logic of the experiments can be illustrated with a hypothetical example referred to as the cross-out experiment. Imagine that two groups of subjects study the same list of words under identical conditions and write down as many as they can recall, drawing a line at 1-minute intervals. Then the experimenter tells one group of subjects that he actually wanted them to recall only half the words in the list, those beginning with the letters N through Z, and that they should go back through all the words they wrote down and cross out ones beginning with the letters

A–M. The experimenter scores their recall, which now consists only of the words not crossed out, and compares the recall performance of the two groups. Assuming that words beginning with the letters A through M constitute a random half of the list, we would expect the cross-out group to recall exactly half as many words as the other group in each interval of the recall period. Alternatively, if both groups of subjects were scored on the same words—those beginning with N through Z—the recall performance of the two groups would be identical. Obviously, we could not expect the cross-out group either to recall a larger proportion of words than the complete recall group or to recall the words in the smaller target set at a faster rate, because the exclusion instruction was specified *after* recall. A postrecall cross-out strategy clearly cannot benefit or penalize the subjects in any way.

The question of interest concerns the effect of exclusion instructions given *before* recall. The hypothetical cross-out experiment suggests a method of evaluating the effect of exclusion instructions given prior to recall. If subjects were told to exclude half the words from a list, facilitation would be reflected by their recalling more than half as many words as control subjects asked to recall the entire list. Alternatively, if the exclusion instruction enhanced recall, subjects told to exclude half the list should recall the remaining half better than control subjects scored on the same target words.

The part-list cuing experiments of Slamecka (1968, 1969) and others can be seen as cases where an exclusion instruction provided before recall does not enhance recall of other items in the list. For example, Roediger et al. (1977, Experiment 1) presented subjects with lists of 48 unrelated words and then provided 0, 16, or 32 words as cues at recall for the remaining words. In the case where subjects are given *n* words as cues, they have to exclude the *n* words from recall. As in other part-list cuing experiments, Roediger et al. (1977) found that subjects provided with 16 or 32 list items as cues were unable to recall more of the remaining words or recall them faster than free recall subjects scored on the same targets.

With the wisdom of hindsight, it is not too surprising that the instructions to exclude from recall a randomly chosen set from a list of unrelated words does not enhance the rate of recall of the remaining words, since it is difficult to imagine what recall method they could use that is more efficient than a postretrieval cross-out strategy. They may retrieve every accessible word, but then exclude it from recall if it is a member of the cue set. A more efficient method might be used if the structure of the study list permits specification in terms of encoded subsets of words. If we tell subjects they need not recall a whole class of words, they might be able to avoid retrieving them and concentrate instead on retrieval of the target categories.

In the first two experiments reported here subjects studied a 64-word list under conditions where we assume they encoded words into eight conceptual categories, but did not encode the words in terms of their initial letters. The eight words in each category began with eight different letters and the same initial letters were used across all categories. Thus the list can be conceived as the orthogonal crossing of eight conceptual categories and eight first letters. After presentation subjects were asked to recall words identified by either semantic categories or beginning letters. It was assumed that the semantic exclusion instruction tapped an encoded dimension, while the graphemic exclusion instruction did not. We expected that semantic exclusion instructions would aid recall of target words relative to free recall, but that the graphemic instructions would not be similarly effective. Subjects told to exclude encoded information should perform better than the hypothetical subjects in the cross-out experiment, while recall of subjects told to exclude unencoded information should approximate that of the cross-out subjects. Thus the purpose of this experiment was to show that Slamecka's (1968) results were due in part to the difficulty in effectively using an exclusion instruction with unrelated words. We thought there would be little or no exclusion difficulty if subjects were allowed to exclude groups of related words, since these are assumed to be stored and retrieved together.

Experiment 1

Method

Subjects. The subjects were 72 Purdue undergraduates who participated to fulfill a course requirement. They were tested in small groups.

Materials. Two categorized lists were constructed that contained eight categories and eight items per category. The 16 categories used in the experiment were selected from the Battig and Montague (1969) category norms. The primary restriction was that the eight items selected from each category must begin with the letters *a*, *b*, *c*, *g*, *l*, *p*, *s*, or *t*. Thus, the 64 items in each of the two lists can be thought of as the orthogonal crossing of the eight categories of the list with the eight initial letters. For example, the items selected from the category *Vegetables* were *asparagus*, *broccoli*, *cucumber*, *garlic*, *lettuce*, *parsley*, *squash*, and *turnip*. When more than one item beginning with one letter was available from a particular category, the item chosen was one of medium frequency to minimize chance hits due to guessing that might have occurred from using high-frequency instances (carrot, pea), while at the same time avoiding somewhat exotic low-frequency instances (collards, parsnips). The arrangement of items within a category was randomly determined and subjects were not told anything about initial letters.

Design. Three groups of subjects heard the same two lists under the same presentation conditions. Subjects recalled the lists under one of three experimental conditions determined by the instructions and the nature of the cues on their recall sheets. In one group subjects received no cues and were instructed to recall as many items as possible in any order (free recall or FR). In the second group subjects were given four category names from the list and were told *not* to recall items from these four categories; that is, they were to recall only items from the other four categories for which names were not provided. Since these subjects could forget or exclude the cued categories, this condition is referred to as "Exclude—4 Categories" or Ex-4C. The other group of subjects received four letters on their recall sheets and were instructed to recall only items other than those beginning with the letters on their sheets (Ex-4L).

Procedure. All subjects were given instructions at the beginning of the experiment explaining its general nature. They were told about the categorized nature of the lists that they were to remember, but they were not told the number of items in each category or the fact that all items in the categories contained the same eight letters. They were simply told that they would be asked to recall as much of the list as possible later, after they had performed an interpolated task and read a brief instruction at the top of their recall sheet telling them how to recall the list. The subjects were also informed that they would be asked to draw a line every 90 seconds during the recall period.

Items in the lists were presented via tape recorder at a 2-second rate and were blocked as to category, with items from each category preceded by the category name. Category names were presented in cadence with category instances but were distinguished by intonation. Half the subjects in each condition heard one list first, while the other half heard the other list first. Following presentation of each list all subjects were engaged in an interpolated task for 3 minutes. The task was recall of presidents of the United States or recall of the 50 states.

After the period used for the interpolated task, subjects were told to turn to their recall sheets and read the instructions at the top. The instructions at the top of all sheets were of roughly equal length. Free recall subjects were told to recall as many items as possible in any order. Subjects in the Ex-4C and Ex-4L conditions were told to recall as many items as possible from the list "except for items in the four categories (beginning with the four letters) listed below. You may try to forget items from these categories (beginning with these letters) and just concentrate on recalling other items from the list."

Each subject was tested with two lists in succession under the same set of instructions. Subjects did not know the specific recall instructions until they turned to their recall sheets after the first list was presented. When subjects were given instructions in preparation for presentation of the second list, they were not told whether or not they would be tested in the same manner as on the first list. In fact, all subjects were tested in the same manner on the two lists.

There were 24 subjects in each of the three conditions. Four categories or letters were randomly selected anew from the pool of eight for each subject in the Ex-4C and Ex-4L conditions. The recall period was 7.5 minutes with subjects instructed to draw a line under the last word they had written every 1.5 minutes.

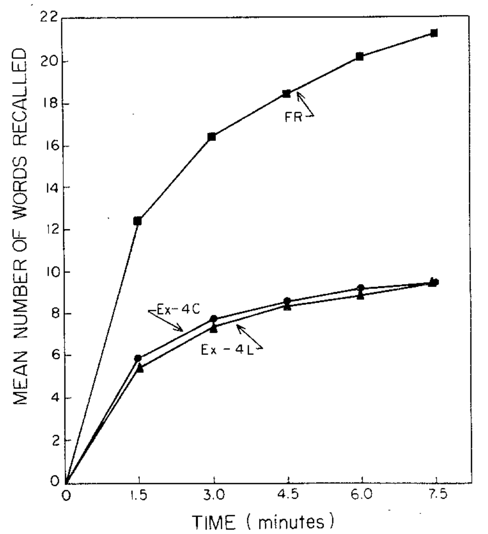

FIG. 1. Mean cumulative recall in the three conditions of Experiment 1.

Results and Discussion

The same pattern of data was obtained in recall of both lists, hence the results were combined. Presented in Figure 1 are the mean cumulative recall curves for the three groups of subjects in Experiment 1. Subjects given both Ex-4C and Ex-4L instructions recalled many fewer words than did free recall subjects, which is to be expected since these subjects were told to recall only half as many words as were FR subjects. The important finding shown in Figure 1 is the similarity between recall of Ex-4C subjects and that of Ex-4L subjects. Exclusion of list words from recall on the basis of an encoded dimension (category names) led to the same level of recall as did exclusion of words on the basis of an unencoded dimension (initial letters). This similarity in recall between the Ex-4C and Ex-4L conditions occurs despite the fact that the target items in the two conditions were not identical, though they were drawn from overlapping sets.

Table 1 shows the ratio of words recalled in the Ex-4C and Ex-4L conditions relative to the free recall condition for all recall intervals. The "expected ratios" listed in Table 1 are the ratios of words recalled in the exclusion conditions to recall that in the free recall condition would be expected if subjects were behaving as in the cross-out experiment. Facilitation in recall from the exclusion instructions would be reflected by ratios higher than the expected ratio. As can be seen in Table 1, subjects in both the Ex-4C and Ex-4L conditions recalled slightly fewer than half as many words as free recall subjects, the proportions ranging from .44 to .47. If subjects were behaving as in the cross-out experiment, the expected ratio for these conditions would be .50. We can conclude, therefore, that exclu-

TABLE 1

THE RATIO OF WORDS RECALLED BY SUBJECTS PROVIDED WITH AN EXCLUSION INSTRUCTION TO THOSE RECALLED BY OTHER SUBJECTS UNDER FREE RECALL CONDITIONS AS A FUNCTION OF TIME[a]

			Time (min)									
Experiment	Condition	Expected ratio	0.5	1.0	1.5	2.0	2.5	3.0	4.5	6.0	7.5	10.0
1	Ex-4L	.50			.44			.45	.45	.44	.44	
	Ex-4C	.50			.47			.47	.46	.44	.44	
2	EX-2L	.75			.64			.70	.71	.72		
	Ex-2C	.75			.69			.71	.68	.67		
	Ex-6L	.25			.22			.23	.22	.22		
	Ex-6C	.25			.26			.24	.23	.23		
3	Fr-30	.50	.99	.97	.84	.74	.72	.69	.64	.62		.60
	Ex-6C	.50	.66	.56	.53	.49	.47	.47	.46	.45		.45

[a]If subjects given the exclusion instruction were using the postrecall cross-out strategy, the expected ratio would be obtained.

sion instructions do not permit subjects to retrieve selectively stored list items, even if whole subsets of list items are to be excluded from overt recall on the basis of encoded dimensions that are known to provide an effective access route under cued recall conditions (e.g., Tulving & Pearlstone, 1966).

The same conclusion follows from an analysis in which free recall subjects were scored on the same set of items as Ex-4C and Ex-4L subjects were to recall. This analysis involved rescoring the FR subjects' protocols twice against two different sets of "critical words", the words that the Ex-4C and Ex-4L subjects were not asked to exclude from recall. The results of this analysis are presented in the two panels of Figure 2. Recall of critical words for FR and Ex-4L groups is shown at the top, that for FR and Ex-4C groups on the bottom. These results confirm the earlier findings: Permitting subjects to exclude half the words from the list did not facilitate recall of the remaining set of words, relative to recall of the same words under free recall conditions. In fact, subjects instructed to exclude information recalled reliably fewer words than did free recall subjects. Subjects in the Ex-4C condition recalled a mean of 9.40 critical words (standard deviation of 4.09)

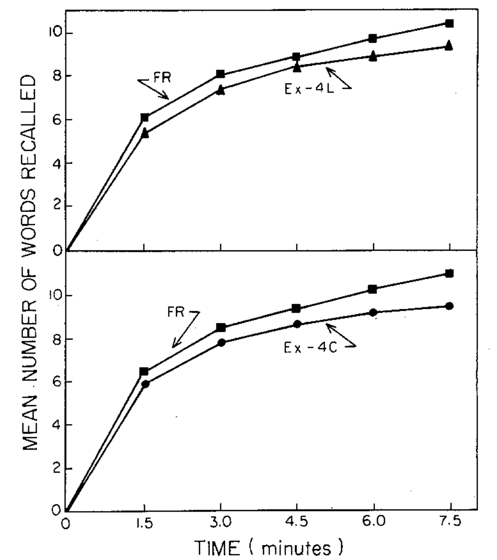

FIG. 2. Mean cumulative recall of critical words for subjects given a letter exclusion instruction (top panel) or a category name exclusion instruction (bottom panel) compared to free recall of the same words in Experiment 1.

while FR subjects recalled 10.77 (2.97), $t(22) = 2.49$, $p < .05$. The Ex-4L subjects recalled 9.40 critical words (3.24) while FR subjects recalled 10.35 (2.50), $t(22) = 1.90$, $p < .10$. (All confidence levels reported in this paper are based on two-tailed tests.)

Thus, rather than facilitating recall relative to free recall, exclusion instructions inhibited recall when provided with either initial letters or categories.

Subjects in the free recall and Ex-4C conditions were also scored on recall of categories, where a subject is given credit for recalling a category when at least one word from the category has been recalled (Cohen, 1966). If cumulative recall of categories were presented, the functions would parallel those for recall of words in Figures 1 and 2. At the end of the recall period FR subjects had recalled words from 5.9 of the 8 categories, while Ex-4C subjects had recalled words from 2.8 categories. When scored on the same critical categories as were to be recalled by Ex-4C subjects, FR subjects recalled 3.0. Recall of words within recalled categories was about the same for the two conditions, 3.62 for FR subjects and 3.46 for Ex-4C subjects. Thus the exclusion instruction, even when provided on an encoded dimension, did not facilitate recall of either the remaining list categories or the words within those categories, relative to free recall.

Experiment 2

The results of the Ex-4C subjects in the first experiment were unanticipated. It seemed perfectly reasonable to expect that when relieved of recalling certain well-encoded subsets of list words, Ex-4C subjects would recall the remaining target words more efficiently than either free recall subjects who were required to recall the entire list or Ex-4L subjects who were told to exclude information on an unencoded dimension. There was no evidence even pointing in the direction of this initial expectation. Because the results were unexpected, it seemed worthwhile to confirm the major finding before considering possible theoretical implications. In the second experiment reported here subjects were presented with the same lists and told, in the critical conditions, to exclude words beginning with two or six of the eight letters or to exclude words from two or six of the eight categories. (Actually Experiment 2 was conducted prior to Experiment 1, which served as the replication. They are reported in the reverse order for ease of exposition.)

Method

The method was quite similar to that of Experiment 1. Sixty graduate and undergraduate Yale University students were paid for their participation. Subjects were tested in small groups under one of five experimental conditions. All subjects heard the same two lists used in Experiment 1 under the same presentation conditions and, after performing an interpolated task involving retrieval from semantic memory, they were told to turn to their recall sheets. A brief instruction at the top of the recall sheet defined the different conditions. Free recall subjects were instructed to recall as many list words as possible. The Ex-2C and Ex-6C subjects were instructed to recall as many words as possible except for those in the two or six categories for which names were listed. Similarly, Ex-2L and Ex-6L subjects were instructed to recall words except for those beginning with two or six letters.

For each subject in the Ex-6C and Ex-6L conditions six categories or six letters were randomly chosen to be excluded. Then for each subject in the Ex-2C and Ex-2L conditions two of the six categories or six letters for one of the Ex-6C or Ex-6L subjects were randomly selected to be excluded. Thus the 16 target words for Ex-6C and Ex-6L subjects were embedded in the larger target sets of Ex-2C and Ex-2L subjects. All subjects in the semantic and graphemic exclusion conditions could then be scored on the same set of critical words, as could free recall subjects.

Subjects were instructed to draw a line under the last word recalled every 1.5 minutes during the 6-minute recall period. As in the previously reported experiment, subjects in the exclusion conditions were informed only about the set of items they were to exclude; they were not cued as to the categories or initial letters of the to-be-recalled set.

Results and Discussion

Results from the two lists were combined since the same pattern of data was obtained with both lists. Presented in Table 1 are the ratios of recall in the four experimental conditions to that in the free recall control at each timed interval. As in the first experiment, subjects in the four exclusion conditions recalled fewer items than expected on the basis of the postretrieval cross-out hypothesis. That is, Ex-2C and Ex-2L subjects did not recall the expected 75%, while Ex-6C and Ex-6L subjects recalled less than the expected 25%. Once again, whether the exclusion instruction was provided on the basis of an encoded dimension (category names) or an unencoded dimension (initial letters) seemed to be immaterial.

Presented in Figure 3 is the mean cumulative recall of the 16 critical words for both the letter and category exclusion instructions (top and bottom panels, respectively). Critical words were those that constituted the target set for all subjects. Recall of critical items was somewhat lower in all four exclusion conditions than under free recall conditions, although not reliably so in the case of category exclusion instructions. The mean critical item recall (with standard deviations in parentheses) was 5.38 (2.06), 5.04 (2.49), and 5.17 (2.49) for FR, Ex-2C, and Ex-6C conditions, respectively. No pair of means differed reliably, $t < 1$ in each case. Mean critical item recall in the FR, Ex-2L, and Ex-6L conditions was 6.17 (1.95), 5.08 (1.61), and 5.04 (1.81), respectively. The superiority of FR subjects to subjects given the graphemic exclusion instruction was of marginal reliability, $t(10) = 1.91$ for FR versus Ex-2L and $t(10) = 1.95$ for FR versus Ex-6L, $p < .10$ in both cases. The difference between Ex-2L and Ex-6L subjects was not reliable, $t < 1$.

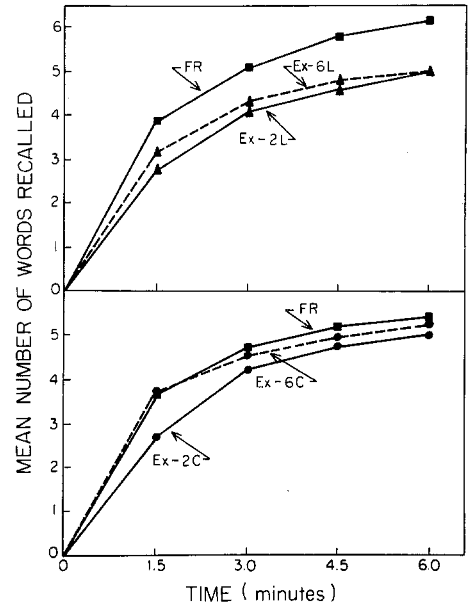

FIG. 3. Mean cumulative recall of critical words for subjects given a letter exclusion instruction (top panel) or a category name exclusion instruction (bottom panel) compared to free recall of the same words in Experiment 2.

The results of Experiment 2 closely replicated those of Experiment 1. Subjects told to exclude a set of words from a recently presented list were unable to recall the remaining target words more efficiently than free recall subjects required to recall the entire list, regardless of whether the exclusion instruction was given on the basis of an encoded dimension (conceptual category) or unencoded dimension (initial letter). Results of this experiment also showed that rather wide variations in the proportion of information to be excluded did not seem to affect either the rate or asymptotic level of critical item recall.

EXPERIMENT 3

The results of Experiments 1 and 2 are interesting when they are compared with the results of experiments examining the effects of list length on recall where we know that the reduction of the size of a target set facilitates the probability of recalling items in the set (e.g., Murdock, 1960; Roberts, 1972). For example, if we presented one group of subjects (the "Long" group) with a long categorized list and a second group of subjects (the "Short" group) with a list composed of half the categories presented to the first, the Short group should recall a greater proportion of words from their list than subjects in the Long group would recall from theirs. Alternatively, the Short group should recall more words than the Long group if the Long group were scored only on the set of words included in the short list. Subjects receiving the short list would, presumably, have less difficulty "searching for" the target set since they would not have to exclude any list items from recall. It is also likely that if recall were measured cumulatively, subjects receiving the short list would approach asympote at a greater rate than subjects receiving the long list (Bousfield & Sedgewick, 1944; Indow & Togano, 1970).

In Experiment 3 we explicitly tested these notions. In addition to the two groups (Long and Short) already described, a third group was given the long list and then told at recall to exclude half the words in the list. Thus their target set was identical to that of subjects given the short list. We were interested in whether subjects, when given short categorized lists, would recall a greater proportion of words from these lists (and approach asymptotic levels of recall at a greater rate) than they would when given the long lists. Assuming this to be the case, when told after presentation of the long list to exclude half of it on the basis of an encoded dimension would subjects be able to recall the list as if the to-be-excluded information had not been presented? Of course, the results of the prior experiments had led us to expect that the answer to this last question would be no: Telling subjects, in effect, to try to shorten a list after it was presented did not alter recall to resemble that of short lists. This experiment allowed a direct comparison of recall of subjects given a short list to recall when these same subjects were trying to "shorten" the list on the basis of an instruction given at recall.

Method

Subjects. Thirty-six University of Toronto undergraduate and graduate students participated for pay.

Materials. Thirty-six categories, and five exemplars per category, were selected from the Battig and Montague (1969) norms. The categories were assigned at random to six sets of six. Thus each set contained 30 words.

Design. All subjects studied three lists and were tested under one of three conditions after each list. The conditions were (a) 60 words presented (12 categories) with subjects asked to free-recall the entire list, (b) 30 words presented (6 categories) with subjects asked to free-recall the list, and (c) 60 words (12 categories) presented with subjects told to exclude six categories and to recall words from the remaining set of (unspecified) categories containing 30 words. These conditions are referred to as FR-60, FR-30, and Ex-6C, respectively.

The sets of categories were allocated to lists in such a way that every set was used equally often in each condition. Thus each set appeared once as a 30-word list, once as the six to-be-excluded categories in an Ex-6C list, once as the words to be recalled in an Ex-6C list, and twice as part of a 60-item list for free recall. In lists in the Ex-6C condition the to-be-excluded categories were always the 2nd, 4th, 5th, 7th, 10th, and 11th categories in the list.

Procedure. The subjects were tested in six groups of six. They were told that they would hear a list of words and that their task would be to recall the words in any order after an interpolated task at the end of the list. Words from each category were presented consecutively in the list, but items from a particular category were not preceded by category names. The words were read at the rate of one word every 2 seconds. After each list was presented subjects were distracted for 1 minute by the requirement to recall, as fast as possible, as many U.S. states, Canadian provinces, or names of rivers as they could. Subjects then turned to their recall sheet and read a brief instruction at the top. Depending upon the condition they were told to recall as many words as possible from the list they had just heard (FR-60 and FR-30 conditions), or to recall words except those from six specified categories. The exact instructions for Ex-6C subjects were: "Now please try to recall as many of the words you have just heard as you can. Write them down below, one under the other, in the order that they occur to you. However, *do not* write down any of the words from the following categories: (listed in a column). Please do not start writing until I tell you to." Subjects were allowed 10 minutes for recall and were instructed to draw a line under the last word recalled after specified intervals. The instruction to draw a line was given every 0.5 minute for the first 6 minutes and at 1-minute intervals thereafter.

The order of conditions was balanced among the six groups of subjects according to two Latin squares. Thus the order of conditions for each group of six subjects was unique.

Results and Discussion

Presented in Figure 4 are the cumulative recall curves for the three conditions of Experiment 3, combined across lists. After the first minute, recall was greatest in the FR-60 condition, at an intermediate level in the FR-30 condition, and poorest in the Ex-6C condition. The interesting observation is that at every point in time subjects tested under FR-30 conditions recalled more words than those tested under Ex-6C conditions, despite the fact that the target sets for the two conditions were identical. We can conclude therefore that the presence of other words in the list in the Ex-6C condition, words that were to be excluded in recall, interfered with the recall of the words in the indirectly specified target set. It is as if subjects cannot ignore

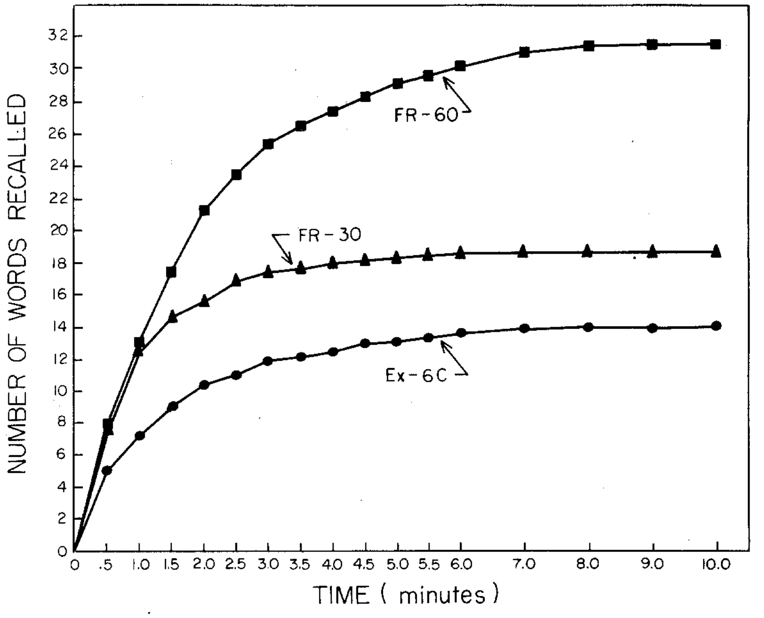

FIG. 4. Mean cumulative recall under the three conditions of Experiment 3.

the words they are instructed to exclude as they "search the store" for the to-be-recalled words.

The purpose of this experiment was to show that under conditions where selective retrieval from a small target set from a categorized list can occur (because the excluded set is never presented), recall would be more efficient than recall of items from a larger target set. That this is so can be seen from the first row of numbers presented in Table 1 for Experiment 3. Here are presented the ratios of recall under the FR-30 condition to that under the FR-60 condition at each timed interval. The ratios of .99 and .97 at the 30-second and 1-minute intervals indicate that initially under FR-30 conditions subjects recalled virtually the same number of words as FR-60 subjects, despite the fact that their target set was only half as large as that under FR-60 conditions (see also Figure 4). This proportion decreased steadily as the recall period increased, but even after 10 minutes when the cumulative recall curves had reached asymptote, recall under FR-30 conditions was 60% of that under FR-60 conditions, still well over half as great. Thus with the short lists subjects recalled proportionately more words than with long lists, and also reached asymptotic levels of recall at a greater rate.

The results from this experiment, if we assume that they generalize to other situations, have implications for the list length/difficulty relationship usually found in free recall. As the length of a list is increased, one typically finds that the number of words recalled increases but that the proportion of the list recalled decreases (Murdock, 1960; Roberts, 1972) when subjects are allowed a fixed amount of time per to-be-recalled item. The results in Figure 4 indicate that the correlation between recall and list length depends on the amount of time subjects are allowed for recall. With short recall periods subjects presented with short lists are likely to recall as many words as subjects given long lists.

The results in the Ex-6C conditions are slightly discrepant from those obtained in Experiments 1 and 2. In the previous experiments we found that subjects told not to recall a certain part of the list recalled slightly less than the ratio of remaining target items to free recall target items at every timed interval. On this basis we would expect that under Ex-6C conditions in Experiment 3 subjects should recall about 50% or slightly less of items recalled under FR-60 conditions. It can be seen in the Ex-6C row in Table 1 that the proportion at the 30-second interval was higher than expected, .66. The ratios at the 1.0- and 1.5-minute intervals are also higher than expected, but this comes about only because of the nonindependence of points in cumulative recall curves. That is, the ratios are higher at the second and third timed intervals only because of the relatively great recall during the first 30 seconds in the Ex-6C condition. Between the 0.5-minute interval and the 1.5-minute interval subjects in the Ex-6C condition recalled 3.9 words while FR-60 subjects recalled 9.2. Thus the ratio of Ex-6C recall to FR-60 recall was .42, below the .50 of the expected ratio.[1]

It is unclear as to why the 0.5-minute point is higher than expected. It seems to be reliable since we obtained the same outcome in a replication of Experiment 3. In the first two experiments the first timed interval was at 1.5 minutes, while in this experiment the discrepant point is at 0.5 minute. It may be that with written recall the free recall subjects simply cannot write down all retrieved items so that rather than Ex-6C performance being "too high" it may be that FR-60 recall is too low, constrained by writing time. It should be possible to examine this conjecture by testing subjects individually with oral recall

[1] Actually the expected ratio for recall after the first 0.5 minute should be .48 rather than .50, an adjustment which reflects the fact that the number of new items that could be recalled after 0.5 minute were no longer in a 2:1 ratio for the FR-60 and Ex-6C conditions. The obtained ratio of .42 is still smaller than the adjusted expected ratio of .48.

where time to produce overt responses is minimal.

Whatever the reason for the aberrant point in the Ex-6C condition early in the recall period, the primary conclusion from the prior experiments is not altered. After 0.5 minute subjects in the Ex-6C condition recalled fewer than half as many words as FR-60 subjects. By the end of the recall period, Ex-6C subjects had recalled 13.97 words (standard deviation of 5.70) while FR-60 subjects recalled 15.79 (4.73) when scored on the same set of critical items, $t(34)=3.31$, $p<.01$. Thus, as in Experiment 1, subjects given a semantic exclusion instruction showed inhibition in recall relative to free recall subjects who were responsible for recalling all 60 words. Subjects in the FR-30 condition also recalled more of the 30 critical target words (mean of 18.78, standard deviation of 5.10) than did either Ex-6C subjects, $t(34)=8.59$, $p<.001$, or FR-60 subjects, $t(34)=5.64$, $p<.001$. The FR-30 condition serves as a useful comparison to the Ex-6C condition, since it allows one to examine recall under a limiting case of selective retrieval when the "excluded information" is not even presented. If subjects in the Ex-6C condition could follow the direction (as given in Experiment 1) to actually forget the to-be-excluded information, recall should be identical to that in the FR-30 condition.

General Discussion

The experiments reported here have produced two primary results. First, it has been shown that exclusion instructions produce a temporal pattern of recall similar to that which would be produced by the hypothetical post-recall cross-out strategy. That is, telling subjects what not to recall did not aid recall of target items, relative to free recall of the same critical items. Second, the effects of the exclusion instructions were the same for the encoded and unencoded dimensions of list items. Subjects told to exclude information from conceptual categories recalled the target words no better than subjects told to exclude items on the basis of their initial letters. Any account of the first result should also fit the second, or at least not be inconsistent with it.

The exclusion conditions not only failed to produce facilitation in recall of critical items relative to free recall, but actually seemed to produce some inhibition. Although this inhibition was small and in some cases not statistically reliable, by pooling t values across experiments by the method suggested by Winer (1971, p. 50), the inhibition produced by both the semantic exclusion instructions and the graphemic exclusion instructions becomes statistically reliable, $z=3.08$ and $z=3.03$, respectively, $p=.002$ in both cases. It may be that the facilitation of recall under the exclusion conditions is masked by this inhibition. We leave this possibility open, but are skeptical of this hypothesis for several reasons. First, it is not at all clear why we should obtain the same amount of facilitation and inhibition with both the semantic and graphemic exclusion instructions. Second, by the usual mechanism that is invoked to account for inhibition in recall that is produced from providing part-list cues (Roediger, 1973, 1978; Rundus, 1973), one would expect inhibition from the semantic exclusion instruction but not with the graphemic exclusion instruction. Finally, one might also expect greater inhibition from presentation of six category names than from two, but such an outcome did not occur in Experiment 2.

Based on the results at the 0.5-minute interval in Experiment 3 one might also speculate that our temporal unit of analysis was too gross and that with smaller units—scoring recall, say, every 15 seconds—some advantage might have been observed for the semantic exclusion conditions. Again, we are willing to leave this possibility open, but the hypothesis seems unlikely because any initial advantage in the rate of recall in one condition over another must then have been cancelled out by a subsequent slower rate, since at 90 seconds and throughout the recall period thereafter the recall rates were the same in both the semantic and graphemic exclusion conditions and no faster than in the free recall control condition.

Assuming that the results cannot easily be interpreted in the ways we have just discussed, why do exclusion instructions not aid recall of other list items? The fact that graphemic exclusion instructions do not aid recall is not at all surprising. Since we know from the research of Lauer (1974; Lauer & Battig, 1972) that initial letters are not effective retrieval cues under the general conditions of our experiment, we have good reason to assume that initial letters were not encoded as features of the list items, or at least were not encoded as salient features useful for retrieval. If letter cues are not effective cues for retrieval, it seems unlikely that they would be useful for nonretrieval under the graphemic exclusion conditions of our experiments. List items under the graphemic exclusion condition can be excluded from overt recall only *after* they have already been retrieved by other means. This postretrieval exclusion strategy on the part of the subjects given the graphemic exclusion instructions would be like the strategy of the hypothetical subjects in the cross-out experiment, except that the "crossing out" would occur after subjects retrieved information but before it was overtly recalled.

But why should the results be the same as in the graphemic exclusion condition when the exclusion instructions are specified in terms of an encoded dimension that we know is useful for retrieval under the conditions of our experiment (that is, under the semantic exclusion conditions)? Why does an intelligent learner who understands the instructions (certainly in learning the second of two lists) "waste time" in retrieving a to-be-excluded category? Or why should one waste time in retrieving another word from a to-be-excluded category immediately after having retrieved one word and found it to be unacceptable in terms of a postretrieval decision based on the list of to-be-excluded categories? That subjects do "waste time" in this manner is strongly suggested by the data, indicating that the temporal pattern of recall is much the same in the exclusion and free recall control conditions. Thus it seems reasonable to assume that every time the subject in the control condition retrieves and overtly recalls a word, the subject in the exclusion condition in which half the items are excluded also retrieves a word but overtly recalls it only if the postretrieval decision is favorable. As we have said, this is probably what happens with the graphemic exclusion instruction. But since the data from the semantic exclusion conditions were quite similar to those in the graphemic conditions, we are led to the hypothesis that subjects given both types of exclusion instructions were engaged in the same type of process of eliminating items from overt recall following their retrieval. The slight inhibition in recall under the exclusion conditions may be due to this postretrieval process.

On the basis of our results, then, we are led to the assumption that retrieval of list words proceeds at the same rate in all three of the basic conditions (semantic exclusion, graphemic exclusion, and standard free recall) and that exclusion represents a postretrieval operation. This postretrieval operation might be regarded as analogous to the strategy of subjects in the postrecall cross-out experiment, except that exclusion subjects "cross out" or eliminate items after they have been brought into consciousness (retrieved) but before they have been overtly recalled. This exclusion process may slow recall slightly.

The hypothesis that the exclusion of learned material at recall on the basis of an encoded memory attribute is a postretrieval operation does not fit well with the customary ideas about organization and retrieval of word lists (e.g., Tulving, 1962, 1964; Mandler, 1967). It is generally assumed that related words, such as members of a conceptual category which cluster in free recall and can be readily recalled in the presence of an appropriate category name cue, are stored together and that their recall consists of a search through the memory store (e.g., Anderson, 1972). If both of these assumptions are made and if the search process is nonrandom and is guided by information in the retrieval environment, why should it operate on a region in the search space that is specified as to be excluded from recall? And why should the search operation persist in this unproductive mode, as the data suggest it sometimes does?

There may be no good answers to these questions, because the questions themselves may be the wrong ones to ask. They may be based on wrong assumptions about the structure of storage and the retrieval process. It may not be profitable to regard items as stored dependently (together) and to regard retrieval as a search through, or activation of, subregions of the storage space or the associative network. It may be more useful to think that list items are stored in the form of a collection of features (e.g., Bower, 1967; Underwood, 1969). Retrieval occurs if a critical subset of these features is matched or complemented by corresponding features of cue information in the retrieval environment (e.g., Flexser & Tulving, 1978; Kintsch, 1974; Pellegrino & Salzberg, 1975; Tulving, 1976). In cued recall the cue information is controlled to some extent by the experimenter; in free recall it is generated by the retrieval system, probably in part from semantic memory. By this view exclusion instructions, even when provided on an encoded semantic attribute, do not increase the rate of retrieval for the simple reason that they do not provide any useful cue information. The exclusion instructions do not provide any information that will complement or match the to-be-recalled information in the target set. Even when exclusion is based on encoded semantic features of to-be-remembered list words, recall of other list words in the target set must proceed on the same basis as in free recall. The result is that exclusion conditions yield retrieval rates much like those obtained in free recall for the same critical items. The question of whether traces of individual words are stored dependently (Anderson, 1972; Mandler, 1967; Tulving, 1962) or independently (Slamecka, 1968) makes little sense when retrieval is conceptualized as a consequence of a conjunction of information from two sources, the trace and the cue.

Finally, a comment about the relation between the experiments reported here and other research on directed forgetting (e.g., Bjork, 1972). In directed forgetting research subjects are typically presented with material and then told either to remember or forget part of it, with the effect of this instruction measured by a test of recall on the other part. For example, if subjects are presented with a list of words and halfway through it are given an instruction to forget the words already presented, they recall more of the words presented later than if no forget instruction is given (e.g., Waugh, 1972). What the present results indicate is that subjects cannot benefit from an instruction to forget some material when it is presented at recall, at least under the conditions of the present experiments. It seems likely that the only situation in which a postinput cue to forget will produce facilitation in recall of other material is when the exclusion or forget instruction specifies directly or indirectly what the subject is to recall (e.g., Epstein, 1972). If subjects studied items from only two categories and were told to exclude one, they would quite likely recall the other category more rapidly than would free recall control subjects because of the absence of output interference for the critical items.

References

Anderson, J. R. FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press, 1972. Vol. 5.

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*, 1969, **80** (3, Part 2).

BJORK, R. A. Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: V. H. Winston, 1972.

BOUSFIELD, W. A., & SEDGEWICK, C. H. W. An analysis of sequences of restricted associative responses. *Journal of General Psychology*, 1944, **30**, 149–165.

BOWER, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. New York: Academic Press, 1967. Vol. 1.

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

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

FLEXSER, A. T., & TULVING, E. Retrieval independence in recognition and recall. *Psychological Review*, 1978, **85**, 153–171.

HUDSON, R. L., & Austin, J. B. Effect of context and category name on the recall of categorized word lists. *Journal of Experimental Psychology*, 1970, **86**, 43–47.

INDOW, T., & TOGANO, K. On retrieving sequence from long term memory. *Psychological Review*, 1970, **77**, 317–331.

KINTSCH, W. *The representation of meaning in memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

LAUER, P. A. Encoding specificity in the cued and free recall of categorically and alphabetically organized words. *Bulletin of the Psychonomic Society*, 1974, **4**, 496–498.

LAUER, P. A., & BATTIG, W. F. Free recall of taxonomically and alphabetically organized words lists as a function of storage and retrieval cues. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 333–342.

MANDLER, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. New York: Academic Press, 1967. Vol. 1.

MURDOCK, B. B. The immediate retention of unrelated words. *Journal of Experimental Psychology*, 1960, **60**, 222–234.

PELLEGRINO, J. W., & SALZBERG, P. M. Encoding specificity in cued recall and context recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 261–270.

ROBERTS, W. A. Free recall of word lists varying in length and rate of presentation: A test of the total time hypothesis. *Journal of Experimental Psychology*, 1972, **92**, 365–372.

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. Recall as a self-limiting process. *Memory and Cognition*, 1978, **6**, 54–63.

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.

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

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.

TULVING, E. Subjective organization in the free recall of "unrelated" words. *Psychological Review*, 1962, **69**, 344–354.

TULVING, E. Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychological Review*, 1964, **71**, 219–237.

TULVING, E. Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition*. New York: Wiley, 1976.

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

UNDERWOOD, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

WAUGH, N. C. Retention as an active process. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 129–140.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

(Received October 5, 1978)