Reprinted from JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR
Volume 5, Number 4, August 1966

JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 5, 381-391 (1966)

Availability Versus Accessibility of Information in Memory for Words[1]

ENDEL TULVING AND ZENA PEARLSTONE

University of Toronto, Canada

The *S*s learned, on a single trial, lists of words belonging to explicitly designated conceptual categories. Lists varied in terms of length (12, 24, and 48 words) and number of words per category (1, 2, and 4). Immediate recall was tested either in presence or absence of category names as retrieval cues. Cued recall was higher than noncued recall, the difference varying directly with list length and inversely with number of items per category. This finding was interpreted as indicating that sufficiently intact memory traces of many words not recalled under the noncued recall conditions were available in the memory storage, but not accessible for retrieval. Further analysis of the data in terms of recall of categories and recall of words within recalled categories suggested two independent retrieval processes, one concerned with the accessibility of higher-order memory units, the other with accessibility of items within higher-order units.

If a person is shown a long list of familiar words and is then asked to recall the list, he can recall some words, but not all of them. It can be assumed that the person learns each single word at the time of its presentation, in the sense that the probability of recall of the word rises from a value near zero immediately before the presentation to a value near unity immediately after the presentation. The failure to recall some of the words, therefore, reflects intratrial forgetting (Tulving, 1964).

1 This research was supported by the National Science Foundation, under grant GB-810. It was prepared for publication during the senior author's tenure of a National Research Council of Canada Senior Research Fellowship at the Institute of Human Learning, University of California at Berkeley. We are grateful to the Director and the Board of Education of the Township of Scarborough, and to the Forest Hill Collegiate Institute and its Principal, who permitted us to test many students as subjects in their schools. We are also grateful to many teachers in the high schools in Scarborough and Forest Hill, who kindly made class time available for the experiment. Our special thanks goes to Dr. Howard Russell and Mr. Vernon Trott for their generous assistance.

Intratrial forgetting is a descriptive label that carries no implications as to the fate of the memory traces associated with nonrecalled words. It may be attributable to the decay of traces as a consequence of passage of time between the presentation and attempted recall of an item (Brown, 1958), or to the displacement of some of the items stored earlier by subsequently presented items (Waugh and Norman, 1965). In either case, failure to recall a certain item would be interpreted to mean that the trace of the item is no longer available in the memory storage at the time of recall. It is also possible, however, that intratrial forgetting represents a failure to "find" otherwise intact traces in the storage. According to an information processing model of memory described by Feigenbaum (1961), for instance, forgetting occurs not because information in storage is destroyed, but because learned material becomes "inaccessible in a large and growing association network." Thus, to interpret intratrial forgetting, it is useful to draw a distinction between what information or what traces are *available* in the memory storage and what are *accessible*. This distinction parallels the distinction between retention and recall, or the distinction between trace storage and trace utilization (Melton, 1963).

The present paper is concerned with a conceptual and experimental analysis of nonrecall of learned items in terms of such a distinction between availability and accessibility. It describes an experiment whose primary purpose was to explore the hypothesis that a substantial part of nonrecall of familiar words under typical experimental conditions is attributable to inaccessibility of otherwise intact memory traces.

Experimental demonstrations of the distinction between availability and accessibility of information require that critical experimental treatments be administered at the time of the recall test, rather than at some earlier stage in the sequence of events involved in any memory task. Only if conditions of the experiment are held constant until the beginning of the recall period can differences in observed recall scores be attributed to differences in accessibility. While scattered examples exist in the literature of expermients satisfying these requirements (e.g., Fox, Blick, and Bilodeau, 1964; Peterson and Peterson, 1962, Exp. IV), there have been no systematic attempts to distinguish between availability and accessibility of mnemonic information. Experiments in which various "measures of retention," such as unaided recall and recognition, have been compared (e.g., Luh, 1922; Postman and Rau, 1957) lend support to the proposition that unaided recall does not tap all of the information that is available about previously learned material, but the interpretation of data in these experiments with respect to the distinction between availability and accessibility is complicated. Unaided recall requires the *S* to reproduce the whole item, while in recognition the correct item is given to the *S* and his task is to decide whether or not it occurred in the list. To distinguish between availability and accessibility of information that is sufficient for *reproduction* of a given item, comparisons between recognition and recall are only partly relevant and other methods must be used.

The experiment described in this paper uses one such other method. Categorized word lists were presented to *S*s for learning, and recall of words was tested in the presence or absence of category names as retrieval cues. It was expected that a large proportion of words not accessible for recall under the unaided conditions would become accessible as a consequence of experimental presentation of such retrieval cues, thus indicating that sufficient information was available in the storage for the reproduction of these words, but that this information was not accessible. The results of the experiment thus were expected to clarify the nature of intratrial forgetting as defined earlier. As the results turned out, they also illuminated the retrieval processes involved in a memory task such as the one used in the experiment, and had several interesting implications for other types of experiment.

Method

Design

Categorized word lists, consisting of (a) *category names,* and (b) *words* representing instances of categories, were presented to *S*s once. Immediately after the presentation, two recall tests were given in succession. The *S*s were instructed to try to remember as many *words* as possible.

Three independent variables were manipulated: (a) list length—L (12, 24, and 48 words), (b) number of words or items per category—IPC (1, 2, and 4 words), and (c) conditions of recall in the first recall test—cued recall (CR) and noncued recall (NCR). The second recall test was always given under the conditions of CR.

All possible combinations of L and IPC were used to yield nine lists. Lists are designated in terms of the values of these two variables. For instance, List 24-2 refers to a 24-word list in which there are two items per each of 12 categories.

All combinations of nine lists and two conditions of recall in the first recall test were used to yield 18 experimental conditions. Experimental conditions are designated in terms of the list and recall condition. For instance, condition 12-4 CR refers to List 12-4 recalled under the conditions of cued recall. Thus, the design of the experiment was $3 \times 3 \times 2$ factorial. With respect to the first recall test the independent variables were L, IPC, and recall condition; with respect to the second recall test they were L, IPC, and recall condition of the first test. Since the second recall test was always given under identical conditions (CR), experimental groups can be uniquely defined in terms of list characteristics and recall condition of the first test. For instance, Group 48-1 NCR designates the sample of *S*s who learned List 48-1 and who were first tested under the conditions of noncued recall.

Subjects and Experimental Groups

The *S*s were high-school students of both sexes from Grades 10 to 12 from a number of different schools in two school systems in the Metropolitan Toronto area.

A total of 948 *S*s were tested in the experiment. Data from 19 *S*s had to be discarded because of incompleteness of recall protocols. The data discussed in this report are thus based on the records from 929 *S*s. The age of *S*s ranged from 14 to 21 years, with a great majority (94%) of *S*s being between 15 and 18 years of age.

The *S*s were tested in groups during a regular class period. Each of nine lists was learned by *S*s in four classes. Within each class, all *S*s were presented with identical material under identical conditions, but half the *S*s were tested first under the conditions of CR while the other half was tested first under the conditions of NCR. The second recall test of the material, as mentioned earlier, occurred under the conditions of CR for all *S*s.

The sizes of the 18 experimental groups, each composed of *S*s from four different school classes, ranged from 48 to 56.

Lists

A practice list, consisting of 24 common adjectives, was administered under the typical single-trial free-recall conditions to all *S*s prior to the presentation of the experimental list.

Two different sets of nine experimental lists were constructed with the aid of the Connecticut word associations norms (Cohen, Bousfield, and Whitmarsh, 1957) and with the aid of norms from a small pilot study patterned after the procedure used by Cohen *et al.* (1957). Two groups of *S*s under each of the 18 experimental conditions learned a list from the first set, while the other two groups learned a corresponding list from the second set.

Corresponding lists in the two sets contained identical categories but different words. The words in List 48-1 represented 48 different categories, 40 taken from the Connecticut norms and eight from the pilot study. Twenty-four categories were selected randomly for Lists 24-1 and 48-2. The 12 categories represented in Lists 12-1 and 48-4 in turn were selected randomly from those occurring in Lists 24-1 and 48-2, respectively. The same general procedure was followed in the selection of categories for other lists.

Words in a given category of a list in which IPC = 4 were, in the first set, usually the second, fourth, sixth, and eighth ranking words in the norms, and in the second set, the third, fifth, seventh, and ninth ranking words in the norms, but some deviations from this general rule occurred. Words for categories containing two items or one item were selected randomly from such sets of four words.

The order of categories in a list and the order of words within categories were determined randomly. All the words within a category occurred in immediately adjacent positions. The lists presented to *S*s thus consisted of a number of category names, each category name being followed by one, two, or four items appropriate to the category. For instance, List 12-2 in the first set was as follows: four-footed animals—*cow, rat;* weapons—*bomb, cannon;* crimes —*treason, theft;* forms of entertainment—*radio, music;* substances for flavoring food—*cinnamon, pepper;* professions—*engineer, lawyer.*

Procedure

The *S*s recorded their recall in specially prepared recall booklets that were distributed at the beginning of the experimental session. Instructions about *S*s' task, and about the use of recall booklets, as well as all lists were presented to *S*s by means of a high-fidelity tape-recorder. The *S*s were first informed that they were going to take "a test to find out how people remember words," and that although *E* was not interested in how well each of them did individually, they should do their best in the test. The standard free-recall instructions were then given for the practice list, followed by the presentation of the practice list, at the rate of 2 sec per word. Two min were given for recall.

The instructions for the second part of the test, the experimental list, informed *S*s that they would next hear and try to memorize a list of nouns, or "names of various things," pairs of nouns (in case of IPC = 2), or groups of four nouns (in case of IPC = 4), and that each word (or pair of words or group of four) would be "preceded by another word or phrase that describes the word (words) to be remembered, but which in itself does not have to be remembered." Next, an illustrative list of the kind that *S*s in a particular group had to learn was given as part of the instructions. This short list contained five categories (country in Europe, boy's name, city in U.S., name of a river, and statesman of our day), each category being accompanied by one, two, or four names, depending on the IPC of the experimental list. The illustrative list was read and the *S*s reminded that "we want you to remember only the word (words) that followed each descriptive phrase, or category." These words that *S*s had just heard, but not the category names, were then read again and referred to as the part of the list *S*s would have to remember. The *S*s were then told the number of words, number of categories, and number of words per category in the list they were going to learn.

Apart from the general instructions to recall as many words as possible, no information was given to *S*s exactly what the conditions of the recall test were going to be nor were they told that there would be different recall conditions for different *S*s in the same group.

The duration of presentation of the list varied for different lists according to the formula: $T = 3\ NoC + L$, where T is the total duration of presentation in seconds, NoC is the number of categories (L/IPC), and L is list length. The amount of time given for recall also varied for different lists, depending on L. The *S*s had 1, 2, or 4 min to recall lists of 12, 24, or 48 words, respectively.

For the condition of NCR, the recall booklets contained L consecutively numbered lines. For the condition of CR, the recall booklet listed all category names that had occurred in the list, in the same order as in the input list, and each category name was followed by one, two, or four lines, depending on IPC.

At the end of the first recall test of the experimental list, all *S*s recalled all the words they could remember a second time under the conditions of CR.

Results

The mean number of correctly recalled words on the practice list for the total sample of 929 *S*s was 9.48 ($SD = 2.27$). The breakdown of these recall scores in terms of the 18 experimental groups showed the means to range from 8.81 to 10.06. A one-way analysis of variance of these data yielded an $F(17, 911)$ of 2.53 which is unexplainably significant at the .01 level. Since the median correlation coefficient between practice and experimental list recall was only + .228 for the nine CR groups and + .284 for the nine NCR groups, possible differences in ability among the groups suggested by differences in practice-list scores probably had only a minor effect on the evaluation of the effects of experimental treatments.

Recall of Words

The first analysis of the data was concerned with the number of words recalled under various experimental conditions. The stability of these data was tested in the following manner. In each of the 18 experimental groups, the *S*s were randomly divided into two subgroups, the mean recall score on the first recall test computed for each subgroup, and an intraclass correlation coefficient (McNemar, 1962) between the 18 resulting pairs of means calculated. This coefficient turned out to be .997, indicating a high degree of stability of the mean recall scores for various experimental groups.

First Recall Test. Mean number of words recalled on the first recall test of the experimental lists is shown by filled (CR) and unfilled (NCR) circles in Fig. 1 as a function of L and IPC. An overall analysis of variance of the number of words recalled in the first recall test showed all three main effects and all three double interactions to be significant at better than the .001 level. The triple interaction among R, L, and IPC was not significant.

Recall of words was higher under the condition of cued recall than under the conditions of noncued recall for all nine lists. The smallest numerical difference between CR and NCR was found for List 12-4. This was not significant by *t*-test ($t = 1.88$), but all other differences were significant at better than the .01 level. As can be seen from Fig. 1, the superiority of CR over NCR was an increasing function of list length and a decreasing function of IPC. The largest difference (19.8 items, or 126%) was found for List 48-1.

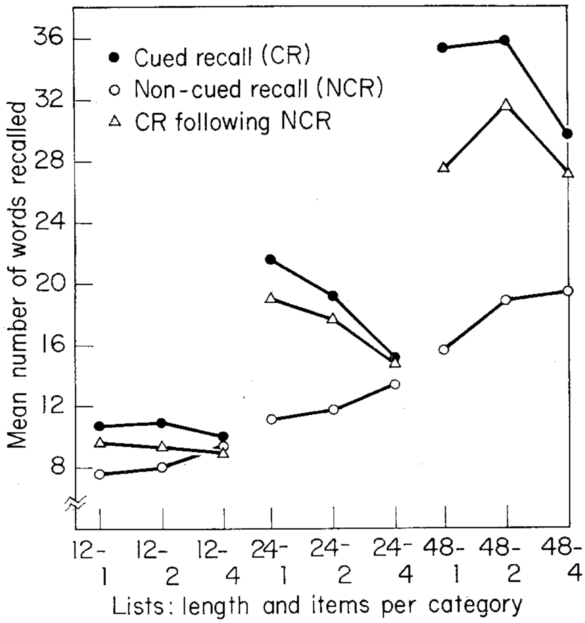

FIG. 1. Mean number of words recalled in the first recall test (circles) and the second recall test (triangles) as a function of list length and number of items per category.

When we consider CR and NCR separately, we find in Fig. 1 that NCR increases with IPC at all three levels of list length, but under the conditions of CR the effect of IPC depends on L. The inverse relation between CR and IPC is quite clear for the 24-word list and is also obvious when we compare recall for IPC = 1 with that for IPC = 4 for the other two lists, but there was no decrease in cued recall from IPC = 1 to IPC = 2 for the 12-word and 48-word lists. All six possible comparisons of mean recall scores between IPC = 1 and IPC = 4 yielded significant differences when tested by means of *t*-tests, five being significant at the .001 level and one (CR for the 12-word list) at the .05 level.

Second Recall Test. The second recall test was administered to all *S*s under the conditions of cued recall, where category names were available on recall sheets. For *S*s in all nine CR groups the mean number of words recalled on the second test was practically identical with the mean number of words recalled on the first test. The overall mean word-recall in all nine groups on the first test was 21.17, and on the second test 21.20. Thus there was neither any forgetting nor "reminiscence" from the first to the second test.

The mean recall scores on the second test for the NCR groups are shown by triangles in Fig. 1. These means were significantly higher than the means on the first test for all lists except List 12-4. But for none of the nine lists did the mean second test recall score in the NCR groups equal that of the CR groups, as can be seen in Fig. 1.

The second recall test was included in the design, and the data from the second test are included in this report, primarily in order to illustrate that subsequent presentation of category names as retrieval cues in the NCR groups would result in an increase in the number of retrieved words. More detailed analyses of these data, however, are not warranted, since no safe assumptions can be made about availability of information in the memory storage after different treatments in the first recall test. For this reason, data from the second recall test will be ignored in the rest of this paper.

Error Data and Guessing Bias. Errors of recall were classified into three categories: repetitions of list words, noncategorical intrusions, and categorical intrusions. Errors falling into the first two classes were few in number. On the first recall test, for instance, a total of 24 repetitions and a total of 73 noncategorical intrusions were found in all 929 recall protocols.

Categorical intrusions are extralist intrusions that are members of one of the categories used in a given list. Mean numbers of such intrusions are shown in Table 1. Three observations are of interest. First, the frequency of categorical intrusions tended to increase with IPC at all levels of L. Second, the frequency of intrusions increased with L at all levels of IPC. Third, the number of intrusions for a given list was always greater for CR than NCR, with the exception of List 12-4.

Since the frequency of categorical intrusions seems to be related to the treatment variables, "correct" recall scores may be somewhat inflated. The *Ss* may have received credit for recall even when they arrived at the correct word through a free association to the category name, remembered under the NCR conditions and explicitly given on recall sheets under the CR conditions. One might argue, therefore, that the differences in recall between the CR and NCR conditions might in fact be smaller than the data depicted in Fig. 1 indicate.

The extent of such possible bias can be roughly estimated by considering the probability of occurrence of our list-words as free associations in the norms of Cohen *et al.* For the nine lists, such mean probability varied over the range of .052 to .078, with an overall average of .065. To illustrate the small extent of the bias in recall scores attributable to guessing of words from given categories under the CR conditions, consider List 24-4 for which the difference between CR and NCR was the smallest of all lists. (We will ignore List 12-4 for which the difference was not significant.) The mean number of words recorded by *Ss* in Group 24-4 CR was 17.7, of which 15.1 were correct and 2.6 were categorical intrusions. The same mean for Group 24-4 NCR was 14.6, of which 13.4 were correct and 1.2 were intrusions. Thus, *Ss* in Group 24-4 CR put down, on the average, 3.1 more responses than *Ss* in Group 24-4 NCR. Since the average word in List 24-4 had occurred in the norms with the probability of .052, only .16 words of the 3.1 "extra" words given by Group 24-4 CR would be expected to match the list words and credited to correct recall. The actual difference in mean correct recall between Groups 24-4 CR and 24-4 NCR, however, was over ten times as large. It is clear, therefore, that even for the list showing the smallest difference between CR and NCR the difference could not be accounted for in terms of free associations to category names, and hence must be attributed to the facilitative effect of category names as retrieval cues.

TABLE 1

MEAN NUMBER OF CATEGORICAL INTRUSIONS IN THE FIRST RECALL TEST OF EXPERIMENTAL LISTS

List length (L)	First recall	Items per category (IPC)		
		1	2	4
12	CR	.66	.59	.77
	NCR	.21	.31	1.23
24	CR	.91	2.12	2.58
	NCR	.22	.92	1.23
48	CR	4.22	3.87	5.52
	NCR	1.59	1.29	2.92

Category and Words-Within-Category Recall

The analysis of word-recall data in an experiment such as the present one can be regarded as a first-level analysis only. It indicates the gross effects of the variables manipulated, but it does not provide much insight into the underlying relations. Some such insight, however, can be obtained from an analysis of the data in terms of two further response measures.

The first of these is referred to as category recall. This is defined in terms of the number of categories from which *at least one word* is recalled. The measure has been used earlier by Cohen (1963). We designate this measure as R_C. In lists where IPC $= 1$, R_C is identical with the number of words recalled (R_W), but in lists where IPC > 1, the two measures do not necessarily covary and usually yield different values.

The second measure is words-within-category recall, or words recalled per category recalled. It is defined in terms of the ratio of the number of words recalled to the number of categories recalled. This measure has been referred to as "mean word recall per category" by Cohen (1966). We designate this measure as $R_{W/C}$. In lists where IPC = 1, $R_{W/C}$ is always 1.00 by definition, given that *S* recalls at least one word from the list, but for higher levels of IPC, $R_{W/C}$ can assume all values between 1.00 and IPC.

The word-recall score (R_W) is a simple multiplicative function of category recall score (R_C) and words-within-category recall score ($R_{W/C}$), i.e., $R_W = R_C \cdot R_{W/C}$. The word-recall data that we considered in the two preceding sections thus reflected the effects of the independent variables on both of the two components of R_W. We will now examine the data from the first recall test with respect to the two components of R_W. Table 2 shows mean R_C scores for all experimental conditions. It can be seen that R_C varies systematically with all three independent variables. It is less under the NCR conditions than under the CR conditions for all lists, but the magnitude of this difference depends on both L and IPC. At a given level of IPC the difference is an increasing function of L, and at a given level of L it is a decreasing function of IPC. In Table 2, the values of R_C for lists in which IPC = 1 are in parentheses to remind the reader that they are identical with the corresponding R_W values.

The mean recall scores of words recalled per category recalled ($R_{W/C}$) are shown in Table 3. Again the scores for lists where IPC = 1 are included for the sake of completeness, although they are always unity by the definition of the $R_{W/C}$ measure.

Table 3 shows that while $R_{W/C}$ is systematically related to IPC, it seems to be independent of recall conditions and also independent of list length for lists of 24 and 48 words. When $R_{W/C}$ scores are averaged over all six lists for which IPC > 1, the overall means are identical at 2.32 for both CR and NCR. None of the differences in $R_{W/C}$ between CR and NCR for the six lists approaches significance by *t*-tests. And when the data are averaged over both recall conditions and IPC levels of 2 and 4, the mean $R_{W/C}$ for 24-word lists is 2.21 and the mean for the 48-word lists is 2.18.

Interpretation of Findings on Word-Recall.

To aid in the interpretation of some of the findings pertaining to word-recall (R_W), the data on R_W, R_C and $R_{W/C}$ are summarized in Table 4 as mean proportions of these measures relative to maximum possible scores. These proportional measures are designated as $P(R_W)$, $P(R_C)$, and $P(R_{W/C})$.

Values of $P(R_W)$ in the right-hand panel of Table 4 were obtained by dividing each of the mean R_W scores by its respective L, but these scores can also be arrived at by multiplying the corresponding $P(R_C)$ and

TABLE 2

MEAN NUMBER OF CATEGORIES RECALLED (R_C) IN THE FIRST RECALL TEST OF EXPERIMENTAL LISTS

List length (L)	Recall condition	Items per category (IPC)		
		1	2	4
12	CR	(10.70)	5.88	2.98
	NCR	(7.70)	4.42	2.92
24	CR	(21.70)	11.16	5.79
	NCR	(11.18)	6.84	4.87
48	CR	(35.35)	20.49	11.36
	NCR	(15.57)	10.52	7.29

TABLE 3

MEAN NUMBER OF WORDS RECALLED PER CATEGORY RECALLED ($R_{W/C}$) IN THE FIRST RECALL TEST OF EXPERIMENTAL LISTS

List length (L)	Recall condition	Items per category (IPC)		
		1	2	4
12	CR	(1.00)	1.86	3.35
	NCR	(1.00)	1.84	3.19
24	CR	(1.00)	1.73	2.61
	NCR	(1.00)	1.73	2.75
48	CR	(1.00)	1.75	2.61
	NCR	(1.00)	1.79	2.65

TABLE 4
RECALL OF CATEGORIES (R_C), OF WORDS PER CATEGORY RECALLED ($R_{W/C}$), AND OF WORDS (R_W)[a]

Recall condition	List length	R_C 1	R_C 2	R_C 4	$R_{W/C}$ 1	$R_{W/C}$ 2	$R_{W/C}$ 4	R_W 1	R_W 2	R_W 4
CR	12	.892	.980	.993	(1.00)	.930	.837	.892	.912	.832
	24	.904	.930	.965	(1.00)	.865	.652	.904	.804	.630
	48	.736	.854	.947	(1.00)	.873	.652	.736	.745	.617
NCR	12	.642	.737	.973	(1.00)	.920	.797	.642	.678	.776
	24	.466	.570	.812	(1.00)	.864	.687	.466	.492	.558
	48	.324	.438	.608	(1.00)	.893	.663	.324	.391	.403

Column heading: Number of items per category.

[a] Table entries are proportions based on group means relative to maximum scores possible.

$P(R_{W/C})$ scores given in the two left-hand panels in Table 4.

An important fact reflected in the data in Table 4 is that the relations between $P(R_C)$ and $P(R_{W/C})$ on the one hand, and IPC, on the other hand, are all monotonic, for all levels of L and for both conditions of recall, while the relation between $P(R_W)$ and IPC is not. $P(R_C)$ is always an increasing function of IPC, and $P(R_{W/C})$ is always a decreasing function of IPC, but the relation between $P(R_W)$ and IPC cannot be stated as simply. $P(R_W)$ is an increasing function of IPC under the conditions of NCR, while under the conditions of CR it is a decreasing function of IPC for the 24-word list and, taking the sample means in Table 4 literally, it increases from IPC = 1 to IPC = 2 and decreases from IPC = 2 to IPC = 4 for both the 12-word and 48-word lists.

The relations between $P(R_W)$ and IPC become somewhat more meaningful if we remember that any change in $P(R_W)$ depends on changes in both $P(R_C)$ and $P(R_{W/C})$. An increase in $P(R_W)$ as a function of IPC means that under certain conditions $P(R_C)$ increases at a faster rate as a function of IPC than $P(R_{W/C})$ decreases. Conversely, a decrease in $P(R_W)$ as a function of IPC means that under certain conditions $P(R_{W/C})$ decreases at a faster rate as a function of IPC than $P(R_C)$ increases.

These considerations suggest that the decrease of R_W as a function of IPC under the conditions of CR, shown in Fig. 1, is probably an artifact related to lists of limited length and to limited number of categories. The $P(R_C)$ score is already so high for lists in which IPC = 1 that there can be relatively little further improvement in this measure with higher levels of IPC. Even for List 48-1, $P(R_C)$ is so high (.736, as shown in Table 4) that the maximum possible increase of .264 in this measure which would bring it to unity would not be sufficient to outweigh the decrease in $P(R_{W/C})$ from 1.00 to .652 over the range of IPC values used in this experiment.

This "ceiling effect" on $P(R_C)$ and the role it plays in determining the relations between R_W and IPC at levels of L used in this experiment is made explicit as a result of the breakdown of the word-recall measure into its two components. Inspection of the R_W curves plotted against IPC would not readily lead to the conclusion that we are dealing with an artifactual limit imposed on the *S*s' recall performance, since the R_W curves have a negative slope.

The reversals in the R_W curves plotted as a function of IPC, for 12-word and 48-word lists under the cued recall conditions, can also be understood in terms of the two components of R_W, in an analogous fashion, and will not be elaborated further.

Order of Recall. Two further findings, having to do with order of recall under the conditions of NCR, where the order of recall was free to vary, will be briefly mentioned.

The first was the tendency for the words from a given category to be recalled together despite the absence of the experimentally presented category name. A measure reflecting this trend is provided by the proportion of times that a word from a given category was followed in recall by another word from the same category. This proportion varied between .92 and .95 in the three lists of IPC = 2, and between .89 and .97 in the three lists of IPC = 4.

The second finding concerned the order in which words were recalled within a given category. The general tendency was for the words to be recalled in the same order in which they appeared in the input list. As an illustration we only mention some data from List 48-2. In those cases where the *S*s recalled *both* words from a category the order of recall was the same as in the input list 78% of the time (311/397) and reverse to that in the input list 22% of the time (86/397). These data show that even in the case of the longest list used in this study the *S*s apparently retain a fair amount of information about the order in which two words from the same category occurred in the input list.

Discussion

The most important finding of this experiment was higher recall under the conditions of cued recall than under the conditions of noncued recall. Since the experimental treatment administered to the *S*s in the two recall conditions was the same, both the amount of information and the organization of this information in the memory storage at the beginning of the recall test must have been identical for the CR and NCR groups. The superiority of cued recall over noncued recall thus suggests that specific information about many words must be available in the storage, in a form sufficient for the reproduction of words, even when this information is not accessible under a given set of recall conditions.

Intratrial forgetting, defined in terms of nonrecall of words learned in the input phase of a trial, thus does not necessarily reflect the loss of relevant information from the storage, but only its inaccessibility. Accessibility of the information clearly depends on its availability, but it also depends on retrieval cues. While the present findings do not rule out the possibility that some information stored in memory in the course of presentation of a list decays over intratrial retention intervals or is erased by other incoming information, they do make clear that inferences about what is available in memory cannot be made on the basis of what is accessible.

Retrieval cues obviously constitute an extremely important factor in determining the level of recall. The presence of a single experimentally manipulated retrieval cue, the category name, resulted in large increments in the number of recalled words, particularly for longer lists. It is entirely within the realm of possibility that additional and more powerful retrieval cues would produce an even greater facilitation of recall. Experimental work on memory has largely ignored recall conditions as an important source of variance in recall. Melton (1963) has discussed three broad theoretical problems concerned with retrieval and utilization of traces, but only one of these—dependence of the retrieval on the completeness of reinstatement at the time of recall of the stimulating situation present at the time of input—involves the analytical separation of conditions affecting storage and those related to retrieval, and very little experimental work has been done on this problem.

The analysis of recall data in the present experiment in terms of the logically definable components of word recall, namely category recall and words-within-category recall, showed that category recall was greater under the conditions of CR than NCR and that it increased directly with the length of the list, while words-within-category recall was independent of recall conditions and remained invariant when list length increased from 24 to 48. The latter finding confirms the data reported by Cohen (1966) who found that mean word recall per category was constant for lists of 35, 53, and 70 words.

The fact that variations in recall conditions and list length have an effect on only one component of the word recall measure, but not on the other, suggests that the two components represent two independent processes of recall. One of these has to do with the accessibility of higher-order memory units into which material has been organized, while the other is concerned with the accessibility of individual items comprising the higher-order units. Accessibility of higher-order units depends on appropriate retrieval cues and on the total number of stored higher-order units (or list length), while accessibility of items within higher-order units is largely independent of these variables.

In the present experiment, and in other experiments with categorized word lists, the words to be memorized were organized into higher-order units by the *E*. This organization apparently determined the arrangement of words in the storage and their retrieval not only for *S*s working under the CR conditions, but also for those working under the NCR conditions. When two or more words from a given category were recalled by the NCR subjects, almost invariably these words occurred in immediate succession.

Even when the *E* does not impose any particular organization on the material the *S* has to memorize, by selecting words for inclusion in lists randomly and by presenting them without any additional descriptive labels, *S*s can and do organize the words into larger units (Tulving, 1962, 1964). Some of these subjective units (S-units) consist of words from meaningful conceptual categories, but others seem to be based on other principles—associative groupings, structural characteristics, and similarity of sound patterns—and still others appear to be determined idiosyncratically. It has been suggested previously (Tulving, 1964) that the functional significance of S-units, whatever their nature, lies in the increased accessibility of individual items constituting a unit. We do not yet know much about the mechanism underlying the retrieval of a single unit of information, be it an individual word or a larger S-unit, but it appears that if an individual list-item has been stored as a part of a larger unit it does become more accessible for retrieval when other items in the same unit are accessible. Thus organization of material, whether suggested by the *E* or imposed by the *S*, seems to affect recall performance primarily by making the desired information more accessible in an otherwise limited biological retrieval system. It need not have any effect on the availability of the information in the storage.

References

Brown, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, **10**, 12-21.

Cohen, B. H. Recall of categorized word lists. *J. exp. Psychol.*, 1963, **66**, 227-234.

Cohen, B. H. Some-or-none characteristics of coding behavior. *J. verb. Learn. verb. Behav.*, 1966, **5**, 182-187.

Cohen, B. H., Bousfield, W. A., and Whitmarsh, G. A. Cultural norms for verbal items in 43 categories. Tech. Rep. No. 22, Nonr-631(00), 1957, University of Connecticut.

Feigenbaum, E. A. The simulation of verbal learning behavior. *Proc. West. Joint Computer Conf.*, 1961, **19**, 121-132.

Fox, P. W., Blick, K. A., and Bilodeau, E. A. Stimulation and prediction of verbal recall and misrecall. *J. exp. Psychol.*, 1964, **68**, 321-322.

Luh, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, **31**, No. 142.

McNemar, Q. *Psychological Statistics* (3rd Ed.). New York: Wiley, 1962.

Melton, A. W. Implications of short-term memory for a general theory of memory. *J. verb. Learn. verb. Behav.*, 1963, **2**, 1-21.

Peterson, L. R., and Peterson, M. J. Minimal paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 521-527.

POSTMAN, L., AND RAU, L. Retention as a function of the method of measurement. *Univ. Calif. Publ. Psychol.*, 1957, **8**, 271-396.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.*, 1962, **69**, 344-354.

TULVING, E. Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychol. Rev.*, 1964, **71**, 219-237.

WAUGH, N. C., AND NORMAN, D. A. Primary memory. *Psychol. Rev.*, 1965, **72**, 89-104.

(Received January 18, 1965)