FUNCTIONAL UNITS AND RETRIEVAL PROCESSES IN FREE RECALL

ENDEL TULVING AND ROY D. PATTERTON

FUNCTIONAL UNITS AND RETRIEVAL PROCESSES IN FREE RECALL

ENDEL TULVING AND ROY D. PATTERSON

University of Toronto

The Ss recalled lists of 12, 16, 20, and 24 common words on a single trial. 4 types of lists were used. C lists consisted of unrelated words, while E, M, and D lists contained 4 highly related words in addition to unrelated words. The 4 highly related words appeared as a cluster at the end of E lists, in the middle of M lists, and were distributed throughout each list in D lists. Recall of words was highest for M lists, followed by E, D, and C, in this order. The number of recalled functional units, however, was identical for M and C, and lower in E than in M and C. The findings suggested that (a) highly related words are retrieved from secondary memory as a single functional unit, (b) utilization of related words in primary memory occurs only to a small extent, and (c) the number of retrieved functional units is independent of the size of the units. Models of free recall postulating two types of memory store and transfer of information from one store to the other appear to be inconsistent with the data. A more appropriate view seems to be one according to which primary and secondary memory represent different types of retrieval mechanism.

A typical S, who is presented with a list of familiar verbal items, such as common words, and immediately afterward is asked to recall the items in any order they occur to him, cannot recall all items if the list length exceeds a critical value. But if S recalls the list again, he can frequently reproduce items he did not recall in the first output phase, even in absence of an interpolated input phase (Tulving, 1967). The recall in the nth output phase of a certain number of “new” words—words not recalled in the n–1th output phase—is usually accompanied by the failure to recall an equivalent number of “old” words—words recalled in the n–1th output phase.

This finding suggests that the failure to recall some of the presented items in a given output phase does not necessarily reflect the failure of getting appropriate information about list items into the memory store, nor does it necessarily reflect the failure of maintaining this information in the store. Nonrecall of some items rather seems to reflect the limited capacity of the retrieval mechanism to find access to the information available in the store. It is as if in any given output phase the retrieval mechanism has access to a limited number of units of stored material. Access to one unit seems to reduce the probability of access to other units. In other words, it is as if free recall of discrete verbal units is limited because the retrieval mechanism has a limited capacity.

If the capacity of the retrieval mechanism is limited to a fixed number of units of material, how does one interpret systematic variations in the number of items recalled when factors such as list length, rate of presentation, and strength of interitem associations (e.g., Deese, 1959, 1960; Murdock, 1960) are varied? One possibility is to

\[1\] This research was supported by the National Research Council of Canada Grant No. APT-39 and the National Science Foundation Grant No. GB 3710.
assumed that these variables affect the size of the units that are retrieved independently of one another. Two or more nominal list-items may be grouped together into a "chunk," or "unitized" into a higher-order functional memory unit (Miller, 1956a, 1956b; Tulving, 1966b, 1968) on the basis of their associative relations, semantic similarity, or other factors. Even if the number of accessible units remains constant, the overall recall, measured in terms of the number of recalled nominal list-items, does covary with the size of the functional units.

The present experiment was designed to further explore the relation between the size of functional units and the number of recalled functional units under the conditions of a single-trial free-recall (FR) experiment. The Ss were shown lists of common words and asked to recall the words in each list immediately after the presentation of the list. The critical experimental manipulation had to do with the nature of the words in the lists. In addition to control lists, consisting of words randomly drawn from a larger pool, there were other lists each of which included a group of strongly related words, such as father, mother, brother, sister, or cow, horse, dog, cat. It was expected that S would unitize each group of such strongly related words into a higher-order functional unit, particularly if all the related words occurred in contiguous input positions in the list. Unitization of randomly selected words was expected to be more difficult and result in much smaller functional units. It was hoped that the comparison of recall of nominal units, i.e., individual words, as well as functional units, i.e., groups of strongly related words, across different kinds of lists, would provide some evidence relevant to the hypothesis that the number of recalled functional units is independent of the size of these units.

Method

Design.—There were two treatment variables in this experiment: (a) length of the list—12, 16, 20, and 24 words, and (b) type of the list—C, D, M, and E, as described below. The four types of list and the four levels of list length were combined factorially to yield 16 different treatment conditions. In addition, an eight-word C list was used, making a total of 17 treatment conditions. Each condition can be designated in terms of the length of the list, e.g., C-16, D-24, E-12, etc.

Each S was tested once under each of the 17 conditions, with the order of the treatment conditions for each S being determined by means of a table of random numbers.

Lists.—All lists consisted of common English nouns, or words that could be used as nouns, drawn from a basic pool of 296 such words. The pool contained 12 sets of 4 highly related words (R words) plus 248 "unrelated" words (U words). A set of R words consisted of either four instances of an exhaustive conceptual category (Cohen, 1963), such as north, south, east, west, or of four high-frequency members of a nonexhaustive category (Cohen, Bousfield, & Whitmarsh, 1959) such as arm, hand, leg, foot. The U words were included in the pool without any regard to their meaning. It is only in this sense that they are referred to as unrelated. The Thorndike-Lorge (1944) General Count frequencies were matched for R and U words. Approximately 75% of all words had frequencies of 50 words per million or over, the remaining 25% of the words ranged in frequency 4-49 per million.

Sets of R words and individual U words were drawn randomly without replacement from the basic pool to make up lists of the required length and of different types separately and independently for each S. The four types of list were:

1. C (Control) lists, each consisting only of U words, the number of words in each list being determined by list length (L).
2. D (Distributed) lists, each consisting of four R words from one of the 12 sets, plus L-4 U words. The R words were dis-
distributed throughout the list. They occupied serial positions of 3, 6, 9, and 11 in 12-word lists; 3, 7, 12, and 15 in 16-word lists; 3, 8, 15 and 19 in 20-word lists; and 3, 9, 18, and 23 in 24-word lists.

3. M (Middle) lists, each consisting of four R words from one of the 12 sets, plus L-4 U words. The R words were “blocked” in the middle of the list, occupying serial positions 7, 8, 9, and 10, counting from the end of the list, i.e., serial positions L-6, L-7, L-8, and L-9, with L = 12, 16, 20, or 24.

4. E (End) lists, each consisting of four blocked R words plus L-4 U words, as in M lists, except that all R words occupied the last four input positions in each list. Thus, in 12-word lists, R words were presented in serial positions 9, 10, 11, and 12; in 16 word lists they occurred in positions 13, 14, 15, and 16, etc.

Subjects.—Sixty students of both sexes taking psychology courses at the University of Toronto in the summer of 1966 served as Ss in this experiment. Participation as Ss in psychological experiments was a course requirement for these students.

Procedure.—Words were presented to Ss by means of flash cards. Each word was hand-printed on a 3 x 5 in. white index card. The E turned over the cards in a prepared deck constituting the list, at the rate of 1.5 sec. per card. The S read each word aloud as it was presented and at the end of the list—signalled by the appearance of a blank card—recalled the words orally at a self-paced rate. The recall was recorded on magnetic tape which was later used to verify the written record of S’s recall made by E at the time of recall. Sufficient number of blank cards were placed at the end of each deck to prevent S from estimating the length of a particular list from the thickness of the deck.

The amount of time given for recall varied with list length. A maximum of 3 sec. per word in the input list was allowed for recall, but when S indicated that he was unable to recall any more words from the list E proceeded to present the next list. Thus, interlist interval varied from list to list, depending on the length of the list presented, and from S to S. A new list was presented approximately 10 sec. after the conclusion of the preceding recall trial.

The instructions to Ss were to recall as many words from each list as they could, in the order in which the words occurred to them. The instructions did not include any information about the nature of composition of lists, the length of lists, or the number of lists to be learned.

RESULTS

Overall recall data.—As the first step in the analysis of the data, the mean number of words recalled was calculated for each of the 17 lists. These means are plotted as a function of total presentation time (T) per list for C, M, and E lists in Fig. 1. Since words were presented at the rate of 1.5 sec. each, T of 12 sec. corresponds to 8-word lists, T of 18 sec. corresponds to 12-word lists, etc. The vertical lines drawn through all data-points indicate the 95% confidence intervals of the corresponding population means.

Figure 1 shows that the mean number of words recalled was a monotonically increasing and approximately linear function of total presentation time (and hence of list length) for these three types of list. It also shows
that recall was highest for M lists and lowest for C lists.

The straight lines shown in Fig. 1 were fitted to the data by the method of least squares. The equations of these functions were as follows: M lists, \( R = 0.060T + 7.12 \); E lists, \( R = 0.069T + 5.42 \); and C lists, \( R = 0.065T + 4.53 \), where \( R \) refers to the mean number of words recalled and \( T \) represents the total presentation time in seconds. These equations indicate that the slopes of the three functions were practically identical and that the three types of list differed only in absolute levels of recall, a conclusion also apparent from the three functions depicted in Fig. 1.

The mean recall data for D lists yielded a more irregular picture. The means, with standard deviations given in parentheses, corresponding to list lengths of 12, 16, 20, and 24, were 6.77 (1.57), 6.45 (1.82), 7.25 (1.79), and 7.52 (2.17). Thus, mean recall for the D-12 list was higher than for the D-16 list, although the difference was not significant: \( t (59) = 1.17, p > .05 \).

The inversion between 12-word and 16-word D lists in the function relating mean recall to list length is probably attributable to the fact that in the 12-word lists two of the R words occurred in the last four input serial positions (Positions 9 and 11), while in the 16-word lists only one of the R words occurred in the last four positions (Position 15). The relevance of input position to probability of recall, not only in D lists but in others as well, will become apparent in the further analysis of the data.

Serial position curves.—The next analysis of the data had to do with the relation between probability of recall and the serial position of words in input lists. Serial-position curves were calculated for each of the 17 lists. The scope of the paper, however, permits the presentation of only those features of serial-position curves that are directly relevant to the main purpose of the experiment.

Inspection of serial-position curves derived from C, M, and E lists revealed that the relation between probability of recall and serial position at and near the end of the list (the recency effect) was largely independent of list length within a given type of list, although it varied between lists. (Some relevant data in support of this conclusion will be presented later in this paper.) Consequently, the recall data for the last 10 input items in each list were pooled over all levels of list length within each of the three types of lists (C, M, and E). Eight-word C lists were not included in the analysis.

Figure 2 depicts a summary of these data. The abscissa of the graph in Fig. 2 represents serial positions of words in input lists, the leftmost point (L) corresponding to the last word in the list, the next (L-1) corresponding to the penultimate word, and so on to L-9. The ordinate shows the proportion of words recalled, with the
proportion of 1.00 corresponding to the recall of 240 words (60 Ss x 4 Lists).

The terminal part of the serial-position curves depicted in Fig. 2 shows the typical recency effect for C lists: a very high level of recall of the last input word, a steeply descending curve over the last four words, and a gradual leveling off of the curve thereafter. The recency effect for M lists is rather similar in shape to that of C lists over the last six input positions, although R words in Input Positions L-6–L-9 in M lists were recalled at a much higher level.

The somewhat lower level of the recency effect for M lists than for C lists is accounted for by those Ss who started recall of M lists with one of the R words. This happened on 30 out of 240 (4 Lists x 60 Ss) trials. The mean number of words recalled from Input Positions L-3–L-4 for 240 trials in case of C lists was 3.12, and for 210 trials on which S did not first recall an R word in case of M lists the mean was also 3.12.

In E lists the last four input words were always R words. With the possible exception of the R word in Position L-3, these words were recalled almost perfectly. Thus, the recall of the words from the last four input positions is greater in E lists than in either C or M lists. It is important to note, however, that the recency effect in E lists does not appear to extend over a greater range of input serial positions than it does in C or M lists. The differences in the recall of words from Input Positions L-4–L-7 between M and C lists appear negligible. Theoretical implications of this finding will be discussed later in this paper.

The serial-position curves for D lists agreed quite well with those of C lists, with the exception that the R words in D lists were always recalled more frequently than the U words in corresponding serial positions in C lists. Since the R words in D lists of different lengths did not occupy corresponding serial positions, the data cannot be summarized in the same way as was done for M and E lists. Instead, a single example is presented in Fig. 3, which shows the full serial-position curves for Lists C-24 and D-24.

The probability of recall of R words in the D list varied inversely with the order of appearance of these words in the list, the first of the four being recalled at the lowest level and the last at the highest level. The same relation held also in D lists of lengths 16 and 20, but not in List D-12.

Recall of R and U words from primary and secondary memory.—The next step in the analysis of the data involved cross classification of recalled words in terms of their type, R and U, and in terms of their input position, terminal words, (Input Positions L-3–L) and preterminal words (Input Positions 1–L-4). The words in terminal positions are referred to as words recalled from primary memory (PM), in keeping with Waugh and Norman’s (1965) terminology, and to the recall of these words as PM recall.
The words in preterminal positions are referred to as words recalled from secondary memory (SM) and to the recall of these words as SM recall. This usage of the terms PM and SM seems to be justified in that a large majority of terminal words were recalled early in the output sequence. Of all the terminal words Ss recalled in the whole experiment, 87% were recalled in Output Positions 1–4, this percentage being highest for C lists (89%) and lowest for D lists (84%).

The cross classification of recalled words in terms of the type of words, R and U, and the type of recall, PM and SM, defines four additive components of recall: R-PM (read R words recalled from PM), U-PM, R-SM, and U-SM. Only recall of D lists provided estimates for all four components, however. In recall of M lists, the R-PM component was missing, and in recall of both C and E lists, two components did not occur because of the composition of those lists.

The results of the analysis of the recall data in terms of these four possible components of recall are shown in Fig. 4. Each data-point represents the mean number of words of a given class recalled by 60 Ss. The data for D-12 lists are omitted from Fig. 4, because, as stated earlier, the composition of that list differed from that of other D lists. In the D-12 lists the last four input positions included two R words, while in all other D lists they included only one R word.

Figure 4 reveals a rather interesting feature of the data. In each type of list there is one and only one component of recall that varies monotonically and approximately linearly with list length. In all four types of the list that component is the same, namely U words recalled from secondary memory. The three other components—R-PM, U-PM, and R-SM—appear to remain relatively invariant with changes in list length. Thus, PM recall is independent of list length, regardless of whether the PM words are R words or U words, and so is the SM recall of R words.

Recall of functional units.—The final analysis of the data was undertaken in the light of the assumption that when an S recalls one or more highly related words, such as the R words used in this experiment, he in fact recalls a single functional unit. The number of functional units involved in the recall of U words, of course, cannot be estimated, but since unitization of U words probably occurred to the same extent in lists of all types, this uncertainty should not invalidate comparisons across list types.

Thus, Ss recall protocols were rescored in terms of the number of functional units recalled. The S was given
credit for the recall of a functional unit when he recalled at least one R word from a given list or one U word. For example, an S who recalled three R words and five U words from a list received credit for the recall of six functional units. To the extent that unitization among U words took place, this method overestimates the number of recalled functional units, but this possible error presumably affects all list types more or less equally. And our main interest lay in the comparison of the number of functional units for different types of list.

This analysis was not applied to D lists, because the assumption about unitization of all R words in a list did not seem justifiable. While R words from E and M lists were invariably recalled as a group, R words from D lists were recalled in noncontiguous output positions frequently enough to cast doubt on the assumption that they were always processed as members of the same higher-order functional unit.

For the other three types of lists—C, M, and E—the data on the recall of functional units as defined above were pooled over lists of length 12–24. The mean number of functional units recalled was 6.30 for C lists, 6.25 for M lists, and 4.43 for E lists. Thus, the mean number of functional units is virtually identical for C and M lists, and considerably lower in E lists than in C and M lists. The t test comparing the means of E and C lists yielded a t (59) of 11, p < .01.

**Discussion**

Contiguously presented related words seem to operate as a single functional unit of memory in free recall if they occur in the middle of the input list. That is, if they are retrieved from secondary memory (Waugh & Norman, 1965) or from the long-term store (Atkinson & Shiffrin, 1967). The number of words by which recall from M lists exceeded the recall from C lists rather precisely matched the number of related words that a typical S recalled in addition to the first related word he recalled. When all related words recalled by an S from a given M list were considered as constituting a memory unit equivalent to that represented by a single unrelated word recalled by S, the number of such units recalled was identical for M and C lists. Thus, the number of functional units of memory that can be retrieved under the conditions of free recall seems to be independent of the size of the units. Recall of constituent elements of a higher-order functional unit appears to occur without any cost to the capacity to retrieve independent units as such.

The finding of independence between size of the functional unit and the number of functional units that can be recalled complements the earlier finding reported by Tulving and Pearlstone (1966). In that experiment, the number of functional units (accessible conceptual categories of words) was greater for lists of 48 words than for lists of 24 words, and greater under the conditions of cued than under the conditions of noncued recall, but the size of recalled units was invariant with these two variables. Variations in the number of higher-order functional units probably reflect the consequences of subjective organization of these units into still higher-order units, in the manner of hierarchical organization described by Mandler (1967), although for the time being this suggestion still represents only a conjecture. The important point, however, is that recall within a higher-order unit is independent of the number of such units that are recalled.

Unitization of the kind observed in M lists did not occur to the same extent in E lists in which the related words occupied four terminal input positions. The Ss memorizing E lists recalled fewer words in addition to the group of four related words than they recalled in addition to a single U word in C lists.
Furthermore, the range of input positions over which the recency effect was observed in E lists was no more extended than it was in C lists (Fig. 2). Some unitization of R words in E lists probably did occur, as indicated by the superior recall of E lists when compared with C lists, but the extent of this unitization was small. Only if it is assumed that the recall of four R words from E lists was approximately equivalent to the recall of three functional units, would the number of functional units recalled from E lists approximate the number of functional units recalled from C lists: 6.43 and 6.30 for E and C lists, respectively, with the data averaged over lists of length 12, 16, 20, and 24. Thus, it appears that related words in terminal input positions are handled more like individual words than members of a higher-order unit.

Some unitization of related words also occurred in D lists, as shown by the higher recall of words from these lists in comparison with C lists, but again the extent of this facilitation was less than that observed in the case of M lists. Whether this difference is attributable to the presentation of the related words in D lists in noncontiguous input positions, to the fact that some of the related words occurred in preterminal positions and some in terminal positions, or to both factors, cannot be determined from the existing data.

The present results do not seem to be readily reconcilable with models of free recall which distinguish between two kinds of memory as representing separate memory stores (e.g., Atkinson & Shiffrin, 1967; Glanzer & Cunitz, 1966; Glanzer & Meinzer, 1967; Waugh & Norman, 1965). If items can be stored in secondary memory (SM) or long-term store (LTS) only through their transfer from primary memory (PM) or short-term store (STS), and if only some items are so transferred, how is the high level of recall of related words in M lists and their unitization in SM or LTS to be explained? In answering this question one must assume either that related words become unitized already in PM or STS, or that they are not unitized in PM or STS and become unitized in SM or LTS. Adoption of the former position would create the problems of explaining why recall of words was less in E lists than in M lists and why recall of units was lower in E lists than in C lists. It would also require the explanation of why the recency effect was not extended over a larger number of terminal input positions in E lists than in C lists and M lists. Adoption of the latter position, on the other hand, makes it necessary to explain how related words not unitized in PM or STS are, almost without exception, transferred into, and how they then become unitized in, SM or LTS. In coping with this problem, one might perhaps argue that the probability of recall of a unit is proportional to the total presentation time for that unit (Waugh, 1967) and that the related words were recalled at a higher level of probability because the unit of which they were members was presented for a time four times longer than that of a single unrelated word. The simple proportional relation between presentation time and probability of recall, however, did not hold under the present conditions.

Because of these apparent difficulties with models postulating two separate stores, and at least until such time that the difficulties are cleared up, the authors prefer to identify PM and SM with different types of retrieval mechanism rather than with different types of store (Tulving, 1966a, 1968). It is assumed that every stimulus word in the list, or, more precisely, information about the fact that the word occurred in the list, is placed into a unitary store in a coded form. Coding refers to storage of additional information with each to-be-remembered word at the time of input (Tulving, 1966a, 1968; Tulving & Osler, 1968). This additional information serves as a source of retrieval cues that provide access to the information about the to-be-remembered words. Certain types of
retrieval cues can be manipulated experimentally (Earhard, 1967; Tulving & Osler, 1968; Tulving & Pearlstone, 1966), but some kinds of retrieval cues must be operating even if E is not aware what they are and even if he has no direct control over them. The explanation of differential accessibility of related and unrelated words, and of words from different parts of the input list, therefore, must be sought in differential effectiveness of different kinds of retrieval cues.

The terminal list items, for instance, may be stored with information about their distinctive input positions or their distinctive time tags (Yntema & Trask, 1963), and this additional information may serve as a retrieval cue. The S knows already before the presentation of a list that some items will occur in the last few input positions and hence this type of retrieval cue itself is always accessible to him. Retrieval of pre-terminal items, on the other hand, may be mediated by list as a retrieval cue. List refers to a given collection of items, differentiated from other possible lists in terms of its own temporal date or demarcation. If the items in the list, at the time of their presentation, can be readily classified by S into two distinctive categories, such as "related words" (e.g., color names—blue, green, yellow, red) and "other words," list can serve as a general retrieval cue for related words and other words which then in turn serve as more specific retrieval cues for words associated with the two categories, in keeping with the characteristics of hierarchical organization (Mandler, 1967). Since it is known that probability of recall of an item associated with a given retrieval cue varies inversely with the number of items associated with the same cue (Earhard, 1967; Tulving & Pearlstone, 1966), and since the potency of the retrieval cue varies directly with the strength of pre-experimental association between it and its associated words, the high probability of recall of related as compared with unrelated words is to be expected.

Although, at the level of experimental operations, S's task in an FR experiment can be viewed as one of recall of individual items, retrieval of stored information must be conceptualized, at the theoretical level, as always involving mediation by retrieval cues. The present speculations as to the nature of retrieval cues in free recall may be woefully inadequate, but they are amenable to experimental tests. At the very least, it seems clear that the identification of functional units of memory and the retrieval cues which provide access to these units in the memory store under the conditions of free recall constitutes a pressing theoretical problem. In the long run, nothing much can be gained by postulating a homunculus searching through one or more types of memory store for desired mnemonic information.

REFERENCES


Earhard, M. Cue recall and free recall as a function of the number of items per cue. J. verbal Learn. verbal Behav., 1967, 6, 257-263.


Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, 63, 81-96. (b)


Tulving, E. Short and long term memory: Different retrieval mechanisms. Paper presented at the meeting of the XVIII International Congress of Psychology, Moscow, August, 1966. (a)


(Received May 29, 1967)