Reprinted from PSYCHOLOGICAL REVIEW, Vol. 71, No. 3, May 1964
Printed in U. S. A.

INTRATRIAL AND INTERTRIAL RETENTION: NOTES TOWARDS A THEORY OF FREE RECALL VERBAL LEARNING[1]

ENDEL TULVING

University of Toronto

Trial-to-trial analysis of recall of individual items in a multitrial free recall learning experiment shows that the traditional learning curve can be expressed as an additive function of intratrial and intertrial retention. Intertrial retention, measured in terms of number of items retained from one trial to the next, increases as a logarithmic function of trials, while intratrial retention, measured in terms of number of items retained from the input phase of a given trial to the output phase of that trial, remains practically invariant over trials. Intertrial retention is positively correlated with subjective organization, giving support to the hypothesis that the increase in intertrial retention as a function of practice reflects the growth in the size, but not necessarily the number, of subjective units of material that the subject can retrieve from the memory storage.

One of the most universally known and reliable phenomena of memory lies in the improvement in recall of verbal materials under conditions of practice. Experiments in verbal learning, as in many other areas of psychological inquiry, often yield inconsistent and contradictory data, but no experimenter has ever reported that his human subjects, appropriately instructed, failed to increase their recall over trials of an experiment.

In spite of, or perhaps because of, the wide generality of the phenomenon, empirical and conceptual analyses of it have seldom been reported in the literature. A great deal of experimental evidence and attendant theoretical speculation is available about conditions and variables that influence the rate at which recall changes over trials, that is, about interaction effects involving trials as one of the independent variables, but systematic attempts to account for the effects of trials per se have not been popular.

The present paper reports an analysis of the trial-by-trial memorization of a list of words under the method of free recall (FR). In a typical FR experiment with which we are concerned the subject is exposed to a list of words in the "input" phase of a trial and is then asked to recall as many of these words as he can, in any order he wishes. Then the same words are presented again, usually in a different order, and again at the end of the presentation, in the "output" phase of the trial, the subject recalls the words he remembers. This procedure, alternating input and output phases, can be repeated for any number of cycles or trials.

The subject's task in the FR experiment appears to be relatively simple. Response learning in the sense of response integration (Mandler, 1954) is minimized whenever more or less

[1] This research has been supported by the National Research Council of Canada, under Grant No. APA-39, and by the National Science Foundation, under Grant No. GB-810. A preliminary version of the paper was read as an invited paper at the annual meeting of the Canadian Psychological Association in Montreal, June, 1961. The writer is indebted to George Mandler for constructive comments and cogent criticisms.

familiar words are used, and associative learning is not necessary since the experimenter gives credit to the subject for words recalled regardless of their order. Memorization of a list of words in the FR experiment is rather reminiscent of that aspect of response learning that has to do with strengthening of responses specified by the input list so that they have "greater response strength than the many other responses in the repertoire which are not in the list the subject is to learn [Underwood & Schulz, 1960, p. 93]." If this reasoning is correct, then FR learning must involve only some of the processes underlying acquisition in serial and paired-associated learning experiments. It thus constitutes a simpler task for the subject to master and a simpler situation for the experimenter to analyze.

Probably because of its apparent simplicity, free-recall learning has failed to arouse much curiosity among experimenters. Only a handful of studies having to do with multitrial FR learning have been reported in the literature (Bruner, Miller, & Zimmerman, 1955; Horowitz, 1961; Murdock, 1960; Waugh, 1961). But it has appealed to several theorists who have constructed various stochastic models to describe its data (Bush & Mosteller, 1955; Miller & McGill, 1952; Waugh & Smith, 1962). All of these models have succeeded in providing good or even excellent mathematical descriptions of certain quantitative aspects of empirical findings. Yet they have not been primarily concerned with isolation and identification of processes involved in memorization of verbal material. The Waugh and Smith (1962) model perhaps goes farthest of the three in that it refers to three hypothetical processes—"labelling," "selecting," and "fixing"—identified with the three parameters of the model, but these processes are not related to any "classical psychological functions [p. 142]."

Another characteristic that the stochastic models have in common lies in their assumption that individual list items are memorized independently of one another. The recall of a given list item is supposed to have no effect on the recall or nonrecall of any other item. The fact that such an assumption is not tenable is well known to the authors of the statistical models, although the implications of the correct assumption of item relatedness have not yet been fully explored. In considering the problem briefly, Miller and McGill (1952) suggested that "associative clustering should affect the variability, not the rate, of memorization [p. 390]," but this somewhat startling hypothesis has not been subjected to an empirical test.

Finally, the description of FR learning provided by the stochastic models is difficult to reconcile with the fundamental fact that an individual list item, if it has been previously integrated, can be "learned" by the subject on one trial of a very short duration, in the sense that the subject can always recall it immediately after seeing or hearing it once. What is it that permits the single item to be labeled, selected, and fixed with a probability of unity, to use Waugh and Smith's language, if it is presented alone, but probabilities much smaller than unity when the item appears in a list? This particular criticism, of course, can be directed at almost any theoretical account of memorization, both formal and informal, that has ever been proposed, but the commonality of an error does not justify it.

Our present analysis of FR learning takes as its starting point the observation that a small unit of verbal material, such as an individual list word, is always "learned" at the time of its presentation to the subject. It then proceeds to examine to what extent these individual items are retained or forgotten, both within a single trial and over successive trials. The traditional learning curve will be described as a composite of intratrial and intertrial retention curves. Finally, intertrial retention as the main component of the subject's total recall performance will be shown to be related to the extent to which subjects organize their recall. It will be argued that the trial-by-trial improvement in recall is a consequence of the development of higher-order units of material which mediate the retrieval of the information from the memory storage.

ONE-TRIAL LEARNING AND INTRATRIAL FORGETTING OF INDIVIDUAL ITEMS

Let us begin by considering what happens on the first trial in a typical FR experiment. Each subject will be able to recall some but not all words in the output phase of the first trial. Such limitation of first-trial recall has traditionally been ascribed to some limitation in the learning mechanism, that is, it has been interpreted as reflecting the fact that the subject, for some reason or other, cannot *learn* all words on the first trial. For instance, in the well-known debate on whether associations in paired-associate learning tasks are formed in an all-or-none or incremental fashion, the participants, although disagreeing on the main issue, seem to agree on interpreting nonrecall of some response items as evidence for absence of, or incomplete formation of associations in the input phase preceding the recall test (Estes, 1960, 1961; Postman, 1963; Rock, 1957; Underwood & Keppel, 1962).

Such an interpretation, if taken literally, may prove somewhat misleading in pursuing the problems involved in verbal learning. Recent evidence shows clearly that a small unit of well-integrated verbal material, or an association between two such units, can practically always be recalled immediately following its presentation (Brown, 1958; Murdock, 1961a, 1961b, 1963; Peterson, Peterson, & Miller, 1961; Peterson, Saltzman, Hillner, & Land, 1962; Tulving & Arbuckle, 1963). When we use probability of response as a measure of learning, therefore, we must conclude that learning of a small unit of material is always complete on a single trial. Learning, in this sense, is neither incremental nor all-or-none, it is always "all."[2] Whether the small unit is presented alone or in a series of other units of the same class is immaterial as far as its one-trial acquisition is concerned. If we presented a list of words to the subject and warned him in advance that we might stop the presentation at any time and test him on the word seen last, we would have every reason to believe that his recall would be practically perfect.[3]

[2] Rock, in considering the problem of "why, in multiple-item learning situations, it is not possible as a rule to learn more than a few associations on any one trial," mentioned the possibility that an association is always formed and "that the failure to get all items right on a test following the entire series is a matter of forgetting of already formed associations [Rock, 1957, p. 192]." He rejected this interpretation for reasons that are not entirely clear to this writer.

[3] It is possible, of course, that subjects in certain types of experiments may in effect "ignore" some input items at the time of their presentation and concentrate their attention on a few specific items. One such selective strategy has been proposed by Feigenbaum and Simon (1962) in their account of the serial position curve. These strategies might indeed explain why the immediate recall of single items or associations in some experiments has been somewhat less than perfect, but their existence cannot change

This "learning" of an individual item—whether we call it learning, perception, fixation, registration, or what not, does not matter, as long as we know the operational referent of the term—depends solely upon variables operating prior to and at the time of the presentation of the item. It must be distinguished from the retention of the item which depends on the conditions prevailing at the time of the learning as well as on certain variables operating in the retention interval, however short. Thus the fact that an individual item is always "learned" following a single presentation does not mean that it might not be forgotten later. Small amounts of material, well within the classical memory span of the individual, often become unavailable for recall in a matter of a few seconds (Brown, 1958; Murdock, 1961a, 1961b, 1963; Peterson & Peterson, 1959; Tulving & Arbuckle, 1963). Nor does the conception of one-trial learning in any way preclude the possibility that additional presentations or rehearsal of the once learned and recalled item may change its resistance to forgetting. Just as overlearning of larger units of material, such as lists of nonsense syllables, has been long known to improve the retention of material (Ebbinghaus, 1885), smaller units have also been shown to benefit from repeated presentation (Peterson & Peterson, 1959; Peterson, Saltzman, Hillner, & Land, 1962; Waugh, 1962b).

It seems reasonable to argue, in view of the foregoing considerations, that limited recall on the first trial does not necessarily reflect incomplete learning, but rather, incomplete retention. Although the intratrial retention intervals, between the presentation and the attempted recall of each individual item, are short, many learned items can be forgotten within that interval. Whether such intratrial forgetting is interpreted in terms of the decay of memory traces (Broadbent, 1958; Brown, 1958; Conrad & Hille, 1958), in terms of input and output interference (Tulving & Arbuckle, 1963), or in terms of yet other factors, is not immediately relevant to our present purposes. What is important is the empirical fact that some individual items, although always recallable immediately following their presentation, do become unavailable for recall well within the interval occupied by a single trial. The recall scores on the first trial, therefore, reflect the combined effects of one-trial learning and intratrial forgetting, or simply of intratrial retention.

We may, if we wish, regard the subject's recall score on the first trial as a measure of his "immediate memory," provided that we realize that such "immediate" memory is limited for the simple reason that it is not immediate. In the FR experiment, it is only the last item that the subject could recall immediately after its presentation, but in a learning task in which the subject has to recall the items in the order of their presentation, recall is not immediate for any single item.

The notions of one-trial learning and intratrial forgetting suggest that it is not the fact of "storage" of list items that is at issue in the FR experiment, but rather the form of storage, or accessibility, of items (cf. Miller, Galanter, & Pribram, 1960). The list items have been "stored" in the subject's memory a long time before he appears for the experiment and the input list serves simply as a set of instructions as to which of the stored items the subject has to retrieve. In

the argument that one-trial learning can, and usually does, occur.

this sense, then, what the subject has to do in the FR experiment is to increase the accessibility of those items that are specified by the input list so that they can be readily retrieved from the storage.

Be it as it may, there are certain advantages in thinking of the first-trial recall as a measure of retention rather than learning. First, it might be easier to deal with questions such as why some items are retained and others forgotten within a single trial (e.g., Tulving & Arbuckle, 1963) than questions such as why some items are and others are not learned, or why each item is learned only to a certain extent and not completely. In terms of our existing conceptual tools as well as available empirical evidence we are probably better equipped to handle problems of retention and forgetting than of original acquisition. Second, and more important for the task at hand, the concept of intratrial retention leads naturally to the concept of intertrial retention and thus suggests a potentially useful way of looking at trial-by-trial recall data. The relation of intratrial and intertrial retention to the overall recall performance will become clear if we consider the pattern of recall of individual list items from trial to trial. This will be done in the next section.

TRIAL-TO-TRIAL ANALYSIS OF RECALL

We have argued that under appropriate conditions each list item is always learned on the first trial and that the subject's performance in the output phase of that trial is to be regarded as a measure of intratrial retention. The subject's performance on the second trial, and on all succeeding trials, can similarly be regarded more profitably as a measure of retention. On these trials, however, there are two sources of retained items. Some items could be recalled in the output phase of the second trial, for instance, even without the benefit of their re-exposure in the input phase of the second trial: the subject remembers some of the responses he made in the output phase of the first trial. The number of these responses provides a measure of intertrial retention. In addition to these responses, the subject, in the output phase of the second trial, also remembers some of the responses he made in the input phase of the second trial. Their number provides a measure of intratrial retention. The subject's overall recall performance on all trials except the first, therefore, consists of two components—intertrial and intratrial retention.

The conception of recall performance in terms of the process of retention or, looking at the other side of the coin, of forgetting, is not a novel one. McGeoch, for instance, more than 20 years ago, pointed out the relations involved:

> The changes in behavior (verbal responses) acquired during the first trial are retained, at least in part, until the second trial. There new ones are added to those retained; some or all of the results of the first and second trial are retained until the third trial, when more are added, and so on, until practice stops. . . . Not all of the acquisitions at each successive trial are carried over to the next; some are forgotten and must be refixated. A curve of learning represents a progressively greater balance in favor of retention, so that it is, in part, a retention curve. . . . Fixation and retention thus mutually interact in the course of what we call learning [McGeoch, 1942, pp. 4–5].

A single practice trial, as well as a series of trials, thus serves at least two functions. It provides the learner with the opportunity to study and retain individual items, but certain events occurring on the trial must also be responsible for forgetting of some other items. The traditional procedure used in verbal learning experiments, in which the experimenter assumes some kind of an equivalence of different list items and expresses the subject's performance in terms of the total number or proportion of items recalled, perhaps simplifies computations, but it clouds the two roles played by the trial.

The traditional method for describing the subject's performance in a verbal learning experiment is based on what might be called the trial-by-trial analysis of recall. On every trial, the experimenter partitions the total set of L items in the list into two mutually exclusive subsets, P_n and N_n. P_n (Performance) refers to the number of items that the subject has recalled on Trial n, while N_n refers to the number of items that the subject has failed to recall on that trial. $P_n + N_n = L$, hence the specification of P_n completely determines its complement N_n and nothing further can be gained from measuring N_n. The traditional learning curve is simply the plot of the size of the subset P_n against trials.

The trial-by-trial analysis of recall is to be contrasted with what will here be referred to as the trial-to-trial (TTT) analysis. This analysis is an extension of that used in Estes' (1960) "miniature experiments" to the typical multitrial experiment. In the TTT analysis, the Set L is partitioned with respect to pairs, triplets, etc., of successive trials. In the present paper, however, we shall be concerned with the TTT analysis as applied only to successive pairs of trials.

The subject's recall on two successive trials, $n - 1$ and n, defines within the Set L two subsets, P_{n-1} and P_n, thus placing each element of the Set L (each word in the list) into one of four mutually exclusive subsets. These four subsets are shown in the Venn diagram in Figure 1. The large circle represents the Set L, the small circle

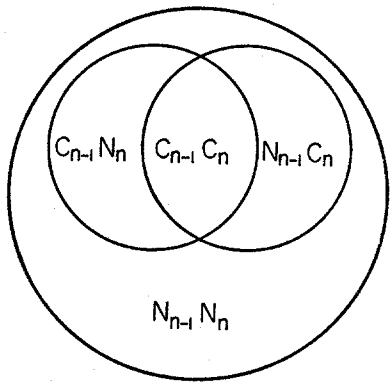

Fig. 1. Venn diagram representing four components of performance derived from the application of the trial-to-trial analysis to consecutive pairs of trials.

on the left represents the subset P_{n-1}, consisting of items that the subject recalled on Trial $n - 1$ and the small circle on the right represents subset P_n, consisting of items that the subject recalled on Trial n.

The four mutually exclusive subsets shown in the diagram are designated as follows:

$C_{n-1}C_n$ (or simply CC) is the intersection of the subsets P_{n-1} and P_n. It consists of items that occur in the subject's recall both on Trial $n - 1$ and on Trial n. The size of this subset can be regarded as an estimate of intertrial retention.

$N_{n-1}C_n$, (NC), consists of items that occur in the subject's recall on Trial n, but not on Trial $n - 1$. Its size provides an estimate of intratrial retention.

$C_{n-1}N_n$, (CN), consists of items recalled on Trial $n - 1$, but not recalled ("forgotten") on Trial n. Its size can be thought of as an estimate of intertrial forgetting.

$N_{n-1}N_n$, (NN), consists of items that the subject fails to recall both on Trial $n - 1$ and on Trial n. Its size can be regarded as an estimate of intratrial forgetting.

The logical definition of the four

subsets is straightforward, but the identification of the subsets with psychological processes of intratrial and intertrial retention and forgetting involves assumptions that may not be quite correct. For instance, referring to the *CC* component of performance as an estimate of intertrial retention ignores the possibility that an item in Subset *CC* was recalled on Trial n, forgotten by the time it was shown to the subject again in the input phase of Trial $n + 1$, "relearned" then, and retained until the output phase of Trial $n + 1$. Thus, both intertrial and intratrial retention may contribute items to the *CC* component which therefore may overestimate the amount of material that the subject can retain from one trial to the next without any intervening input. For the same reason the *NC* component may underestimate intratrial retention. Moreover, the extent of these errors of estimation may vary systematically over trials. These relations clearly constitute problems for further research. For the time being, however, it does not seem completely unreasonable to regard the four measures—*CC, NC, CN,* and *NN*—as estimates of intertrial and intratrial retention and forgetting as designated above.

If the TTT analysis is extended over all consecutive pairs of trials (Trials 0 and 1, 1 and 2, 2 and 3, . . . $N - 1$ and N, where N refers to the total number of trials), and if the size of each of the four subsets is plotted against trials, four curves are obtained: an intertrial retention curve, an intratrial retention curve, an intertrial forgetting curve, and an intratrial forgetting curve.

These four curves contain all the information that is available in the trial-by-trial "learning curve," but they also lay bare the anatomy of such a curve. Any two of the four curves can be mathematically derived from the other two, even though for a given pair of trials only one of the four measures is redundant with the other three. This fact reflects the complementarity of the definitions of the two processes, retention and forgetting: whatever the subject does not retain, he has forgotten, and vice versa. It also means that the traditional learning curve, in which P is plotted against trials, can be described in terms of any two of the four components.

In this paper, we shall be mainly concerned with the following relation:

$$P_n = C_{n-1}C_n + N_{n-1}C_n \qquad [1]$$

Equation 1 is a simple mathematical statement to the effect that performance is an additive function of two components, *CC* (intertrial retention) and *NC* (intratrial retention). Performance, either on a single trial or over a number of successive trials, can be analyzed into the *CC* and *NC* components, and these two components, again either on a single trial or over successive trials, can be synthesized into overall performance.

Empirical Synthesis and Mathematical Analysis of the Free Recall Learning Curve

Let us next examine some data from a simple FR experiment and demonstrate, first, how the learning curve can be derived from the empirically determined measures of *CC* and *NC*, and second, how the two components can be estimated from the overall performance data.[4]

Thirty-two summer school students served as subjects in the experiment. Each subject learned a list of 22 words on 22 trials by the method of FR. Words were presented to the subject

[4] Thanks are due to Albert S. Bregman who collected these data and wrote the electronic computer programs for analyzing them.

by means of a Gerbrands memory drum, at the rate of one word per second. Subjects recalled words by speaking them out aloud. Their responses were recorded on tape and later transcribed. No definite amount of time was given for recall in the output phase; when the subject had been silent for 10 seconds, the input phase of the next trial was begun.

There were four different, but formally equivalent stimulus lists. Each list was learned by eight different subjects. The lists are shown in Table 1. The words were selected from the Thorndike-Lorge (1944) word book, according to the following criteria: (*a*) All words were two-syllable nouns, or words that could be used as nouns. (*b*) No two words in the same list began with the same letter. (*c*) In each list, the words were distributed approximately equally among four frequency-of-occurrence categories. These four categories were: words occurring more than 100 times per million, those occurring 14 to 16 times per million, those occurring 4 times per million, and those occurring 4 times per 18 million. Apart from these restrictions, the words were assigned to different lists in a haphazard fashion.

TABLE 1

LISTS OF WORDS USED IN THE EXPERIMENT

1	2	3	4
accent	amice	action	answer
barrack	bridle	bandage	buyer
center	canard	country	cherub
drumlin	daddy	dipper	despot
entry	express	effort	ether
finding	flower	farrow	fasces
garden	gable	gambler	gorget
hoyden	hormone	hamlet	hermit
issue	impact	island	journal
jungle	juror	kitchen	letter
kernel	kitty	legion	mantel
lagoon	lactose	miser	natron
maxim	midden	noodle	orphan
newell	novice	octroi	person
office	ocean	pilgrim	question
pomade	quarter	quinsy	rennin
quillet	rumor	rennet	satin
relique	trochee	stamen	tempest
surtout	union	trollop	umbra
treason	virgin	voter	vulture
valley	wafer	waiver	windrow
walker	zither	zenith	xylem

The order of words within a given list varied systematically from trial to trial. Twenty-two different orders were constructed according to the method used in a previous paper (Tulving, 1962a). Each word appeared in each serial position just once over the total block of 22 trials, and was preceded and followed immediately, as well as by lags of 1, 2, 3, etc., by every other word in the list just once. This arrangement provides for complete absence of second-order sequential constraint among items in input lists over the block of 22 trials, and it minimizes all higher-order redundancies. The sequence in which different input orders of a given list were presented to subjects on successive trials was also systematically varied. Every subject received the input orders on successive trials in a different sequence.

The sizes of each of the four subsets defined by the TTT analysis were determined for all 22 pairs of trials (Trials 0 and 1, 1 and 2, 2 and 3, . . . 21 and 22), separately for each subject. The mean sizes of the four subsets, based on data from all 32 subjects, were then calculated. In Figure 2, these mean sizes of the four subsets are plotted against trials, yielding four different curves.

The intertrial retention (*CC*) curve is the only one that rises systematically over trials. It is a monotonically increasing negatively accelerated curve. The intratrial retention (*NC*) curve is an approximately linear curve with a negative slope. The intertrial forgetting (*CN*) curve is obviously nonmonotonic: it rises for the first few trials and then, apart from the presumably "chance" deflection upward at Trial 8, remains practically parallel with the abscissa. The intratrial forgetting (NN) curve is a monotonically decreasing curve with a negatively accelerated slope.

Both the theoretical and actual values for CC and CN on Trial 1 are 0, since subjects can neither retain nor forget any items from Trial 0 to Trial 1. These values are not shown in Figure 2. The value of NC on Trial 1, on the other hand, is determinable, and so is the value of NN. N_0C_1 corresponds to the performance score on Trial 1: $N_0C_1 = P_1$ and $N_0N_1 = L - N_0C_1$.

On inspection of the data it appeared that, as a first approximation, the intertrial retention (CC) curve could be described as a linear function in log n. The relation appears in Figure 3. The ordinate shows the mean number of items in the CC category in the recall

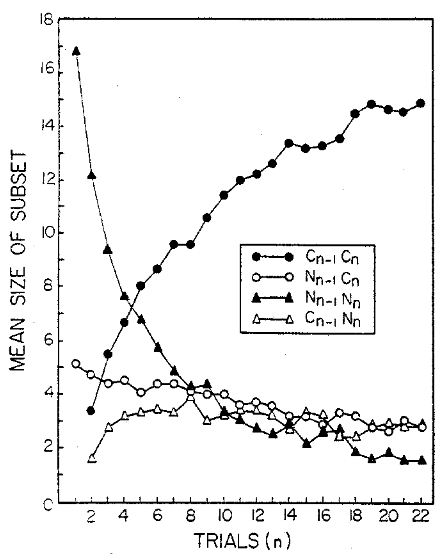

FIG. 2. Mean sizes of subsets corresponding to four different components of performance derived from trial-to-trial analysis of recall data.

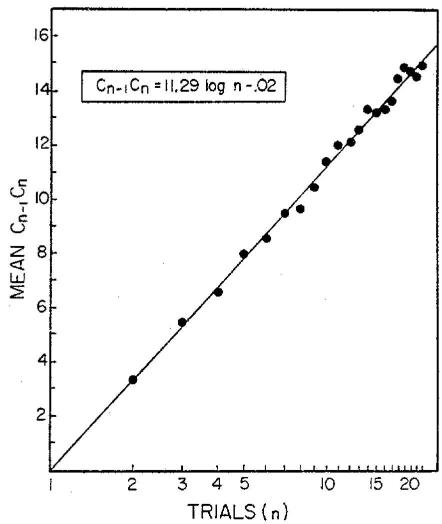

FIG. 3. Intertrial retention as a function of trials. (Trials on the abscissa are presented on the logarithmic scale. The ordinate represents the mean number of words recalled on two consecutive trials, $n-1$ and n. The straight line was fitted to the data by the method of least squares.)

of 32 subjects, the abscissa refers to trials on a logarithmic scale.

The least-squares method yielded the following equation for CC as a function of trials:[5]

$$C_{n-1}C_n = 11.29 \log n - .02 \quad [2]$$

This equation is represented by the straight line drawn in Figure 3. The fit is obviously very good. It is interesting to note that the curve intersects the ordinate corresponding to $n = 1$ at a CC value of 0 as it should, since, by definition, C_0C_1 must be zero. An

[5] All equations fitted to the data can be considered to hold only within the limits of the number of trials used in the experiments. The CC function, for instance, cannot rise indefinitely, since the length of the list will impose a limit on the subjects' recall. This limit, of course, does not characterize the subjects' information-processing ability but only the experimental conditions under which this ability is permitted to operate (cf. Waugh, 1962a).

important implication of this fact is that one could, for the data in the present sample, predict the whole *CC* curve rather accurately from a single point, the *CC* score on Trial 2, by drawing a straight line through the zero intersect and that point.

Intratrial retention (*NC*) is shown as a function of trials in Figure 4. As a first approximation, for the data in this particular sample, the relation between mean *NC* and trials can be represented by a linear function with a slight negative slope. The least-squares method yielded the following equation for *NC* as a function of trials:

$$N_{n-1}C_n = 4.87 - .10n \qquad [3]$$

The next step was to add alegbraically, on the basis of the definition of performance given in Equation 1, the intertrial and intratrial retention functions to obtain the equation of the trial-by-trial learning curve. The expression on the right-hand side in Equation 4 is the algebraic sum of corresponding expressions in Equations 2 and 3:

$$P = 11.29 \log n - .10n + 4.85 \qquad [4]$$

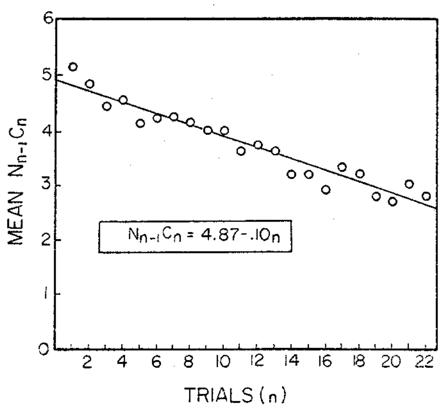

FIG. 4. Intratrial retention as a function of trials. (The ordinate represents mean number of words recalled on Trial n but not Trial $n-1$. The straight line was fitted to the data by the method of least squares.)

We have now "synthesized" the learning curve from two empirically determined components of performance, intertrial and intratrial retention. How well does such a curve fit the performance data? Equation 4 is graphically depicted in Figure 5, together with mean performance data on 22 trials. The fit appears good. The standard error of estimate, $\sigma_{P \cdot n}$ was found to be .246.

Let us now turn to the problem of the mathematical analysis of the learning curve into its two component functions, without going through the relatively tedious operation of counting items in the *CC* and *NC* categories in the subjects' recall protocols. The mathematical analysis is, in a sense, a complementary operation to the synthesis of the learning curve from the empirically determined *CC* and *NC* functions.

The logarithmic function of the general form,

$$P = a \log n + bn + c \qquad [5]$$

suggested by trial-to-trial analysis, is a statement to the effect that P is a linear function of two variables, $\log n$ and n. It is identical with the multiple regression equation of the form, $Y = aX_1 + bX_2 + c$. Its parameters, a, b, and c, can be estimated from the product-moment correlations among P, $\log n$, and n in the same manner as the constants in the multiple regression equation.

Such a multiple regression analysis of the mean performance data over 22 trials in the experiment, based on the method of least squares, yielded the following equation of the learning curve:

$$P = 10.70 \log n - .073n + 5.07 \qquad [6]$$

Equation 6 is to be compared with Equation 4, in which the parameters

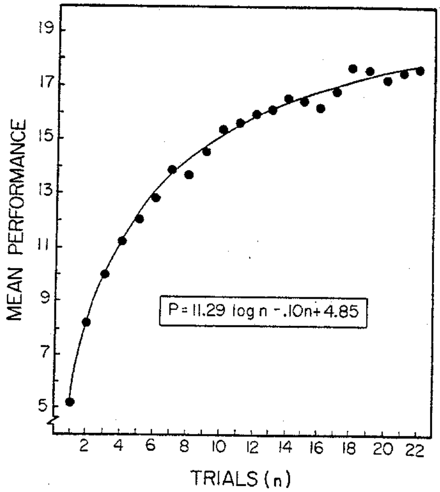

FIG. 5. The logarithmic learning curve. (The smooth curve represents the function obtained by the algebraic addition of the intertrial and intratrial retention functions shown in Figures 3 and 4. The data points represent mean number of words recalled by 32 subjects.)

were estimated from the empirically determined *CC* and *NC* components of performance. The agreement between the two sets of values, while not perfect, is close enough to suggest that they represent the same population parameters and that the differences are attributable to errors of measurement. In fact, it can be shown that with infinitely large samples in which sampling error is eliminated, the two procedures, TTT analysis and regression analysis, would yield identical results, provided of course that the forms of the functions correspond to those in the present sample, linear in log n for *CC* and linear in n for *NC*.[6]

[6] While it can be argued that the main value of equations fitted to learning curves lies in new insights they generate, rather than in their ability to describe the data, it is not without interest that the logarithmic curves do seem to fit the FR data quite well. The multiple regression method was used to generate learning curves, of the form given by Equation 5, for five other sets of FR data collected at Toronto. The goodness of fit of these curves was then compared with that provided by exponential functions of the form, $P = a - be^{-cn}$ (Murdock, 1960). In each case, the standard error of estimate was much lower for the logarithmic than the exponential curve, the difference being primarily attributable to the superior ability of the logarithmic curve to describe data from the first few trials.

INTRATRIAL RETENTION AND PRACTICE

We have seen that both logically and empirically performance can be shown to be an additive function of two components, intratrial and intertrial retention. For the data from our demonstration experiment, intratrial retention decreased as a function of trials, while intertrial retention increased. Since the rate of increase in intertrial retention exceeds the rate of decrease in intratrial retention, total recall goes up over trials.

In evaluating the intratrial retention data we must keep in mind a possible artifact that might be responsible for the observed decrements in *NC*. As the subject's intertrial retention increases, there remain fewer and fewer items in the list that the subject could retain from the input phase of a given trial. This would be particularly true for short lists that are mastered relatively quickly. It is possible, therefore, that the negative slope of the mean *NC* function, whenever it occurs, is attributable to the combined effects of limited list length and rapid learning by some subjects, and that it does not reflect the relation between practice and intratrial retention faithfully.

Three observations bearing on this issue should be mentioned. The first concerns the *NC* functions for "fast" and "slow" learners in the experiment described above. The equation of the *NC* function for the subgroup of 16 fast learners (those with the mean P scores above the group median) turned out to be $N_{n-1}C_n = 5.4 - .18n$, while for the 16 slow learners it was $N_{n-1}C_n = 4.4 - .03n$. Thus for those subjects whose performance is presumably less affected by the list length, the NC function is almost flat.

The second source of evidence supporting the hypothesis that the negative slope of the NC function represents an artifact lies in the findings of an unpublished experiment conducted at Toronto. In this experiment, 24 subjects learned three successive lists of 52 words each by the method of FR. The list words covered a wide range of Thorndike-Lorge (1944) frequencies. They were presented to the subjects on eight trials, at the rate of one second per word. The subjects had 104 seconds at the end of each input phase to recall as many words as they could. Their oral recall was recorded on tape and later transcribed.

In this experiment, no subject ever recalled more than 37 items correctly on any single trial, and mean P on the last trial, Trial 8, was only 26.2. The NC data from this experiment, even when averaged over all subjects, should thus be relatively free from any artifactual constraints. These NC data, pooled for the 24 subjects over all three lists, are shown in Figure 6.

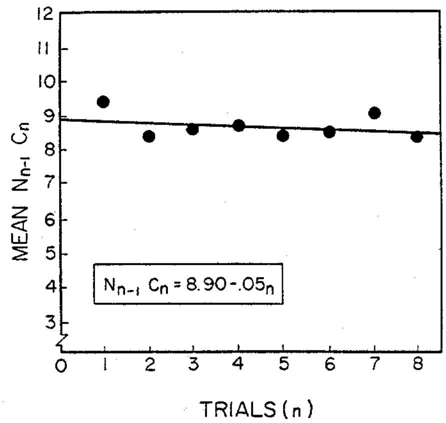

FIG. 6. Intratrial retention as a function of trials for the experiment involving lists of 52 words. (The ordinate represents mean number of words recalled on Trial n but not on Trial $n-1$. The straight line was fitted to the data by the method of least squares.)

The slope of the function is still negative, but it is smaller than that in Figure 4, namely $-.05$. More important, the negative slope is largely produced by the mean NC value on the very first trial. There seems to be little systematic change in NC over Trials 2 to 8. The finding that NC on Trial 1 is higher than on all subsequent trials has occurred in all samples of data that we have examined to date, and eventually we will have to come to grips with the theoretical problems that this finding entails. For the time being, however, we can conclude that for long lists intratrial retention seems to be essentially constant over trials.

The third observation relevant to the hypothesis comes from an interesting experiment reported by Murdock and Babick (1961). These investigators presented a "critical word" (CW) to the subjects on a number of successive trials, but always in the context of different lists. They found that the probability of recall of CW was not influenced by the frequency of previous presentations. This probability can be regarded as a measure of intratrial retention, estimated under conditions where list length does not impose any artifactual limits on recall. Since it did not change as a function of trials, we can conclude that intratrial retention is independent of practice.

It must be noted that Murdock and Babick examined only the probability of initial recall of their critical words while our NC measure includes both initial recalls and re-recalls of previously recalled and then forgotten items. The comparability of the two sets of data, therefore, may seem questionable. But the fact that the number of items in the *NC* category—particularly for longer lists that are not mastered very rapidly—remains nearly constant over trials suggests that recall on Trial n of items not recalled on Trial $n-1$ is independent of the frequency of previous recalls of these items and that in this sense original recalls are quite comparable to re-recalls. In a typical FR experiment the number of initial recalls per trial usually decreases as practice proceeds, and the number of re-recalls increases. The total number of all items in the *NC* category, however, remains practically invariant at all stages of practice.

At first glance the conclusion that recalls of items on Trial n following nonrecall on Trial $n-1$ is independent of frequency of previous recalls may appear at variance with the well-known fact that the probability of recall of an item is a direct function of the number of times that it has been recalled before (e.g., Miller & McGill, 1952; Underwood, 1954). Furthermore, Waugh and Smith (1962) have shown that the probability of recall of an item on Trial n following its nonrecall on Trial $n-1$ does depend on the frequency of previous recalls.

The contradiction, however, may be more apparent than real. The probability of occurrence of events of a given class may increase, but if, at the same time, the size of that class decreases, the frequency of occurrence of these events may not change at all, or even change in the opposite direction. For instance, an item that has been recalled, say, on eight successive trials, and then forgotten on the ninth trial, may be recalled again on the tenth trial with a very high probability, but if recall on the ninth trial depends on the frequency of recall on the previous eight trials, there will be very few items that are forgotten on the ninth trial. In short, the apparent contradiction between Waugh and Smith's conclusion and the hypothesis that intratrial retention is independent of frequency of previous recalls remains a contradiction only as long as we confuse relative frequency, that is probability, with absolute frequency of occurrence.

In the light of these observations, then, it seems reasonable enough to entertain the hypothesis that in an ideal FR experiment, in which the subject's performance is determined only by the capacity of his memory, intratrial retention is independent of practice. The possibility, mentioned earlier, that the *NC* component underestimates intratrial retention and that intratrial retention may in fact increase over trials, however, cannot yet be completely ruled out.

INTERTRIAL RETENTION AND SUBJECTIVE ORGANIZATION

There is relatively little that needs to be said about intratrial retention in this paper. We have argued, in the light of empirical data, that intratrial retention remains essentially invariant over trials, and invariances of nature do not require explanations in the same sense as do variances. There are, however, several implications that the alleged invariance in intratrial retention has for the theory of FR learning. One such is provided by the fact that if intratrial retention is independent of trials, it must also be independent of intertrial retention. Regardless of the number of items that the average subject remembers from his previous recall, he always remembers the same number of "new" items from the immediately preceding input, where new items are those that did not occur in the immediately preceding output. Recall of the "old" does not seem to interfere with recall of the new. The exact significance of this and other implications may become clearer when we understand the processes involved in intratrial and intertrial retention. For the present we can only conclude that the limitation found in immediate memory (Miller's magical number seven, 1956b) also seems to apply to memory for new items on later trials.

In this last section of the paper we return to the problem of improvement in recall over trials. In the light of the preceding analysis this has now become the problem of increments in intertrial retention over trials. Even though the TTT analysis has contributed to the conceptual clarification of the nature of the traditional learning curve, it cannot illuminate the processes responsible for various components of recall performance. Some insight into the process of intertrial retention, however, can be derived from certain additional findings from the demonstration experiment described earlier in this paper. These findings pertain to subjective organization.

Subjects do not only recall more and more items from the input list as practice proceeds in the FR experiment. They also impose an increasingly tighter sequential organization on the recalled material. It is this organization, absent in input lists and present in output lists, that is referred to as subjective organization (*SO*). The concept has been discussed and a method for its measurement has been described in a previous paper (Tulving, 1962a). In that paper it was also argued that trial-by-trial increments in performance are functionally dependent upon organization.

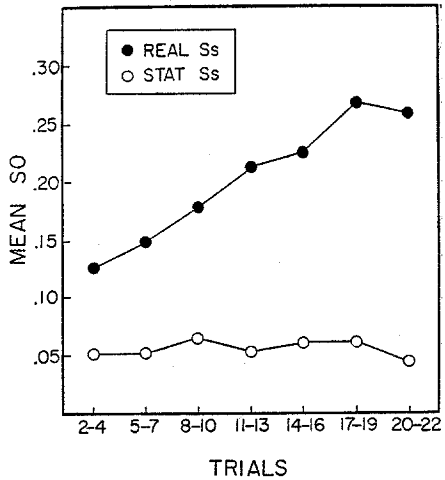

FIG. 7. Mean subjective organization (*SO*) scores on successive blocks of three trials for 32 experimental and 32 statistical subjects.

This argument can now be further evaluated by examining the relations between *SO* on the one hand and the components of performance that have emerged from the TTT analysis, intratrial and intertrial retention, on the other hand. In the demonstration experiment, *SO* (Lag 0) was calculated for each of 32 subjects on successive blocks of three trials (Trials 2–4, 5–7, . . . 20–22), as well as on blocks of seven trials (Trials 2–8, 9–15, 16–22).

Figure 7 shows the mean *SO* (Blocks = 3) for seven successive blocks of trials. It also shows mean *SO* scores for 32 statistical subjects, whose performance was matched with that of the experimental subjects, subject by subject and trial by trial. It can be readily seen that *SO* increases lawfully over trials for the real subjects, but not for the statistical subjects.

Figure 8 shows two parallel effects

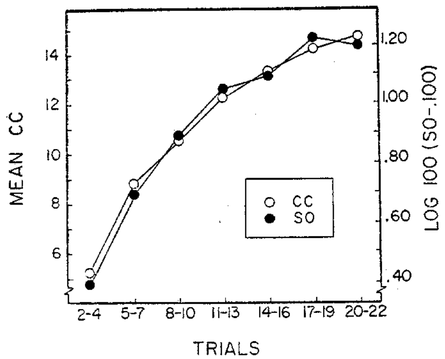

FIG. 8. Mean intertrial retention and a logarithmic transformation of mean subjective organization (*SO*) scores for blocks of three trials. (The left hand ordinate represents intertrial retention scores, the right hand ordinate shows the values of log 100 [mean *SO* − .100].)

of practice, increasing intertrial retention and increasing *SO*. Mean *CC* scores are plotted against blocks of three trials, together with a logarithmic transformation of the mean *SO* scores. The correspondence between the two curves is very good, the product-moment correlation for the seven pairs of mean scores being +.996. The important fact, of course, is that both variables increase monotonically over trials; the close correspondence between the two curves in Figure 8 depends on the transformation of *SO* scores and on the appropriate placing of the values on the two ordinates.

Does this high degree of correlation between *SO* and *CC* still occur when trials are held constant and the correlation coefficient is computed for subjects? Two variables that covary for experimental conditions need not covary for subjects (Mandler, 1959). To examine the interrelations among different response variables available from the experiment, product-moment correlations of various measures were calculated for all 32 subjects. For each subject, mean *P*, mean *CC*, mean *NC*, and log *SO* (Lag 0) were calculated for three successive blocks of seven trials, omitting Trial 1. These data are summarized in Table 2.

In interpreting these correlation coefficients, we must remember that as subjects approach the limit of performance determined by the list length L, their *CC* scores must necessarily become higher and *NC* scores lower. Thus, the negative correlations between *CC* and *NC*, as well as between *NC* and other measures, on Trials 9–15 and 16–22, are probably to a large extent artifactual. Data from the first block of trials may be less biased in reflecting the relations among these response variables. It should also be

TABLE 2

PRODUCT-MOMENT CORRELATION COEFFICIENTS AMONG MEAN *P*, MEAN *CC*, MEAN *NC*, AND LOG *SO* (LAG *O*) ON THREE BLOCKS OF SEVEN TRIALS FOR 32 SUBJECTS

Variables	Blocks of trials		
	Trials 2–8	Trials 9–15	Trials 16–22
Log *SO* and mean *P*	+.506	+.722	+.843
Log *SO* and mean *CC*	+.584	+.769	+.862
Log *SO* and mean *NC*	−.075	−.783	−.839
Mean *CC* and mean *NC*	−.018	−.881	−.943
Mean *CC* and mean *P*	+.948	+.986	+.988
Mean *NC* and mean *P*	+.295	−.792	−.882

remembered that correlations between P on the one hand and CC and NC on the other hand are, in a sense, spurious correlations, since CC and NC are subscores of P. These correlations, as well as some others, are included to complete the table. Since $P = CC + NC$, multiple correlations among P, CC, and NC are necessarily unity.

Considering the initial block of trials only, we note first that the correlation between CC and NC is essentially zero. Subjects' "ability" of retaining items from one trial to the next does not seem to be related to their ability of retaining items within a trial.

The second observation of interest in Table 2 concerns the positive and significant correlation between CC and log SO, and the absence of any notable correlation between NC and log SO, in the initial block of seven trials. For any given subject, therefore, intratrial retention is independent of the amount of organization that the subject imposes on the material, while intertrial retention seems to be directly related to SO.

At the empirical level, SO refers to the subjects' tendency to recall certain items in close temporal contiguity to one another. At the conceptual level, this tendency can be thought to represent the formation and existence of higher-order memory units. It is as if the list items—all already in the memory storage prior to the experiment, as was argued earlier—are rearranged in the storage in the course of trial-by-trial practice. Such rearrangement manifests itself and can be described in a variety of ways—development of associations of the type that define the associative meaning of a word (Deese, 1962); clustering in terms of conceptual (Bousfield, 1953; Cohen, 1963), associative (Jenkins & Russell, 1952), or synonymic (Cofer, 1959) categories; chunking, unitization, or recoding as envisaged by Miller (1956a, 1956b); construction of a plan, or creation of a hierarchical structure (Miller, Galanter, & Pribram, 1960); employment of various "mnemonic aids" as described, for instance, by Balaban (1910) and Bugelski (1962); ordering of items in recall according to a previously learned code such as the alphabet (Tulving, 1962b); and probably many others. Subjective organization is just a general name for all of these processes. To the extent that the higher-order memory units that result from organization are not specified experimentally and to the extent that the nature of the units is determined by the subject's previous experience, it might be justifiable to refer to these units as subjective units (S units). The functional significance of the development of these units lies in the increased accessibility of individual items constituting a unit. An item can be retrieved on its own merits, or through other items in its higher-order S unit.

If we accept the assumption that the measure of SO reflects the extent to which initially "unrelated" list items are formed into higher-order S units, it becomes possible to regard the increasing intertrial retention as reflecting nothing more or less than the increasing size of S units, and to entertain the hypothesis that the number of S units mediated by intertrial retention remains essentially constant over trials. This hypothesis agrees well with Miller's (1956a, 1956b) conception of memory system that is limited by the number of units that can be retrieved in succession without intervening external instructions, but not by the information content of these units. There is already some evidence that large differences in first-trial free recall become very much smaller or even disappear completely when retention is measured in terms of S units rather than in terms of grammatically designated units such as individual words (Tulving & Patkau, 1962). The hypothesis proposed here to account for the observed intertrial retention simply suggests that handling of information over successive trials in a verbal learning experiment is governed by the same laws of invariance as is handling of information within a single trial.

SUMMARY

1. A subject attempting to memorize a list of previously integrated items under the conditions of a multitrial free recall experiment "learns" every single item at the time of its presentation. Because of rapid intratrial forgetting only some items can be recalled in the output phase of the first trial. The number of such items represents a measure of intratrial retention.

2. Recall on the second trial, as well as on all subsequent trials, can be analyzed into two components—intratrial and intertrial retention. Intratrial retention refers to retention of items from the input phase of the trial, intertrial retention refers to retention of items from the output phase of the previous trial.

3. The traditional learning curve can be expressed as an additive function of intratrial and intertrial retention curves plotted against trials. For many samples of data, one of which was considered in detail in the paper, intertrial retention increases as a logarithmic function of trials, whereas intratrial retention decreases as a linear function of trials. The logarithmic learning curve of the form, $P = a \log n - bn + c$, not only describes the recall performance quite adequately, but also makes explicit the two components of performance, $a \log n$ representing intertrial retention and $c - bn$ representing intratrial retention.

4. Although the intratrial retention component of performance has been observed to decrease as a function of trials, it can be argued, on the basis of several kinds of evidence, that this decrease is an artifact attributable to limited list length used in most experiments, and that in an unconstrained task involving memorization of a list of words, intratrial retention is invariant at all stages of practice.

5. Supported by the finding that intertrial retention is positively correlated with the amount of organization that subjects impose on their recall, the hypothesis was advanced that the increase in intertrial retention reflects the increase in the size, but not the number, of the higher-order subjective units of material which can be retrieved from storage.

REFERENCES

BALABAN, A. Über den Unterschied des logischen und mechanischen Gedächtnisses. *Z. Psychol.*, 1910, **56**, 356–377.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

BROADBENT, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

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

BRUNER, J. S., MILLER, G. A., & ZIMMERMAN, CLAIRE. Discriminative skill and discriminative matching in perceptual recognition. *J. exp. Psychol.*, 1955, **49**, 187–192.

BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 409–412.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

COFER, C. N. A study of clustering in free recall based on synonyms. *J. gen. Psychol.,* 1959, **60**, 3–10.

COHEN, B. H. An investigation of recoding in free recall. *J. exp. Psychol.,* 1963, **65**, 368–376.

CONRAD, R., & HILLE, BARBARA A. The decay theory of immediate memory and paced recall. *Canad. J. Psychol.,* 1958, **12**, 1–6.

DEESE, J. On the structure of associative meaning. *Psychol. Rev.,* 1962, **69**, 161–175.

EBBINGHAUS, H. *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie.* Leipzig: Duncker & Humbolt, 1885.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.,* 1960, **67**, 207–223.

ESTES, W. K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika,* 1961, **26**, 73–84.

FEIGENBAUM, E. A., & SIMON, H. A. A theory of the serial position effect. *Brit. J. Psychol.,* 1962, **53**, 307–320.

HOROWITZ, L. M. Free recall and ordering of trigrams. *J. exp. Psychol.,* 1961, **62**, 51–57.

JENKINS, J. J., & RUSSELL, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.,* 1952, **47**, 818–821.

MANDLER, G. Response factors in human learning. *Psychol. Rev.,* 1954, **61**, 235–244.

MANDLER, G. Stimulus variables and subject variables: A caution. *Psychol. Rev.,* 1959, **66**, 145–149.

MCGEOCH, J. A. *The psychology of human learning.* New York: Longmans, 1942.

MILLER, G. A. Human memory and the storage of information. *IRE Trans. Inform. Theory,* 1956, **2**, 129–137. (a)

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)

MILLER, G. A., GALANTER, E., & PRIBRAM, K. *Plans and the structure of behavior.* New York: Holt, 1960.

MILLER, G. A., & MCGILL, W. J. A statistical description of verbal learning. *Psychometrika,* 1952, **17**, 369–396.

MURDOCK, B. B., JR. The immediate retention of unrelated words. *J. exp. Psychol.,* 1960, **60**, 222–234.

MURDOCK, B. B., JR. Short-term retention of single paired associates. *Psychol. Rep.,* 1961, **8**, 280. (a)

MURDOCK, B. B., JR. The retention of individual items. *J. exp. Psychol.,* 1961, **62**, 618–625. (b)

MURDOCK, B. B., JR. Short-term retention of single paired associates. *J. exp. Psychol.,* 1963, **65**, 433–443.

MURDOCK, B. B., JR., & BABICK, A. J. The effect of repetition on the retention of individual words. *Amer. J. Psychol.,* 1961, **74**, 596–601.

PETERSON, L. R., & PETERSON, MARGARET J. Short-term retention of individual verbal items. *J. exp. Psychol.,* 1959, **58**, 193–198.

PETERSON, L. R., PETERSON, MARGARET J., & MILLER, A. Short-term retention and meaningfulness. *Canad. J. Psychol.,* 1961, **15**, 143–147.

PETERSON, L. R., SALTZMAN, DOROTHY, HILLNER, K., & LAND, VERA. Recency and frequency in paired-associate learning. *J. exp. Psychol.,* 1962, **63**, 396–403.

POSTMAN, L. One-trial learning. In C. N. Cofer & Barbara S. Musgrave (Eds.), *Verbal behavior and verbal learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 295–321.

ROCK, I. The role of repetition in associative learning. *Amer. J. Psychol.,* 1957, **70**, 186–193.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

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

TULVING, E. The effect of alphabetical subjective organization on memorizing unrelated words. *Canad. J. Psychol.,* 1962, **16**, 185–191. (b)

TULVING, E., & ARBUCKLE, TANNIS Y. Sources of intra-trial interference in immediate recall of paired associates. *J. verbal Learn. verbal Behav.,* 1963, **1**, 321–334.

TULVING, E., & PATKAU, JEANNETTE E. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canad. J. Psychol.,* 1962, **16**, 83–92.

UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.,* 1954, **51**, 276–282.

UNDERWOOD, B. J., & KEPPEL, G. One-trial learning? *J. verbal Learn. verbal Behav.,* 1962, **1**, 1–13.