MATHEMATICAL MODELS FOR MEMORY AND LEARNING

by

R. C. Atkinson

and

R. M. Shiffrin

TECHNICAL REPORT NO. 79

September 20, 1965

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

MATHEMATICAL MODELS FOR MEMORY AND LEARNING*

by

R. C. Atkinson and R. M. Shiffrin

Stanford University

In recent years a number of models have been proposed to account for retention phenomena, with the emphasis primarily on short-term memory experiments. There has also been an active development of models for verbal learning, with the focus on experiments dealing with serial and paired-associate learning. Except for a few notable exceptions, most of these theoretical developments have been applicable either to memory or learning experiments, and no attempt has been made to bridge the gap. It is our feeling that theoretical and experimental work in these two areas is sufficiently well advanced to warrant the development of a general theory that encompasses both sets of phenomena. This, then, is the goal of the paper. We must admit, however, that the term "general theory" may not be entirely appropriate, for many features of the system are still vague and undefined. Nevertheless, the work has progressed to a point where it is possible to use the general conceptual framework to specify several mathematical models

*This paper was prepared for the "Third Conference on Learning, Remembering, and Forgetting" sponsored by the New York Academy of Science at Princeton, New Jersey, October 3 to 6, 1965. Support for the research was provided by the National Aeronautics and Space Administration, Grant No. NGR-05-020-036. The authors also wish to acknowledge their indebtedness to Gordon Bower who, in discussion, contributed substantially to many of the ideas presented in this paper.

that can be applied to data in quantitative detail.

The theory that we shall outline postulates a distinction between short-term and long-term memory systems; this distinction is based on the coding format used to represent information in the two systems, and on the conditions determining the length of stay. In addition, two process variables are introduced: a transfer process and a retrieval process. The transfer process characterizes the exchange of information between the two memory systems; the retrieval process describes how the subject recovers information from memory when it is needed. As one might conjecture from this brief description, many of the ideas that we will examine have been proposed by other theorists. In particular we have been much influenced by the work of Bower (1964), Broadbent (1963), Estes (1965), Feigenbaum and Simon (1962), and Peterson (1963). However, we hope we have added to this earlier work by applying some of the ideas in quantitative form to a wider range of phenomena.

In presenting the theory we shall begin with an account of the various mechanisms involved, making only occasional references to experimental applications. Only later will models be developed for specific experimental paradigms and applied to data. Thus the initial description will be rather abstract, and the reader may find it helpful to keep in mind the first study to be analyzed. This experiment deals with short-term memory, and involves a long series of discrete trials. On each trial a new display of stimuli is presented to the subject. A display consists of a random sequence of playing cards; the cards vary only in the color of a small patch on one side. The cards are presented at a fixed rate, and the subject names the color of each card as it is presented. Once the card has been named it is turned face down so that the color is no longer visible, and the next card is presented. After presentation of the last card in a display the experimenter points to one of the cards, and the subject must try to recall its color. Over the series of trials, the length of the display and the test position are systematically varied. One goal of a theory in this case is to predict the probability of a correct response as a function of both list length and test position. With this experiment in mind we now turn to an account of the theory.

GENERAL FORMULATION OF THE BUFFER MODEL

In this section the basic model will be outlined for application later to specific experimental problems. Figure 1 shows the overall conception. An incoming stimulus item first enters the sensory buffer where it will reside for only a brief period of time and then is transferred to the memory buffer. The sensory buffer characterizes the initial input of the stimulus item into the nervous system, and the amount of information transmitted from the sensory buffer to the memory buffer is assumed to be a function of the exposure time of the stimulus and related variables. Much work has been done on the encoding of short-duration stimuli (e.g., see Estes and Taylor, 1964; Mackworth, 1963; Sperling, 1960), but all of the experiments considered in this paper are concerned with stimulus exposures of fairly long duration (one second or more). Hence we will assume that all items pass successfully through the sensory buffer and into the memory buffer; that is, all items are assumed to be attended to and entered correctly into the memory buffer. Throughout this paper, then, it will be understood that the term buffer refers to the memory buffer and not the sensory buffer. Furthermore, we will not become involved here in a

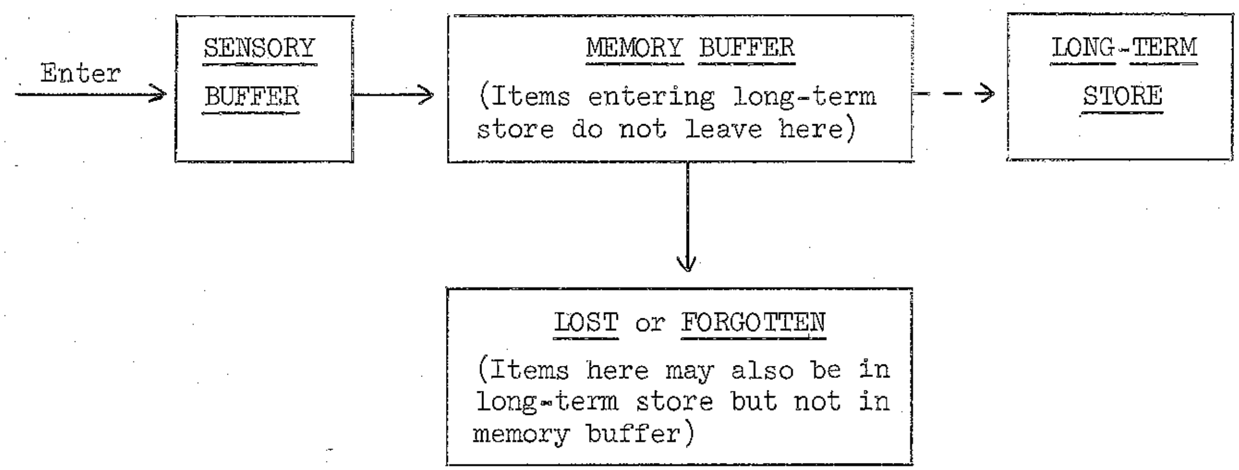

Fig. 1. Flow chart for the general system.

detailed analysis of what is meant by an "item." If the word "horse" is presented visually, we will simply assume that whatever is stored in the memory buffer (be it the visual image of the word, the auditory sound, or some vector of information about horses) is sufficient to permit the subject to report back the word "horse" if we immediately ask for it. This question will be returned to later. Referring back to Fig. 1, we see that a dotted line runs from the buffer to the "long-term store" and a solid line from the buffer to the "lost or forgotten" state.* This is to emphasize that items are copied into LTS without affecting in any way their status in the buffer. Thus items can be simultaneously in the buffer and in LTS. The solid line indicates that eventually the item will leave the buffer and be lost. The lost state is used here in a very special way: as soon as an item leaves the buffer it is said to be lost, regardless of whether it is in LTS or not. The buffer, it should be noted, is a close correlate of what others have called a "short-term store" (Bower, 1964; Broadbent, 1963; Brown, 1964; Peterson, 1963) and "primary memory" (Waugh and Norman, 1965). We prefer the term buffer because of the wide range of applications for which the term short-term store has been used. This buffer will be assigned very specific properties in the following section. Later on, the features of LTS will be considered, but with less specificity than those of the buffer.

A. THE MEMORY BUFFER

Certain basic properties of the buffer are diagrammed in Fig. 2. They are as follows:

---

*The term long-term store will be used throughout the paper and hence abbreviated as LTS.

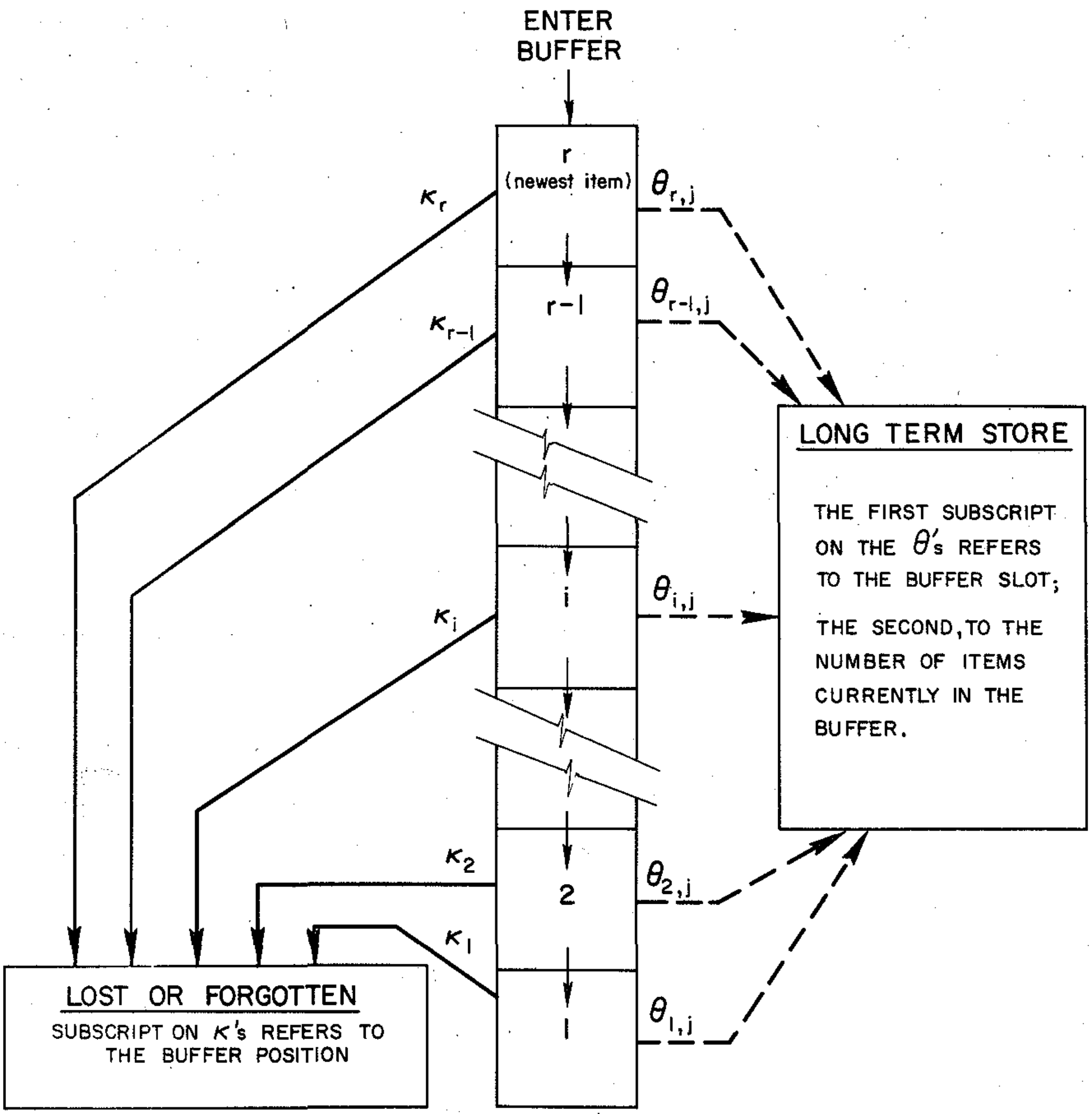

Fig. 2. Flow chart for the memory buffer.

1) Constant size. The buffer can contain exactly r items and no more. We start by supposing that items refers to whatever is presented in the experiment in question, whether it be a paired-associate, a 6-digit number, or a single letter. Thus, for each experimental task the buffer size must be estimated. Hopefully in future work it will be possible to specify the parameter r in advance of the experiment by considering physical characteristics of the stimulus items. For the present, no contradiction arises in these two approaches if we remember that stimulus items for any given experiment are usually selected to be quite homogeneous, and can be roughly assumed to carry equal information. It would be expected that the more complicated the presented item, the smaller r would be. Similarly, the greater the number of alternatives that each presented item is chosen from, the smaller r should be.

2) Push-down buffer: temporal ordering. These two properties are equivalent. As it is shown in the diagram the spaces in the buffer (henceforth referred to as "slots") are numbered in such a way that when an item first enters the buffer it occupies the r^{th} slot. When the next item is presented it enters the r^{th} slot and pushes the preceding item down to the $r-1^{st}$ slot. The process continues in this manner until the buffer is filled; after this occurs each new item pushes an old one out on a basis to be described shortly. The one that is pushed out is lost. Items stored in slots above the one that is lost move down one slot each and the incoming item is placed in the r^{th} slot. Hence items in the buffer at any point in time are temporally ordered: the oldest is in slot number 1 and the newest in slot r.

3) Buffer stays filled. Once the first r items have arrived the buffer is filled. Each item arriving after that knocks out exactly one item already in the buffer; thus the buffer is always filled thereafter. It is assumed that this state of affairs continues only as long as the subject is paying attention and trying to remember all that he can. At the end of a trial for example, attention ceases and the buffer gradually empties of that trial's items. Whether the items in the buffer simply fade out on their own or are knocked out by miscellaneous succeeding material is a moot point. In any event the buffer is cleared of the old items by the start of the next trial. The important point, therefore, is the focus of attention. Though the buffer may be filled with other material at the start of a trial, primacy effects are found because attention is focused solely on the incoming items.

4) Each new item bumps out an old item. This occurs only when the buffer has been filled. The item to be bumped out is selected as a function of the buffer position (which is directly related to the length of time each item has spent in the buffer). Let

κ_j = probability that an item in slot j of a full buffer is lost when a new item arrives.

Then of course $\kappa_1 + \kappa_2 + \ldots + \kappa_r = 1$, since exactly one item is lost. Various schemes can be proposed for the generation of the κ_j's. The simplest scheme (which requires no additional parameters) is to equalize the κ's; i.e., let $\kappa_j = 1/r$ for all j. A useful one-parameter scheme will be described in some detail later on. In general, we would expect the smaller the subscript j, the larger κ_j; that is, the longer the item has been in the buffer the higher the probability of its being lost. The extent of this effect would depend in each experiment upon such things as the tendency toward serial rehearsing, whether or not the subject can anticipate the end of the list, and so on. Once an item has been bumped out of the buffer it cannot be recalled at a later time unless it has previously entered LTS.

5) Perfect representation of items in the buffer. Items are always encoded correctly when initially placed in the buffer. This, of course, only holds true for experiments with slow enough inputs, such as those considered in this paper. This postulate would have to be modified if items entered very quickly; the modification could be accomplished by having an encoding process describing the transfer of information from the sensory buffer to the memory buffer.

6) Perfect recovery of item from the buffer. Items still in the buffer at the time of test are recalled perfectly (subject to the "perfect representation" assumption made above). This and the previous assumption are supported by certain types of digit-span experiments where a subject will make no mistakes on lists of digits whose lengths are less than some critical value.

7) Buffer is unchanged by the transfer process. The contents of the buffer are not disturbed or otherwise affected by the transfer of items from the buffer to LTS. Thus an item transferred into LTS is still represented in the buffer. The transfer process can be viewed as one of copying an item in the buffer, and placing it in LTS, leaving the contents of the buffer unchanged.

This set of seven assumptions characterizes the memory buffer. Next we shall consider the transfer process which moves items out of the buffer into LTS, but before we do this let us examine a simple one-parameter scheme for generating the κ_j's.

We want the probability that the j^{th} item in a full buffer is the one lost when a new item enters. The following process is used to determine which item is dropped: the oldest item (in slot 1) is dropped with probability δ. If that item is <u>not</u> dropped, then the item in position 2 is dropped with probability δ. If the process reaches the r^{th} slot and it also is passed over, then the process recycles to the 1^{st} slot. This process continues until an item is dropped. Hence

$$\kappa_j = \delta(1-\delta)^{j-1} + \delta(1-\delta)^{r+j-1} + \delta(1-\delta)^{2r+j-1} + \delta(1-\delta)^{3r+j-1} + \cdots$$

$$= \frac{\delta(1-\delta)^{j-1}}{1-(1-\delta)^r} . \qquad (1)$$

If we expand the denominator in the above equation and divide top and bottom by δ it is easy to see that κ_j approaches $1/r$ for all j as δ approaches zero. Thus, this limiting case represents a bump-out process where all items in the buffer have the same likelihood of being lost. When $\delta = 1$, on the other hand, $\kappa_1 = 1$ and $\kappa_2 = \kappa_3 = \cdots = \kappa_r = 0$; i.e., the oldest item is always the one lost. Figure 3 illustrates what this process is like. What is graphed is a recency curve; the probability that the i^{th} item from the end of the list is still in the buffer at the time of test. The last item presented is the leftmost point and of course is always 1 since there are no additional items to bump it out. The line labeled $\delta = 1$ represents the case where the oldest item is lost each time. In this case the last r items presented are all still in the buffer at the time of test; no older item is present however. The line labeled $\delta \to 0$

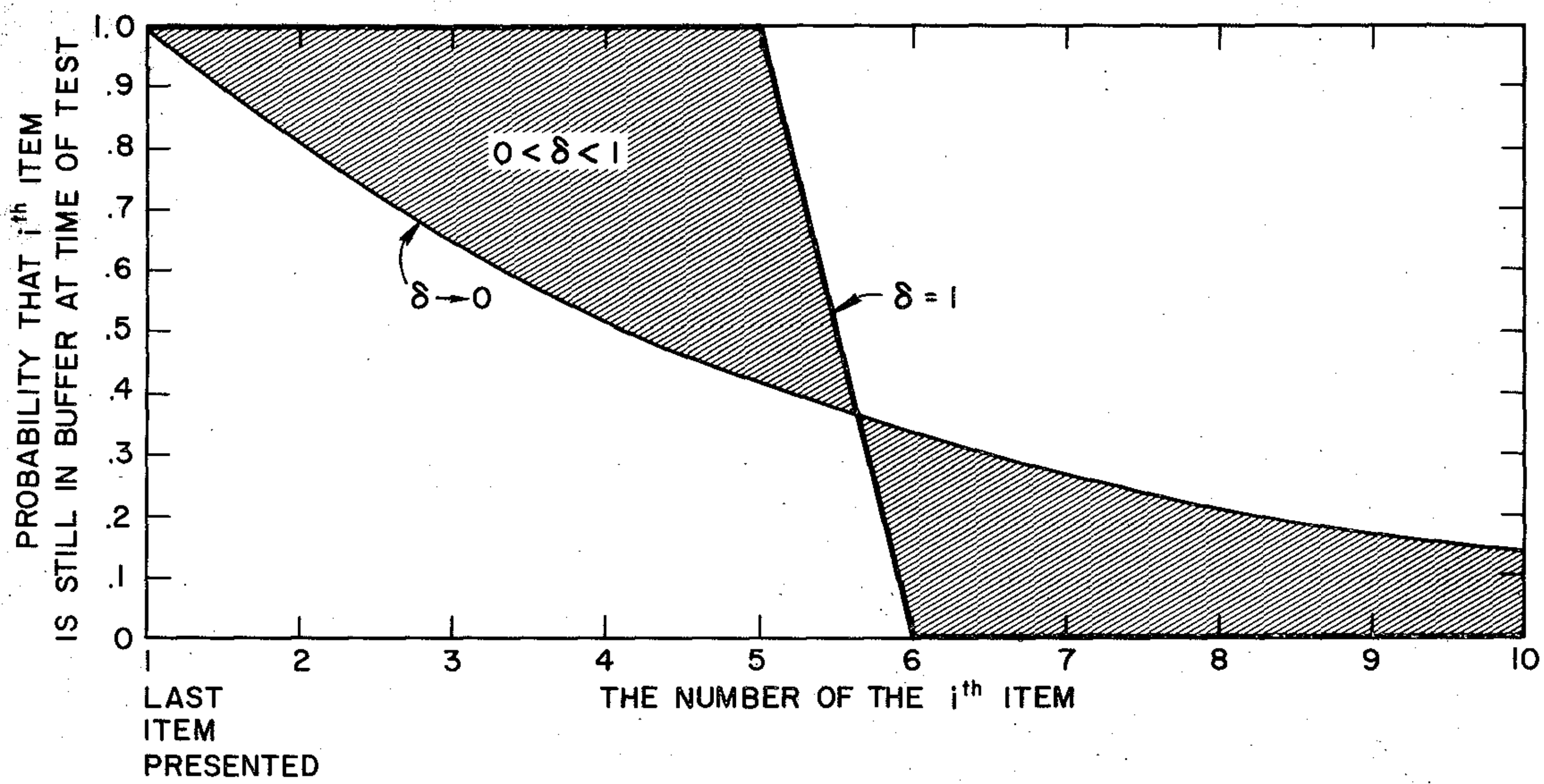

Fig. 3. Recency curves as a function of δ (the functions are computed for $r = 5$).

shows the case when the bump-out probabilities are all equal. This curve is a simple geometric function, since the probability that any item will still be in the buffer when n items follow is $(\frac{r-1}{r})^n$. The shaded region indicates the range in which the recency function must lie for $0 < \delta < 1$. Hence, depending upon the value of δ, either S-shaped or exponential curves can be obtained.

B. THE TRANSFER PROCESS TO LONG-TERM STORE

For now it will suffice to say that the transfer process involves making copies of items in the buffer and then placing them in LTS. Later we will want to think of each item as a mosaic of elements and to view a copy as either a complete or partial representation of the array. Thus the transfer process can be thought of as all-or-none if the initial copy is complete, and incremental if each copy is incomplete and the item's accurate representation in LTS depends on an accumulation of partial copies.

We shall let θ_{ij} be the transfer parameter. In particular θ_{ij} is the probability that an item in the i^{th} plot of the buffer is copied into LTS between one item presentation and the next if there are j items in the buffer during this period. The parameter θ_{ij} thus depends on the number of items currently in the buffer and on the buffer slot. It also depends on the buffer size, the rate at which items are input into the buffer, and such things as the complexity and codability of the items.

C. THE LONG-TERM STORE

The question, "What is stored in long-term memory?" is basic to the theory, and we shall be more flexible in considering it than we were in laying down the postulates for the buffer. A number of different models will be developed in the paper and several more proposed. The first viewpoint, and the simplest, holds that:

1) Items are represented in an all-or-none fashion no more than once in LTS.

In this case the parameter θ_{ij} represents the probability of placing a copy of an item in LTS; once a copy has been placed in LTS no further copies of that item are made. A variation of this version is:

2) Items are represented in LTS by as many copies as were made during the time the item was in the buffer.

In this case θ_{ij} is the same as before except that the process does not end when the first copy is made. (Looking ahead a bit, we note that a simple retrieval scheme, such as perfect recall of all items in LTS will not differentiate between 1 and 2. This is, of course, not the case for more elaborate schemes.) Cases 1 and 2 will be called the "single-copy" and "multiple-copy" schemes, respectively. If the all-or-none assumption is now removed from the multiple-copy scheme we have:

3) Items are represented by partial copies, the number of partial copies being a function of the time spent in the buffer. One partial copy will allow recall with probability less than one.

If items are again viewed as information arrays, then each partial copy can be viewed as a sample from the array characterizing that item. With a partial copy the subject may be able to recognize an item previously presented, even though he cannot recall it. Processes of this type will be considered in greater detail later in the paper. Case 3 leads to its continuous counterpart (the strength postulate):

4) Each item is represented by a strength measure in LTS, the strength being a function of the amount of time the item was in the buffer.

For both cases 3 and 4, θ_{ij} is best considered as a rate parameter. These various storage schemes naturally lead to the question of recall or retrieval from LTS.

D. <u>RETRIEVAL OF ITEMS FROM MEMORY</u>

1) <u>Retrieval from the buffer</u>. Any item in the buffer is recalled perfectly (given that it was entered correctly in the buffer).

2) <u>Retrieval from the lost state</u>. No item can be recalled from this state. It must be noted, however, that an item can be in this state and also in LTS. Thus an item that has been lost from the buffer can be recalled only if it has been previously entered in LTS. If an item is in neither LTS nor the buffer, then the probability of making a correct response is at a guessing level.

3) <u>Retrieval from LTS</u>. Each storage process mentioned in the previous section would, of course, have its own retrieval scheme. Later we will propose retrieval postulates for each storage process, but for now the topic will be considered more generally.

In order to place the problem in perspective, consider the free verbal recall data of Murdock (1962) which is shown in Fig. 4. The experimental situation consists of reading a list of words to a subject and immediately afterward having him write down every word he can remember. The graph shows the probability of recalling the word presented in position i for lists of various lengths and input rates. The two numbers appended to each curve denote the list length and the presentation time in seconds for each word.

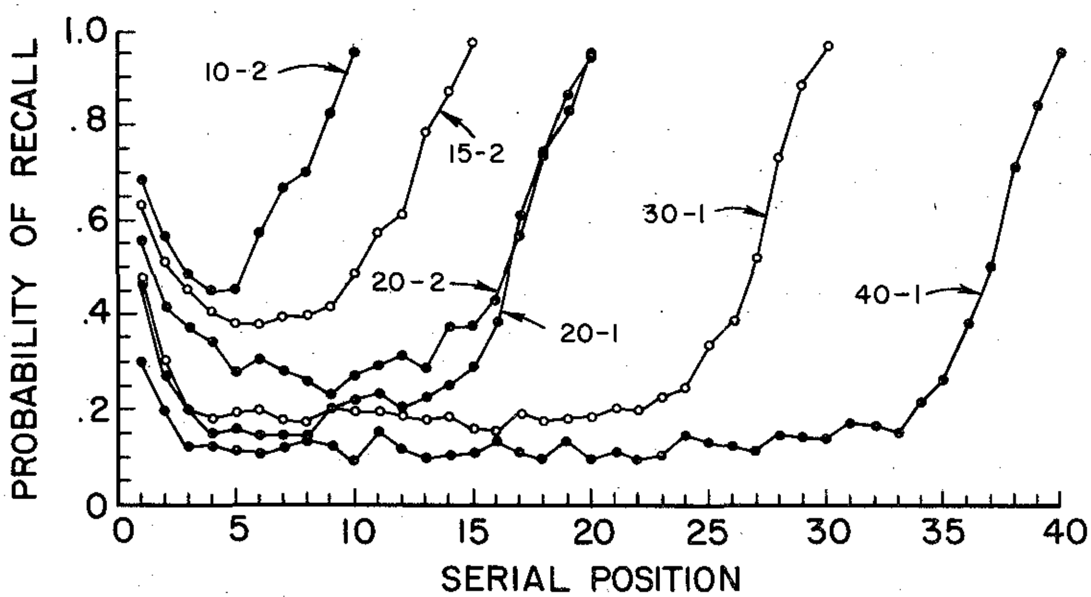

Fig. 4. Serial position curves for free verbal recall (after Murdock, 1962).

In particular consider the data for lists of 30 and 40 items. The first items in the list (the oldest items) are plotted to the left and exhibit a primacy effect; i.e., the probability of recall is higher for these than for the middle items. The last items are plotted to the right and exhibit the recency effect; i.e., the probability of recall is higher for these also. Most important for present purposes is the response level for items in the middle of each list; note particularly the drop in the probability of recall for these items from the 30 to the 40 list. Specifically, why are the middle items in the 30 list recalled more often than the middle items in the 40 list? The effect itself seems reliable since it will be given corroborating support in similar experiments to be reported later. Furthermore, the effect appears intuitively to be what one would expect. For example, imagine presenting lists of lengths 10, 20, 1000, etc. It is obvious that the probability of recalling items in the middle of a list is going to tend to the guessing level as list length increases indefinitely, but what is there in the theory to predict this occurrence?

Two different answers to this question suggest themselves. The historical answer is that of interference. Each item placed in LTS interferes somewhat with each succeeding item placed there (proactive interference), and each item placed in LTS interferes somewhat with each item already there (retroactive interference). The other answer that suggests itself is that retrieval from LTS is less effective as the number of items in LTS increases. In particular we can view the retrieval process as a search of LTS that occurs at the moment of test (we will assume that the search does not take place if the item is in the buffer at the time of test--in that case the item is reported out quickly and perfectly). The notion of a search process is not new. For some time workers in the area of perception and psychophysics have been employing such schemes (e.g., Estes and Taylor, 1964; and Sperling, 1960). Sternberg has presented a search theory based on memory reaction time studies (1963), and Yntema and Trask (1963) have proposed a search scheme for recall studies. In many experimental tasks it is intuitively clear that the subject engages in an active search process and often can verbalize his method (Brown and McNeill, 1966).

Without yet fixing on a specific scheme, two possibilities can be considered under the heading of search processes. First, there can be a destructive process in which each search into LTS disrupts the contents of the store, and second, there can be a stopping rule so that the search may stop before an item actually in LTS is found. Using either of these processes or some combination, the drop in recall probability as list length increases can be explained.

While not denying that an interference theory may be a viable way of explaining certain data, we have decided for several reasons to restrict ourselves to search theories in this paper. First, it is obvious that some manner of search process must be present in most memory experiments. Second, an interference process seems to require a more exact specification of just what is stored than a search theory. Third, a search theory gives a natural interpretation of reaction time data.

Two representative retrieval schemes may now be proposed:

a. The subject makes R searches in LTS and then stops. If there are n items in LTS, then it is assumed that on each search the subject has probability $1/n$ of retrieving the item. Thus, the probability of correctly recalling an item stored only in LTS is

$$1 - \left(1 - \frac{1}{n}\right)^{R} .$$

For greater generality it could be assumed that the number of searches made has a distribution with mean R.

b. On each search the subject samples randomly and with replacement from among the items in LTS. He continues to search until the item is found. Each search, however, may disrupt the looked-for item with probability R', and hence when it is finally found the subject may be unable to reproduce it.

It should be noted that these retrieval schemes are strictly applicable only to a storage process where each item is stored once and only once in an all-or-none fashion. The schemes would have to be modified to be applied to a multiple-copy or a strength process. The central consideration in this regard is the probability of a hit, denoted h_i, which is the probability that the desired item i will be found in a single search. In the single-copy scheme $h_i = n^{-1}$ if there are n items in the store. In the multiple-copy scheme $h_i = n_i/\Sigma n_j$ where n_j is the number of copies of item j. In the strength scheme if the i^{th} item has strength λ_i then $h_i = \lambda_i/\Sigma\lambda_j$. These more complicated schemes will be treated in detail as they occur.

APPLICATION OF MODEL TO SHORT-TERM MEMORY EXPERIMENT

Enough general features of the buffer model have been presented to make it possible to apply certain special cases to data. Consequently, we will now analyze a study reported by Phillips and Atkinson (1965).

The experiment involved a long series of discrete trials. On each trial a display of items was presented. A display consisted of a series of cards each containing a small colored patch on one side. Four colors were used: black, white, blue, and green. The cards were presented to the subject at a rate of one card every two seconds. The subject named the color of each card as it was presented. Once the color of the card had been named by the subject it was placed face down on a display board so that the color was no longer visible, and the next card was presented. After presentation of the last card in a display the cards were in a straight row on the display board: the card presented first was to the subject's left and the most recently presented card to her right. The trial terminated when the experimenter pointed to one of the cards on the display board, and the subject attempted to recall the color of that card. The subject was instructed to guess the color if uncertain and to qualify her response with a confidence rating. The confidence ratings were the numerals 1, 2, 3, and 4. The subjects were told to say 1 if they were positive; 2 if they were choosing from two alternatives, one of which they were sure was correct; 3 if they were choosing from three alternatives, one of which they were sure was correct; and 4 if they had no idea at all as to the correct response.

Following the subject's confidence rating, the experimenter informed the subject of the correct answer. The display size (list length) will be denoted as d. The values of d used in the experiment were 3, 4, 5, 6, 7, 8, 11, and 14. Each display, regardless of size, ended at the same place on the display board, so that the subject knew at the start of each display how long that particular display would be. Twenty subjects, all females, were run for a total of five sessions, approximately 70 trials per session.

Figure 5 presents the proportion of correct responses as a function of the test position in the display. There is a separate curve for each of the display sizes used in the study. Points on the curves for $d = 8, 11,$ and 14 are based on 120 observations, whereas all other points are based on 100 observations. Serial position 1 designates a test on the most recently presented item. These data indicate that for a fixed display size, the probability of a correct response decreases to some minimum value and then increases. Thus there is a very powerful recency effect as well as a strong primacy effect over a wide range of display sizes. Note also that the recency part of each curve is S-shaped and could not be well described by an exponential function. Reference to Fig. 5 also indicates that the overall proportion correct is a decreasing function of display size.

MODEL I (PERFECT RETRIEVAL OF ITEMS IN LTS)

We shall begin our analysis of these data using an extremely simple form of the buffer model. The buffer will be specified in terms of postulates A-1 through A-7, along with the time-dependent bump-out process of Eq. 1. The LTS assumptions are those indicated in C-1; i.e., each item in the list is stored possibly once and no more than once in LTS. The transfer function also will be simplified by assuming that transfer of any item in the buffer to LTS depends only on the number of items currently in the buffer. Thus the first subscript on the θ_{ij} function defined earlier will be

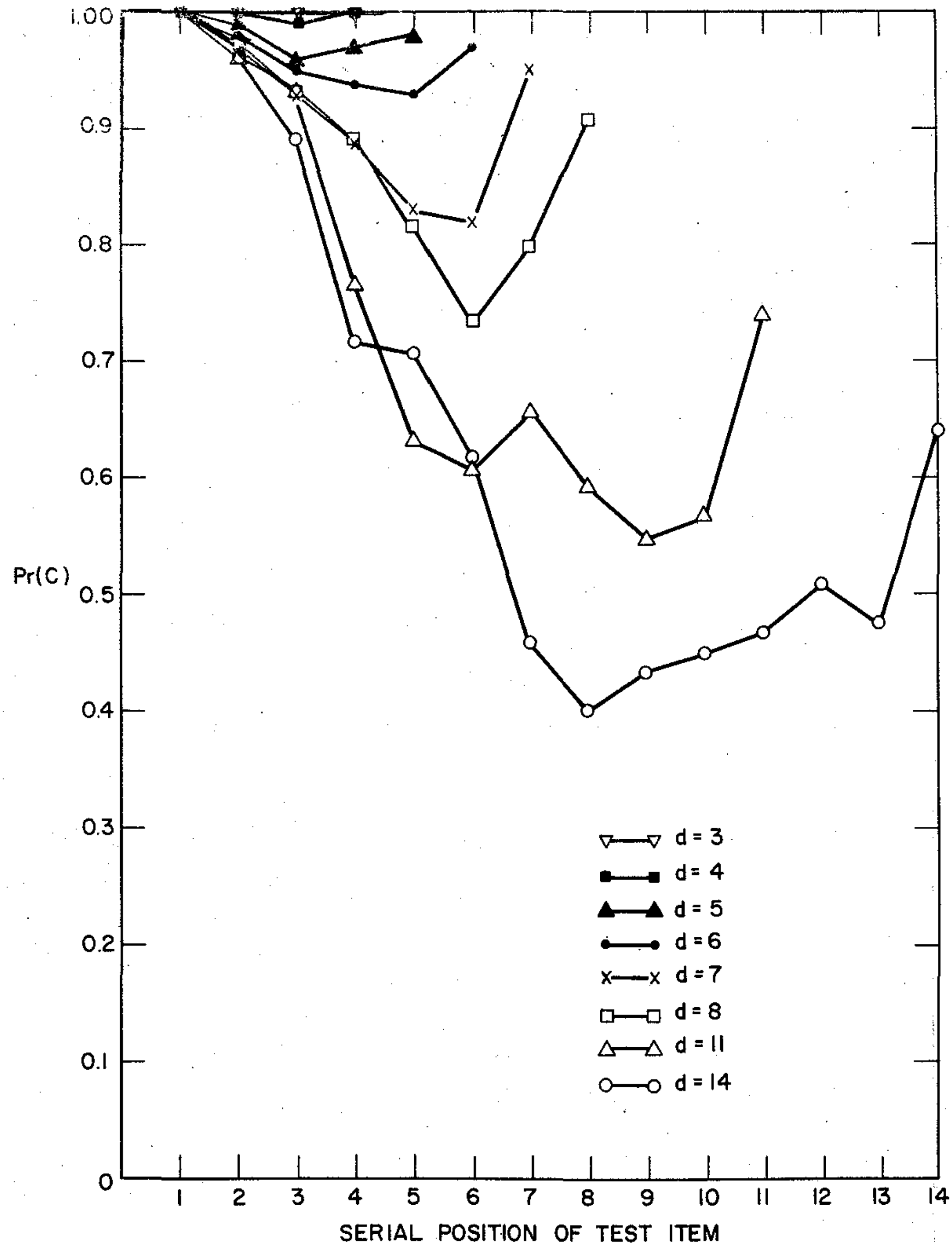

Figure 5. Proportion of items correctly recalled at each serial position for the various display sizes.

dropped, and θ_j will denote the probability that any item in the buffer will be copied into LTS between presentations of successive items, given that there are j items in the buffer during that period. Further, we will assume that

$$\theta_j = \frac{\theta}{j}$$

where θ is an arbitrary parameter between 0 and 1. This assumption is justified by the following considerations: if in each small unit of time the subject attends to just one of the items in the buffer, and if over many of these small units of time the subject's attention switches randomly among the j items currently in the buffer, then the amount of time spent attending to any given item will be linearly proportional to j. We use this argument to justify setting $\theta_j = \theta/j$, but we recognize the arbitrariness of the assumption and later will examine other schemes.

The last feature to be specified is the retrieval scheme. In Model I we will assume simply that any item in the LTS is retrieved correctly with probability 1. Hence the probability of a correct response for an item stored in either the buffer or LTS is 1. The probability of a correct response for an item in neither the buffer nor LTS is the guessing probability, which will be set equal to 1/4 since there were four response alternatives in the experiment.

Mathematical Development of Model I

We begin by defining the following quantities:

$f_i^{(d)}$ = probability that item i in a display of size d is neither in the buffer nor in LTS at the time of test.

$s_i^{(d)}$ = probability that item i in a display of size d is in the buffer at the time of test.

$\ell_i^{(d)}$ = probability that item i in a display of size d is in LTS and not in the buffer at the time of test.

Of course, $f_i^{(d)} + \ell_i^{(d)} + s_i^{(d)} = 1$. It should be emphasized that in our analysis of this experiment, position i denotes items counted from the end of the list; i.e., the last item presented is number 1, the second to last number 2, etc.

In order to facilitate the derivation of expressions for this model, we define the quantity, φ_{ij}. Given that there are j items yet to be presented, φ_{ij} is the probability that an item currently in slot i, which has not yet entered LTS, will be neither in LTS nor in the buffer at the time of test. We note that for the first position of the register $(i = 1)$ these expressions are first-order difference equations of the form

$$\varphi_{1,j} = \kappa_1 + (1-\frac{\theta}{r})(1-\kappa_1)\varphi_{1,j-1} .$$

For $i \geq 2$ the expressions are somewhat more formidable:

$$\begin{aligned}
\varphi_{2,j} &= \kappa_2 + (1-\frac{\theta}{r})[\kappa_1\varphi_{1,j-1} + (\kappa_3 + \kappa_4 + \cdots + \kappa_r)\varphi_{2,j-1}] \\
\varphi_{3,j} &= \kappa_3 + (1-\frac{\theta}{r})[\kappa_1 + \kappa_2)\varphi_{2,j-1} + (\kappa_4 + \kappa_5 + \cdots + \kappa_r)\varphi_{3,j-1}] \\
&\vdots \\
\varphi_{i,j} &= \kappa_i + (1-\frac{\theta}{r})[\kappa_1 + \kappa_2 + \cdots + \kappa_{i-1})\varphi_{i-1,j-1} + (\kappa_{i+1} + \kappa_{i+2} + \cdots + \kappa_r)\varphi_{i,j-1}] \\
&\vdots \\
\varphi_{r-1,j} &= \kappa_{r-1} + (1-\frac{\theta}{r})[(\kappa_1 + \kappa_2 + \cdots + \kappa_{r-2})\varphi_{r-2,j-1} + \kappa_r\varphi_{r-1,j-1}] \qquad (2) \\
\varphi_{r,j} &= \kappa_r + (1-\frac{\theta}{r})(1-\kappa_r)\varphi_{r,j-1} .
\end{aligned}$$

The initial condition for each of these equations is $\varphi_{i,0} = 0$.

The equations above can be derived by the following argument. We want to specify φ_{ij} in terms of the φ's for $j-1$ succeeding items. Thus

φ_{ij} equals κ_i [the probability that the item in slot i is lost when the next item is presented] plus $1-\frac{\theta}{r}$ [the probability that the item does not enter LTS] times the quantity

$$\{(\kappa_1 + \kappa_2 + \cdots + \kappa_{i-1})\varphi_{i-1,j-1} + (\kappa_{i+1} + \kappa_{i+2} + \cdots + \kappa_r)\varphi_{i,j-1}\} .$$

But the quantity in brackets is simply $\kappa_1 + \kappa_2 + \cdots + \kappa_{i-1}$ [the probability that an item in a slot numbered less than i is lost which means that the item in slot i will move down to slot $i-1$] times $\varphi_{i-1,j-1}$ (since the item has moved to slot $i-1$ with $j-1$ items to be presented] plus $\kappa_{i+1} + \kappa_{i+2} + \cdots + \kappa_r$ [the probability that an item in a slot numbered greater than i is lost] times $\varphi_{i,j-1}$ [since the item is still in slot i with $j-1$ items to be presented].

The quantity $f_i^{(d)}$ may now be defined in terms of the φ_{ij}'s. It is clear that any item numbered less than $d-r+1$ will enter the buffer with all the slots filled. Thus, for $i \leq d-r+1$, $f_i^{(d)}$ equals $1-\frac{\theta}{r}$ [the probability of not entering LTS at once] times $\varphi_{r,i-1}$ [since after the i^{th} item there are $i-1$ still to come]. For $i > d-r+1$ we must consider the probability that the item stays in the buffer until it is full without entering LTS. Specifically, this probability is

$$(1-\frac{\theta}{r})(1-\frac{\theta}{r-1}) \cdots (1-\frac{\theta}{d-i+1}) = \prod_{j=d-i+1}^{r}(1-\frac{\theta}{j}) ,$$

at which time the item will be in slot $d-i+1$ of the buffer. Furthermore, there will now be $d-r$ items to come. Hence, for $i > d-r+1$, $f_i^{(d)}$ will simply be the above product multiplied by $\varphi_{d-i+1,d-r}$. Summarizing these results we have:

$$f_i^{(d)} = \begin{cases} \left[\prod_{j=d-i+1}^{r} \left(1-\frac{\theta}{j}\right)\right] \varphi_{d-i+1,d-r} & , \text{ for } i > d-r+1 \\ \left(1-\frac{\theta}{r}\right)\varphi_{r,i-1} & , \text{ for } i \leq d-r+1 . \end{cases} \quad (3)$$

Now let $C_i^{(d)}$ denote the event of a correct response to item i in a list of length d. Then

$$Pr[C_i^{(d)}] = 1 - f_i^{(d)} + f_i^{(d)}[\tfrac{1}{4}] , \quad (4)$$

where $1/4$ is the guessing probability and $1 - f_i^{(d)}$ is the probability that the item is either in the buffer, LTS, or both at the time of test.

The obvious next step would be to solve the various difference equations and thereby obtain an explicit expression for $Pr[C_i^{(d)}]$ as a function of the parameters θ, r, and δ. This is a straightforward but extremely tedious derivation. Rather than do this we have decided to use a computer to iteratively calculate values of φ_{ij} for each set of parameters θ, r, and δ we wish to consider.

For purposes of estimating parameters and evaluating the goodness-of-fit of data to theory, we now define the following chi-square function:

$$\chi^2(d) = \sum_{i=1}^{d-1} \left\{\frac{1}{NPr[C_i^{(d)}]} + \frac{1}{N - NPr[C_i^{(d)}]}\right\} \left\{NPr[C_i^{(d)}] - O_i^{(d)}\right\}^2 \quad (5)$$

where $O_i^{(d)}$ is the observed number of correct responses for the i^{th} item in a display of size d, and N is the total number of observations at each position of the display. (Recall that N was 120 for $D = 8, 11, 14$, and 100 for $d = 3, 4, 5, 6, 7$.) The sum excludes the first item (item d) because $1 - Pr[C_i^{(d)}]$ is predicted to be zero for all list lengths; this prediction is supported by the data.

Goodness-of-Fit Results for Model I

It seemed reasonable to estimate the parameter r on the basis of data from the short lists. The model predicts that no errors will be made until the display size d exceeds the buffer size. Extremely few errors were made for d's of 5 and less, and we will assume that these are attributable to factors extraneous to the main concern of the experiment. On this basis r would be 5; this estimate of r will be used in further discussions of this experiment.

The estimates of the parameters δ and θ were obtained by using a minimum χ^2 procedure. Of course, the minimization cannot be done analytically for we have not derived an explicit expression for $\Pr[C_i^{(d)}]$, and therefore we will resort to a numerical routine using a computer. The routine involves selecting tentative values of δ and θ, computing the associated $\Pr[C_i^{(d)}]$'s and the $\chi^2(d)$, repeating the procedure with another set of values for θ and δ, and continuing thus until the space of possible values on θ and δ $[0 < \theta \leq 1, 0 < \delta \leq 1]$ has been systematically explored. Next the computer determined which pair of values of θ and δ yielded the smallest χ^2, and these are used as the estimates. When enough points in the parameter space are scanned, the method yields a close approximation to the analytic solution.*

The results of the minimization procedure are presented in Fig. 6, which displays the fits, and gives the parameter estimates and χ^2 values. As noted earlier, the prediction for list lengths less than 6 is perfect recall at all positions. A measure of the overall fit of this model can

---

*For a discussion of this procedure see Atkinson, Bower, and Crothers 1965.

be achieved by summing the X^2's for each list length. The result is a X^2 of 31.8 which is to be evaluated with 38 degrees of freedom. (There are 46 points to be fit and two parameters are estimated for each list length.)

As we can see from an inspection of Fig. 6, the model provides a good account of the data. Also, note that the estimates of δ are reasonably constant as list length varies. Indeed on theoretical grounds there is no reason to believe that δ should vary with list length. Note also that a δ of about .40 gives a slight S-shape to the recency portion of the curve; as indicated in Fig. 3, the higher δ the greater the S-shape effect. As indicated earlier, the S-shape effect depends directly upon the tendency for the oldest items in the buffer to be lost first. One might conjecture that this tendency would depend on factors such as the serial nature of the task, the makeup of the stimulus material, the instructions, and the subject's knowledge of when the display list will end. In the present experiment, the subject knew when the list would end, and was faced with a memory task of a highly serial nature. For these reasons we would expect an S-shaped recency effect. It should be possible to change the S-shape to an exponential by appropriate manipulation of these experimental factors (Atkinson, Hansen, and Bernbach, 1964).

A notable aspect of the fit is the rapid drop in the θ parameter as list length increases. Furthermore, it is intuitively clear that as list length increases, the probability of recall will necessarily tend to a guessing level for all but the most recent items. Thus, to account for the effect with this model, it would be necessary to assume that the θ parameter goes to zero as list lengths increase. However, because Model I is minimized over two parameters, the drop in $\hat{\theta}$ is undoubtedly confounded

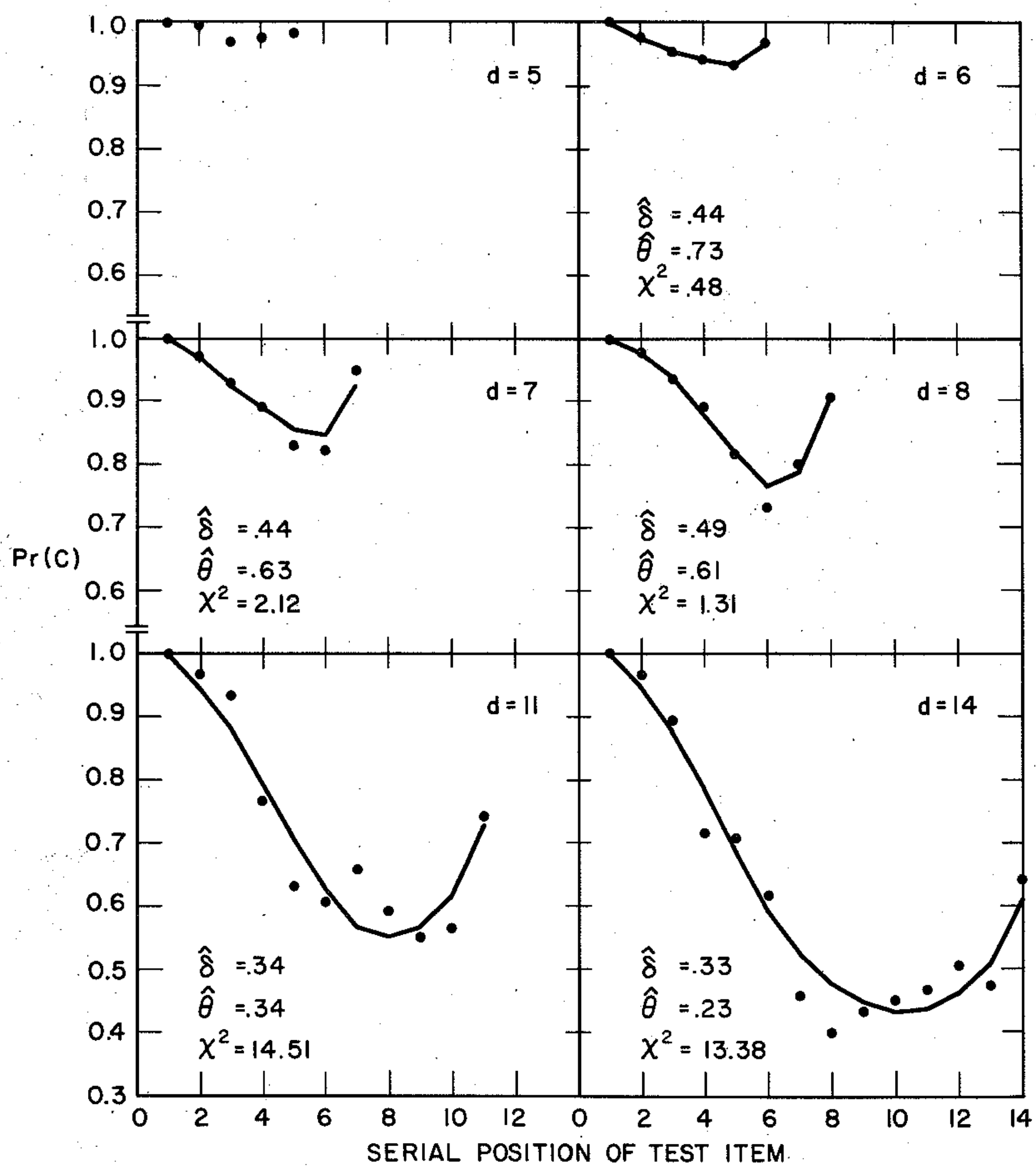

Fig. 6. Goodness-of-fit results for Model I ($\Sigma\chi^2(d)$ = 31.8 on 36 degrees of freedom).

with the variations in $\hat{\delta}$. For this reason the χ^2 minimization was carried out using a single value of δ for all list lengths simultaneously, and selecting an estimate of θ for each list length separately. The fit was about the same as the one displayed in Fig. 6 so it will not be graphed. The minimum χ^2 summed over all list lengths was 39.1 based on 40 degrees of freedom. The estimate of δ was .38 and the various estimates of θ were as follows:

List Length	$\hat{\theta}$
6	.72
7	.61
8	.59
11	.35
14	.24

MODEL II (IMPERFECT RETRIEVAL OF ITEMS IN LTS)

From the above results it is clear that θ is dropping with list length. While attempts to explain this drop could be made in terms of changing motivation or effort as the lists get longer, we dislike such explanations for several reasons. First of all, experiments in which the subject does not know when the display list will end show the same effects (this will be seen in a free recall experiment to be presented later). Also, subjects report that they try as hard, if not harder, on the longer lengths. Finally, the magnitude and orderliness of the effect belie efforts to explain it in such an offhand fashion.

The approach we shall take is that retrieval from the LTS is not perfect. In particular, if the subject does not find the item in the buffer, we assume he engages in a search process of LTS. The probability that this search is successful decreases as the number of items in LTS increases. The next model, Model II, is therefore identical with Model I except that a retrieval function (that described in Postulate D-3-a) is appended to determine the probability that an item is recovered from LTS. With the addition of a retrieval function it is now possible to estimate a single δ and a single θ for all list lengths.

The assumptions are as follows: if at the time of test the sought-after item is not found in the buffer, then a search of LTS is made. The search consists of making exactly R picks with replacement from among the items in LTS, and then stopping. If the item is found, it is reported out with probability 1; if not, the subject guesses.

Mathematical Development of Model II

For Model II it is necessary to determine $s_i^{(d)}$ and $\ell_i^{(d)}$ as well as $f_i^{(d)}$. To do this, define

β_{ij} = probability that an item currently in slot i of a full buffer is still in the buffer j items later.

The difference equations defining β_{ij} are straightforward, being functions solely of the κ_j:

$$\begin{aligned}
\beta_{1,j} &= (1-\kappa_1)^j \\
\beta_{2,j} &= \kappa_1\beta_{1,j-1} + (\kappa_3 + \kappa_4 + \cdots + \kappa_r)\beta_{2,j-1} \\
&\vdots \\
\beta_{i,j} &= (\kappa_1 + \kappa_2 + \cdots + \kappa_{i-1})\beta_{i-1,j-1} + (\kappa_{i+1} + \kappa_{i+2} + \cdots + \kappa_r)\beta_{i,j-1} \\
&\vdots \\
\beta_{r-1,j} &= (\kappa_1 + \kappa_2 + \cdots + \kappa_{r-2})\beta_{r-2,j-1} + \kappa_r\beta_{r-1,j-1} \\
\beta_{r,j} &= (\kappa_1 + \kappa_2 + \cdots + \kappa_{r-1})\beta_{r-1,j-1} \quad .
\end{aligned} \tag{6}$$

The initial conditions are $\beta_{i,0} = 1$. Incidentally, Fig. 3 is a graph of $\beta_{5,j}$ for the δ scheme defined earlier.

The $s_i^{(d)}$ can now be defined in terms of the β_{ij}; namely

$$s_i^{(d)} = \begin{cases} \beta_{d-i+1,d-r} & , \text{ if } i \geq d-r+1 \\ \beta_{r,i-1} & , \text{ if } i < d-r+1 . \end{cases} \qquad (7)$$

We have already obtained an expression for $f_i^{(d)}$, therefore $\ell_i^{(d)}$ can be recovered as follows:

$$\ell_i^{(d)} = 1 - f_i^{(d)} - s_i^{(d)} .$$

Now define

$h_i^{(d)}$ = probability of finding the i^{th} item in a single search of LTS, given that the i^{th} item is in LTS, and not in the buffer.

$\rho_i^{(d)}$ = probability of retrieving the i^{th} item as the result of a search process in LTS, given that the i^{th} item is in LTS, and not in the buffer.

But the number of items in LTS and not in the buffer is the sum of the $\ell_i^{(d)}$. Further, since we select randomly from this set it follows that

$$h_i^{(d)} = \left[1 + \sum_{j \neq i} \ell_j^{(d)} \right]^{-1} , \qquad (8)$$

where j ranges from 1 to d.* (An alternative conception is that the search takes place among all the items in LTS, whether or not they are in the buffer. If this were the case then we would have a smaller $h_i^{(d)}$. We have decided to present the above scheme, however, since the two schemes give little different results in practice. This occurs because the smaller

---

*Equation 8 is actually an approximation, but it greatly simplifies calculations and the error introduced is negligible.

$h_i^{(d)}$ of the second scheme can be compensated for by a higher estimate of R.)

We now define $\rho_i^{(d)}$ in terms of $h_i^{(d)}$; namely

$$\rho_i^{(d)} = 1 - \left[1 - h_i^{(d)}\right]^R , \qquad (9)$$

since, to miss an item entirely, it must be missed in R consecutive picks. Hence

$$\Pr[C_i^{(d)}] = s_i^{(d)} + \ell_i^{(d)}\rho_i^{(d)} + \frac{1}{4}\left\{f_i^{(d)} + \ell_i^{(d)}[1 - \rho_i^{(d)}]\right\} . \qquad (10)$$

We next define

$$\chi^2 = \chi^2(6) + \chi^2(7) + \chi^2(8) + \chi^2(11) + \chi^2(14) \qquad (11)$$

where $\chi^2(d)$ was given in Eq. 5. To apply Model II to our data, we minimized the above χ^2 function over the parameters θ, δ, and R. As before, r was set equal to 5. The parameter estimates were as follows:

$$\hat{\delta} = .39$$

$$\hat{\theta} = .72$$

$$\hat{R} = 3.15 .$$

The predicted curves are given in Fig. 7. The fit of Model II is remarkably good; simultaneously fitting five list lengths, the minimum χ^2 is only 46.2 based on 43 degrees of freedom (i.e., there are 46 points to be fit, but three parameters were estimated in minimizing χ^2). The fit is very nearly as good as that of Model I where each list length was fit separately using 10 parameter estimates. As pointed out earlier, however, there are many possible retrieval schemes which could be suggested. Is it possible on the basis of a χ^2 criterion to distinguish among these? By way of answering this question, we shall consider a second, very different retrieval procedure, to be called Model III.

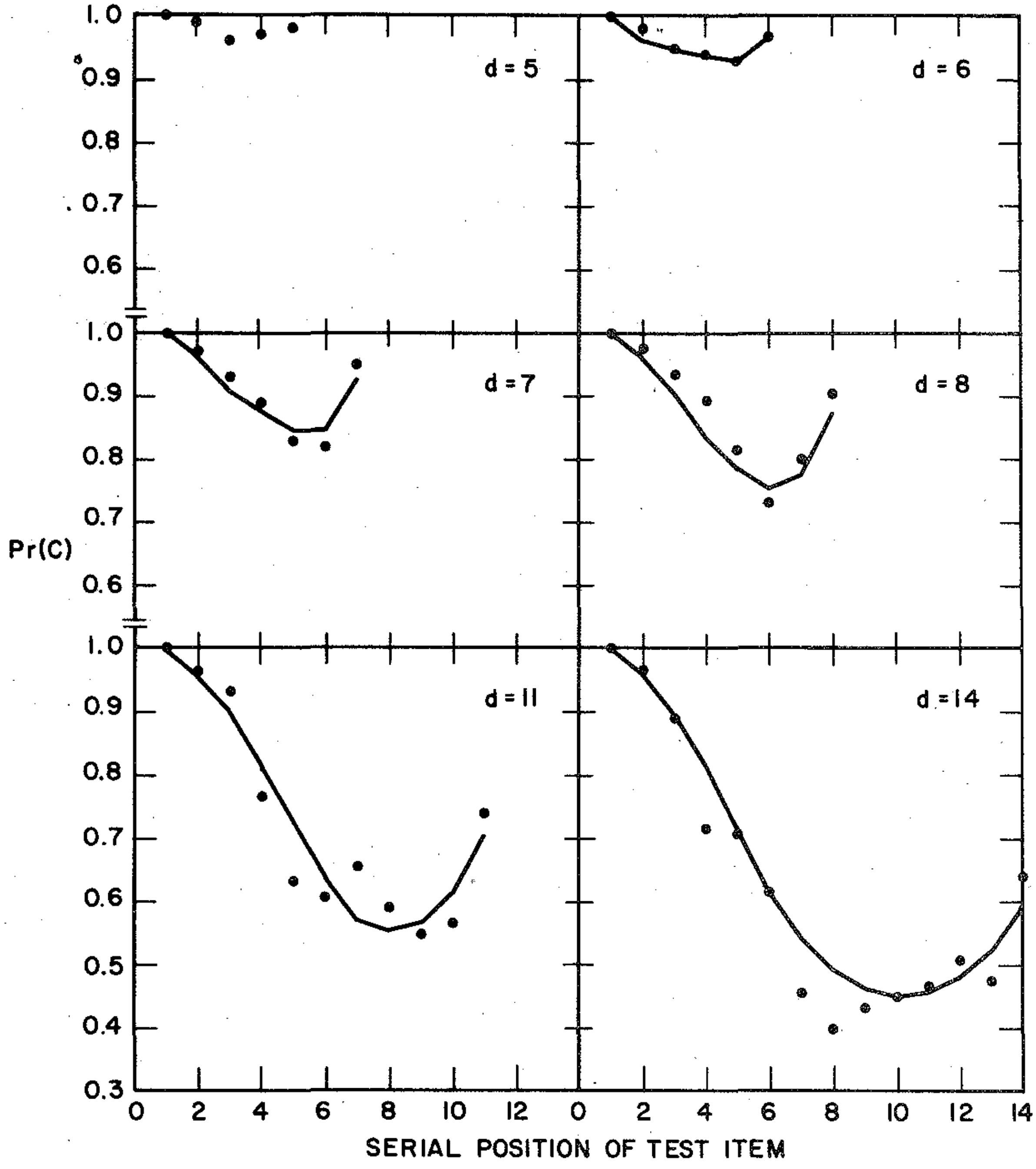

Fig. 7. Goodness-of-fit results for Model II_2 (parameter values: $\delta = .39$, $\theta = .72$, $r = 5$, $R = 3.15$, $X^2 = 46.2$ on 43 degrees of freedom).

MODEL III (IMPERFECT RETRIEVAL OF ITEMS IN LTS)

This model is identical to Model II except for the retrieval process. The proposal is that mentioned in Postulate D-3-b. Searches in the LTS are made randomly with replacement. Each unsuccessful search disrupts the looked-for item with probability R'. If the item is ever disrupted during the search process, then when the item is finally retrieved the stored information will be such that the subject will not be able to recall at better than the chance level. Figure 8 shows the branching tree for this process, where $h_i^{(d)}$ is the probability of finding the item on each search. For this process

$$\rho_i^{(d)} = h_i^{(d)} + [1 - h_i^{(d)}]\ (1 - R')h_i^{(d)} + \left\{[1 - h_i^{(d)}](1 - R')\right\}^2 h_i^{(d)} + \ldots$$

$$= h_i^{(d)} \left\{ \sum_{j=0}^{\infty} [1 - h_i^{(d)}]^j (1 - R')^j \right\} \tag{12}$$

$$= \frac{h_i^{(d)}}{1 - [1 - h_i^{(d)}](1 - R')} .$$

The same method for estimating parameters used for Model II was also used here. The obtained minimum χ^2 was 55.0 (43 degrees of freedom), and the parameter estimates were as follows:

$$\hat{\delta} = .38$$

$$\hat{\theta} = .80$$

$$\hat{R}' = .25.$$

The predicted curves are shown in Fig. 9. The fit is not quite as good as for Model II, but the difference is not great enough to meaningfully distinguish between the two models. Notwithstanding this fact, we shall go on and develop a somehat more sophisticated retrieval model for use later in the paper.

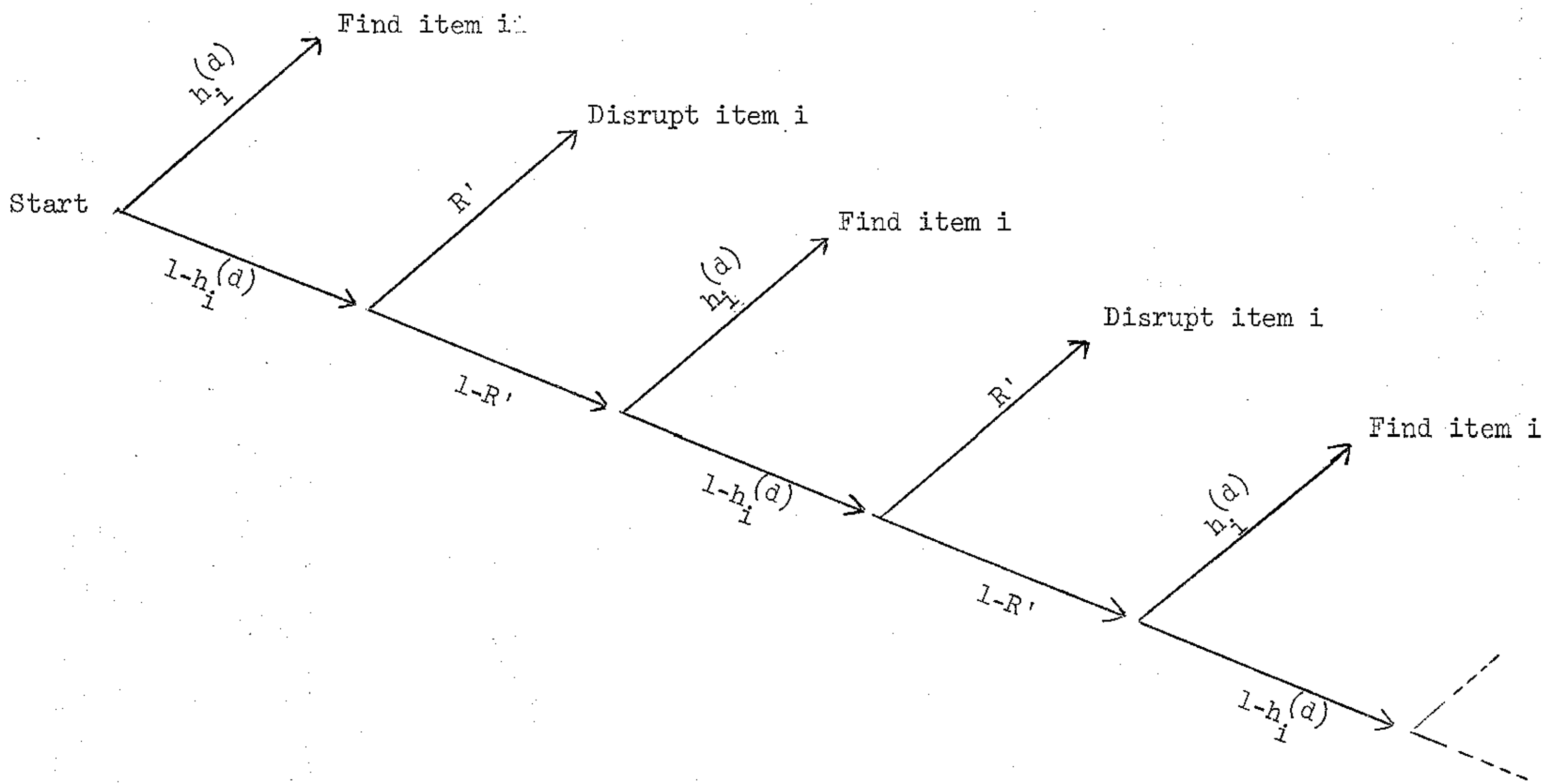

Figure 8. Retrieval process for Model III

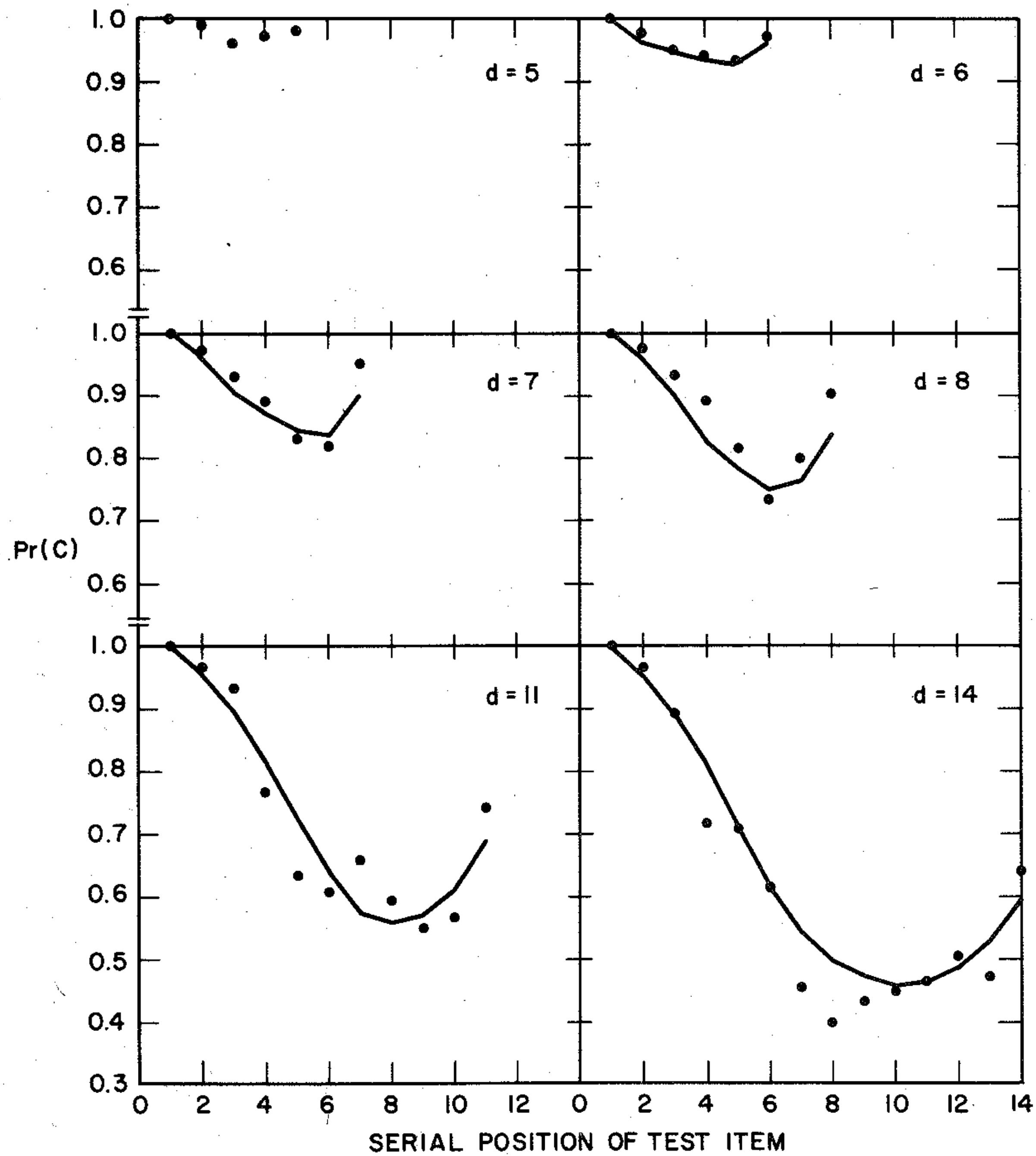

Fig. 9. Goodness-of-fit results for Model III (parameter values: $\delta = .38$, $\theta = .80$, $r = 5$, $R' = .25$, $\chi^2 = 55.0$ on 43 degrees of freedom).

STRENGTH MODELS FOR LTS

Models I, II, and III are all marked by the same assumption concerning what is stored in LTS. In all these models, an item can be stored only once in an all-or-none fashion. We now will develop some of the techniques necessary to deal with more complicated models. There are several reasons that motivate the development: first, the single-copy model gives no reasonable method to deal with confidence ratings; second, there is no particularly good way of dealing with the confusion errors found in certain types of experiments (see Conrad, 1964); and third, the single-copy model does not lend itself well to postulates concerning what happens when items are repeatedly presented as in a paired-associate learning task.

Consider for a moment the problem of confidence ratings. In the Phillips and Atkinson experiment described earlier, subjects were asked to give the confidence rating 1, 2, 3, or 4 depending on their estimate of the number of alternatives from which they were choosing. If they could actually follow these directions, their probabilities of being correct for each confidence rating would be 1.0, 0.50, 0.33, and 0.25, respectively. The results are shown in Fig. 10. What is graphed is the probability of a correct response, given that confidence rating i was made against the inverse of the confidence rating. Since the inverse of the confidence rating is the value the subjects should approximate if they were able to obey the instructions accurately, the points should all fall on a straight line with slope 1.

The fact that the observed response probabilities are quite close to the values predicted on the basis of confidence ratings, indicates that a useful alternative to the "signal detectability theory" view of confidence

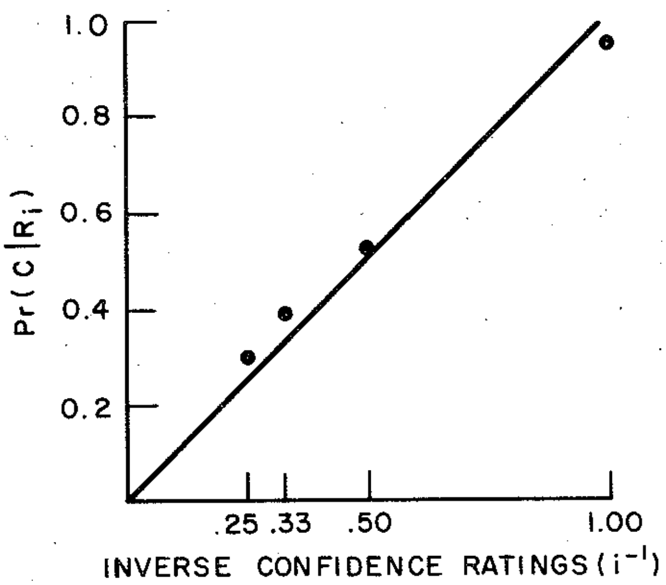

Fig. 10. Probability of a correct recall vs. reciprocal of the confidence ratings.

ratings can be found (De Finetti, 1965; Egan, 1958). In any case it is not unreasonable to assume that the subject does actually choose from among either 1, 2, 3, or 4 alternatives at different times, and that one of the picked-from alternatives is the correct response. We will not try in this paper to present a model capable of explaining these results. Nevertheless it is clear that a model of greater sophistication than the all-or-none, single-copy model is needed. For these and related reasons we would like to analyze some of the implications of buffer models postulating a memory strength in LTS.

Two aspects of the earlier models, the transfer assumptions and the long-term storage ssumptions, will now be re-examined. The basic premise to be considered is that whatever is stored in LTS (the number of copies, a strength measure, etc.) is a function of the time spent by an item in the buffer. At this stage, therefore, some statistics relevant to an item's duration in the buffer are developed.

Define

ζ_{ij} = probability that an item currently in slot i of a full buffer is knocked out of the buffer when the j^{th} succeeding item is presented.

Then

$$
\begin{aligned}
\zeta_{1,j} &= (1 - \kappa_1)^{j-1}\kappa_1 \\
\zeta_{2,j} &= \kappa_1\zeta_{1,j-1} + (\kappa_3 + \kappa_4 + \dots + \kappa_r)\zeta_{2,j-1} \\
&\vdots \\
\zeta_{i,j} &= (\kappa_1 + \kappa_2 + \dots + \kappa_{i-1})\zeta_{i-1,j-1} + (\kappa_{i+1} + \kappa_{i+2} + \dots + \kappa_r)\zeta_{i,j-1} \\
&\vdots \\
\zeta_{r-1,j} &= (\kappa_1 + \kappa_2 + \dots + \kappa_{r-2})\zeta_{r-2,j-1} + \kappa_r\zeta_{r-1,j-1} \\
\zeta_{r,j} &= (\kappa_1 + \kappa_2 + \dots + \kappa_{r-1})\zeta_{r-1,j-1} .
\end{aligned}
\tag{13}
$$

The initial conditions are $\zeta_{i,1} = \kappa_i$. An important function may now be defined in terms of the ζ_{ij}'s. Namely,

$\omega_{ij}^{(d)}$ = probability that the i^{th} item in a list of length d stays in the buffer exactly j units of time (where a time unit is the presentation period per item).

Then

$$\omega_{ij}^{(d)} = \begin{cases} 0 & , \text{ if } i < j \\ 1 - \sum_{j=1}^{j=i-1} \omega_{ij}^{(d)} = s_i^{(d)} & , \text{ if } i = j \\ \zeta_{rj} & , \text{ if } i > j \text{ and } i \leq d - r + 1 \\ \zeta_{d-i+1, j-i+d-r+1} & , \text{ if } i > j,\ i > d - r + 1 \text{ and } j > i - d + r - 1 \\ 0 & , \text{ if } i > j,\ i > d - r + 1 \text{ and } j \leq i - d + r - 1 . \end{cases} \tag{14}$$

The convention is used here that if item i is still in the buffer at the time of test, the number of time units it is said to have been present in the buffer is i.

Our assumptions for the present model go back to the suggestions made in Postulates C and D. Consideration of each item as made up of a large number of bits of information (used here in a loose sense--not necessarily binary bits) lends credence to the postulate that an item's strength in LTS can build up in a gradual continuous fashion as a function of time spent in the buffer. In particular, the assumption is made here that what is stored in LTS is represented by a strength measure.* For example, the

---

*This assumption is actually quite similar to the multiple-copy assumptions, and it would be extremely difficult to differentiate the two on the basis of data. More will be said about this later.

strength could represent the number of bits of information stored. This strength measure will be defined for a list of length d as follows:

$\lambda_{ij}^{(d)}$ = strength of the i^{th} item in LTS, given that it was in the buffer exactly j units of time.

In order to define a transfer function to LTS, we use the notation introduced earlier. However, the θ_{ij}'s are no longer a probability that an item will be transferred. Instead they represent a weighting factor on the time spent in the buffer. For example, an item is weighted more for each time unit it spends in the buffer alone, than when it shares the buffer with several other items. One way of looking at this is to think of the amount of "attention" received by an item in one unit of time; if all items in the buffer are attended to for an equal share of the available time, then an item alone in the buffer for one second would be attended to for the full second, whereas an item sharing the buffer with four others would be attended for only 1/5 second. In this case, then, the item alone would be weighted five times as heavily as the item which shares the buffer with four others.

As before we will make the simplifying assumption that the θ_{ij}'s do not depend on i, the buffer position; hence the first subscript is superfluous and will be dropped leaving θ_j as the weighting function. Thus θ_j represents how much each item is to be weighted, if there are currently j items in the buffer. We can now compute the strength that an item accumulates during its stay in the buffer. To do this simply consider the number of time units an item is in the buffer; multiply each unit by the appropriate θ_j and also by the length of the time unit. To state this mathematically, let $\mu_{ij}^{(d)}$ denote the <u>weighted time</u> that item i accumulates in the buffer, if it remains in the buffer j time units. Then

$$\mu_{ij}^{(d)} = \begin{cases} \theta_r jt & , \text{ for } i \leq d-r+1 \\ \left[\theta_r(j-i+d-r+1) + \sum_{i=d-i+1}^{r-1} \theta_i\right] t & , \text{ for } i > d-r+1 , \end{cases} \tag{15}$$

where t denotes the length of a time unit (i.e., the presentation time per item).

The central assumption, now, is that the strength built up in LTS is a linear function of the weighted time accumulated. Namely

$$\lambda_{ij}^{(d)} = \gamma\mu_{ij}^{(d)}$$

where γ is a dummy parameter. The introduction of γ permits us to convert θ_j to a rate measure; specifically the variable of interest is the rate at which strength accumulates, defined here as $\gamma\theta_j$. Obviously θ_j could have been defined directly as a rate parameter; however, we preferred to have θ_j bounded between 0 and 1 in order to keep its usage in line with earlier developments. What this means, of course, is that in any application of the strength model the quantity θ_1 can be arbitrarily set equal to 1. To make this point entirely clear, note that $\lambda_{ij}^{(d)}$ can be rewritten as follows:

$$\lambda_{ij}^{(d)} = \begin{cases} (\gamma\theta_r) jt & , \text{ for } i \leq d-r+1 \\ \left[(\gamma\theta_r)(j-i+d-r+1) + \sum_{i=d-r+1}^{r-1} (\gamma\theta_i)\right] t & , \text{ for } i > d-r+1 . \end{cases}$$

The strength schema outlined above is somewhat analogous to what has been labeled in the literature a "consolidation process." One view of the consolidation hypothesis holds that a short-term decaying trace lays down a permanent structural change in the nervous system; in turn, our model postulates that a strength measure is laid down in permanent memory during the period that an item remains in the buffer. Whether or not there is anything significant to this similarity, the analogy will not be pursued further in this paper.

An important property of this model is now presented: regardless of any conditionalities, the total strength in LTS of all items in a display of size d is a constant. This total strength will be denoted as $S(d)$, and is as follows:

$$S(d) = \left[r(d-r)\theta_r + \sum_{i=1}^{r} (i\theta_i) \right] t\gamma \ . \tag{16}$$

Thus for the retrieval schemes discussed earlier, the probability of finding item i in a single search, given that the item had been in the buffer for j time units is as follows:

$$h_{ij}^{(d)} = \frac{\lambda_{ij}^{(d)}}{S(d)} ,$$

which simply says that the probability of picking the i^{th} item is its <u>relative</u> strength.

In terms of our earlier analyses, it seems reasonable to assume that whatever the retrieval procedure, the probability of recall will be a function of $h_{ij}^{(d)}$. Thus, if

$\rho_{ij}^{(d)}$ = probability of retrieving item i from LTS, given that it was in the buffer exactly j time units,

then $\rho_{ij}^{(d)}$ will be some as-yet-unspecified function of $h_{ij}^{(d)}$. Taking the next step yields an expression for $\Pr[C_i^{(d)}]$; namely

$$\Pr[C_i^{(d)}] = s_i^{(d)} + [1 - s_i^{(d)}] \left\{ \frac{1}{4} + \frac{3}{4} \sum_{j=1}^{i-1} \omega_{ij}^{(d)} \rho_{ij}^{(d)} \right\} , \qquad (17)$$

where non-retrievals are interpreted as generating correct responses at guessing probability of 1/4.

The stage has now been reached where it is necessary to specify a retrieval process in order to complete the model and apply it to data.* Many processes come to mind, and we have tried several on the Phillips and Atkinson data. However, as one might expect, the data from that experiment do not permit us to distinguish among them. Consequently it will be necessary to analyze other experiments; in particular certain especially contrived studies involving free verbal recall. Before turning to the free verbal recall experiments, however, it will be helpful to examine a paired-associate learning experiment for indications of how to proceed. We do this because a central question not yet considered is how to handle repeated presentations of the same item.

PAIRED-ASSOCIATE LEARNING

Our analysis of learning will be primarily within the framework of a paired-associate model proposed by Atkinson and Crothers (1964) and Calfee and Atkinson (1965). This model postulates a distinction between short-

---

*We still have not considered the problem of confidence ratings, but we have reached a point where suggestions can be made for dealing with them. For example, cut-off points can be defined along the strength dimension, and the retrieval process modified to handle this elaboration.

term and long-term memory and has been labeled the trial-dependent-forgetting (TDF) model because the recall process changes over time. With certain minor amendments the TDF model can be viewed as a special case of the buffer model presented in this paper. Our approach in this section will be to analyze some paired-associates data in terms of the TDF model, with the goal of determining what modifications need to be made in the buffer model to make it a viable theory of learning. To start, let us consider the experimental task.

A Paired-Associate Experiment Manipulating List Length

Three groups of 25 college students were used as subjects. Each subject learned a paired-associate list in which the stimulus members consisted of two-digit numbers, and the response members were one of three nonsense syllables. For group 21 a set of 21 stimulus items was selected on the basis of low inter-item association value. For groups 9 and 15 the experimental lists consisted of a selection of 9 or 15 items, respectively, from this set, a different subset being selected randomly for each subject. Each of the three responses was assigned as the correct alternative equally often for each subject. After instructions and a short practice list, the experiment began. As each stimulus item was presented the subject was required to choose one of the three responses, following which he was informed of the correct response. In order to reduce primacy effects, the first three stimulus-response pairs shown to the subject were two digit numbers that were not in the set of 21 experimental items; these three items did not reoccur on later trials. Then, without interruption, the experimental list (arranged in a random order) was presented. After the entire list had been presented, the second trial then proceeded without interruption in the same manner with the items arranged in a new random order. Thus, the procedure involved continuous presentation of items with no breaks between trials.*

Figure 11 presents the mean learning curves for the three experimental groups. The curves are ordered on the list length variable, with the longer lists producing a slower rate of learning. It should be clear that this effect is a direct consequence of the buffer model, since for the longer lists a smaller proportion of the items is retrieved via the buffer. Figure 12 presents the conditional error curves, $Pr(e_{n+1}|e_n)$, which also are ordered according to list length. Note that the conditional probability is definitely decreasing over trials. Without going into details now, it is clear that a buffer model will also predict this effect because the probability of retrieval would increase with repeated presentations.

Trial-Dependent-Forgetting Model

As noted earlier the TDF model assumes that paired-associate learning is a two-stage process in which a given stimulus item may be viewed as initially moving from an unconditioned state to an intermediate short-term state. In the intermediate state an item may either move back to the unconditioned state or move to an absorbing state. This intermediate state can be viewed as a counterpart of the buffer in our buffer model, and the absorbing state the counterpart of LTS.

To develop the TDF model mathematically, the following notions need to be introduced. Each item in a list of paired-associates is assumed to be in one of three states: (a) state U is an unlearned state in which

*See Calfee and Atkinson (1965) for a detailed account of this experiment.

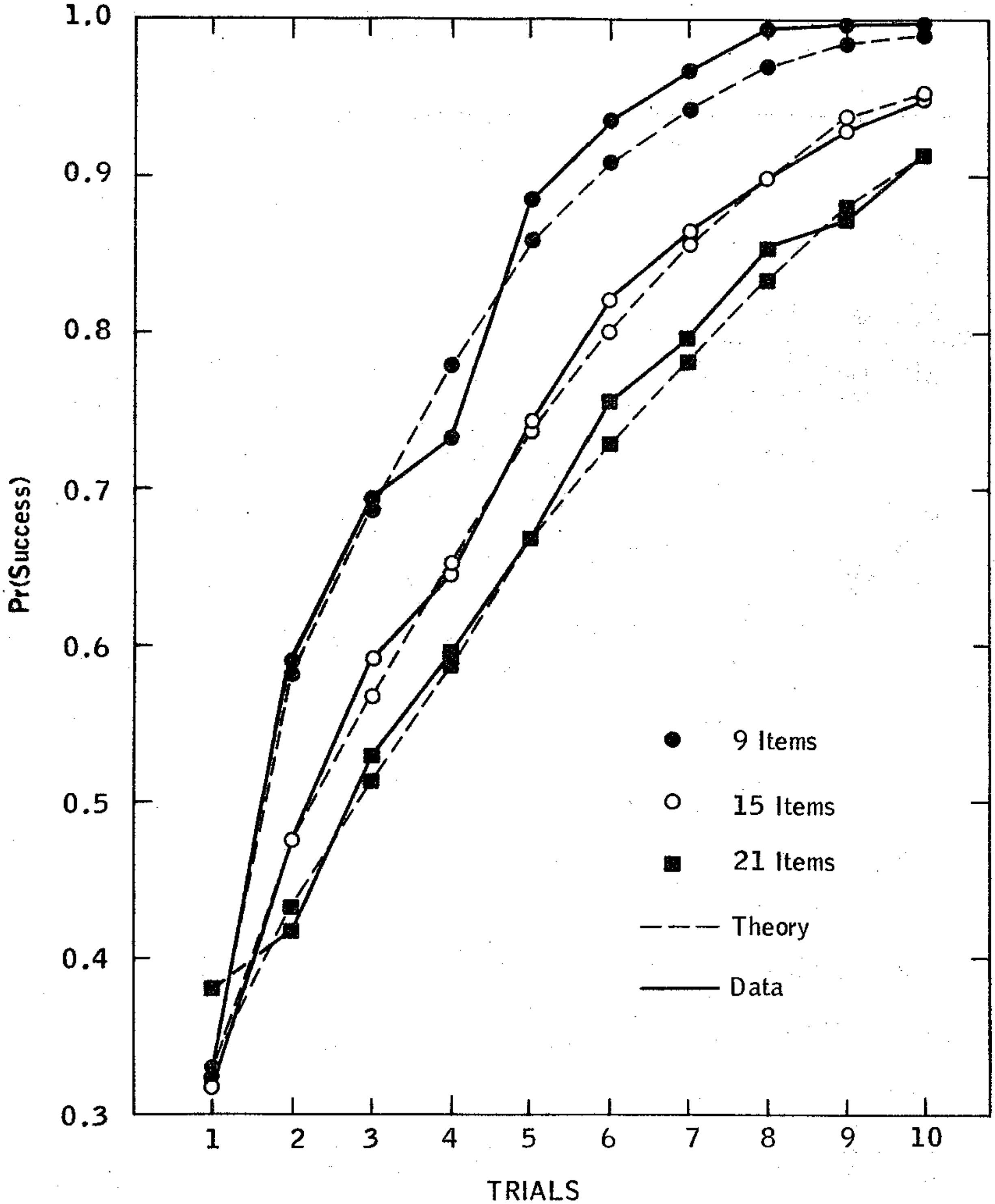

Fig. 11. Average probability of a success on trial n for three groups with different list lengths. See text for description of theoretical curves.

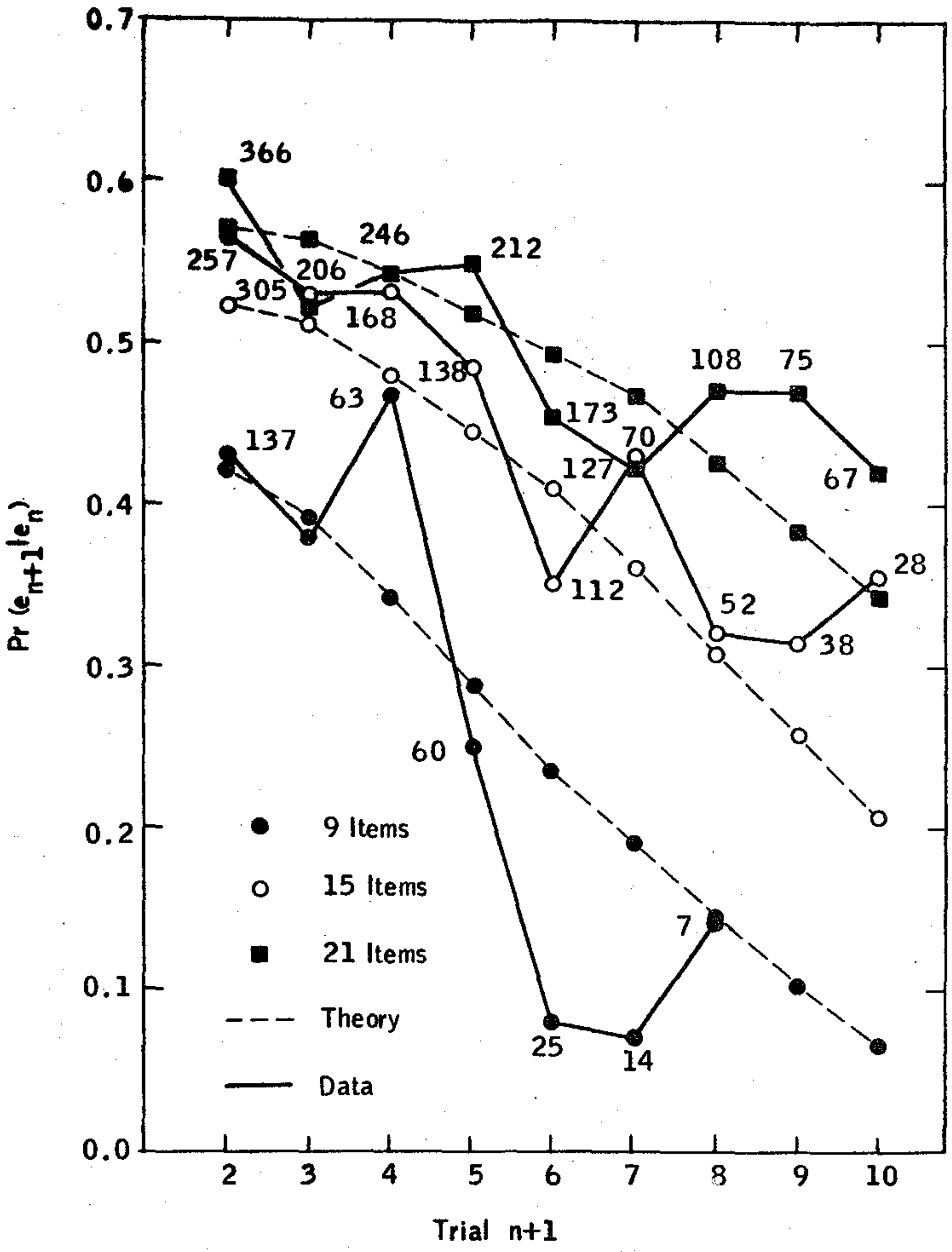

Fig. 12. Average probability of an error on trial n+1, given an error on trial n for three groups with different list lengths.

the subject guesses at random from the set of response alternatives, (b) state S is a short-term-memory state, and (c) state L is a long-term state. The subject will always give a correct response to an item if it is in either state S or state L. However, it is possible for an item in state S to return to the unconditioned state (i.e., be forgotten); whereas, once an item moves to state L it is learned, in the sense that it will remain in state L for the remainder of the experiment.* The probability of a return from state S to state U is postulated to be a function of the number of other items that remain to be learned on any given trial. In terms of the buffer model, this is similar to the statement that the probability of being knocked out of the buffer is related to the number of items still to be presented.

Two types of events are assumed to produce transitions from one state to another in the TDF model: (a) the occurrence of a reinforcement, i.e., the paired presentation of the stimulus item together with the correct response alternative and (b) the presentation of an unlearned stimulus-response pair (an item not in state L) between successive occurrences of a particular item. The associative effect of a reinforcement is described

---

*In order to make the TDF model parallel the buffer model, the reader should assume that U refers to the state in the buffer model where an item is neither in the buffer nor in LTS; that S refers to the state where an item is solely in the buffer and not in LTS; and that L refers to any item which has entered LTS, whether in the buffer or not. Furthermore the recall assumptions imply that a very elementary retrieval scheme is being put forth: any item in LTS is recalled with probability 1.

by matrix A below:

$$A = \begin{array}{c} \\ L \\ S \\ U \end{array} \begin{array}{c} \begin{array}{ccc} L & S & U \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ a & 1-a & 0 \\ b & 1-b & 0 \end{bmatrix} \end{array} \tag{18}$$

Thus if an item is in state U and the correct response is shown to the subject, then the item moves to state L with probability b, or to state S with probability $1-b$. Starting in S it moves to L with probability a or remains in S with probability $1-a$. In either case, if the item were to be presented again immediately following a reinforcement, this model, like the buffer model, makes the plausible prediction that a correct response would be certain to occur.

The effect of the presentation of a single unlearned stimulus-response pair on the state of a particular item is described by matrix F:

$$F = \begin{array}{c} \\ L \\ S \\ U \end{array} \begin{array}{c} \begin{array}{ccc} L & S & U \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-f & f \\ 0 & 0 & 1 \end{bmatrix} \end{array} \tag{19}$$

If a given item is in state S and some other unlearned stimulus-response pair is presented, then the interference produced by the unlearned pair results in forgetting of the item (i.e., transition to state U) with probability f, and otherwise there is no change in state. Furthermore, it is assumed that when a learned stimulus-response pair is presented there is no change in state.* Again drawing a parallel to the buffer model, we should

---

*See Brown and Battig (1966) for experimental work in support of this notion.

note that the above transition matrices require that an item move to LTS only when it is presented. However, the parameters a and b can be interpreted as a rough approximation of the average probability of transfer during an item's stay in the buffer. Parameter a, of course, refers to a process that has not heretofore been considered in the buffer model: a repeated presentation of an item. Similarly, the assumption that the presentation of a learned item will not effect a change in state has not been previously considered. It is clear, however, that assumptions of this nature will have to be proposed in extensions of the buffer model. More will be said about this shortly.

Continuing, however, let T_n be the matrix of the transition probabilities between states for a particular item from its n^{th} to its $(n+1)^{st}$ presentations, and suppose ξ_n is the number of other unlearned items that intervene between these two presentations of the given item. Then T_n is found by postmultiplying A by the ξ_n^{th} power of F; matrix A represents the n^{th} reinforced presentation of the item, and the interference matrix F is applied once for each of the intervening unlearned pairs. Performing the multiplication yields:

$$T_n = \begin{array}{c} \\ L_n \\ S_n \\ U_n \end{array} \begin{array}{c} \begin{array}{ccc} L_{n+1} & S_{n+1} & U_{n+1} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ a & (1-a)(1-F_n) & (1-a)F_n \\ b & (1-b)(1-F_n) & (1-b)F_n \end{bmatrix} \end{array} \qquad (20)$$

where $F_n = 1 - (1-f)^{\xi_n}$.

Unfortunately there is no way of determining from the data the exact value of ξ_n. However, an approximation can be used. Let X denote the number of items in the paired-associate list and remember that a trial consists of a random ordering of these items. Between the n^{th} and the $(n+1)^{st}$ presentations of a given item $(j+k)$ interpolated pairs (IP) may intervene; j on trial n and k on trial $n+1$ (where $j,k = 0, 1, \cdots X-1$). The probability of j IP's on trial n is the probability that the item is in position $X-j$, which is $1/X$; whereas the probability of k IP's on trial $n+1$ is the likelihood that the item is in position $k+1$, which also is $1/X$. Thus for each combination of j and k, the probability of the combination occurring is $1/X^2$. For each of these combinations the average value of ξ_n will be $j(1-\ell_n) + k(1-\ell_{n+1})$, where ℓ_n is the probability of being in state L on trial n. Using this average as an approximation,

$$F_n = 1 - \frac{1}{X^2} \sum_{j=0}^{X-1} \sum_{k=0}^{X-1} (1-f)^{[j(1-\ell_n)+k(1-\ell_{n+1})]}$$

$$= 1 - \frac{1}{X^2} \left\{\frac{1-(1-f)^{X(1-\ell_n)}}{1-(1-f)^{(1-\ell_n)}}\right\}\left\{\frac{1-(1-f)^{X(1-\ell_{n+1})}}{1-(1-f)^{(1-\ell_{n+1})}}\right\} . \qquad (21)$$

During the early trials of an experiment, ℓ_n will be small (all items are assumed to be in state U initially, and so ℓ_1 is 0); hence F_n, the probability of forgetting while in state S, will be relatively large. As n increases, ℓ_n approaches 1 and so F_n goes to 0. As a consequence of the decrease in F_n over trials, the model predicts a non-stationary learning process. For example, consider the probability of an error on the $n+1^{st}$ presentation of an item conditional on an error on its n^{th} presentation. The error on trial n indicates that the item is in state U, so the probability of an error on the next trial is the joint probability of (a) no learning, (b) forgetting, and (c) an incorrect response by chance; namely

$$Pr(e_{n+1} | e_n) = (1-b)F_n(1-g) ,$$

where g denotes the probability of a correct response by guessing. In other words, $Pr(e_{n+1} | e_n)$ is predicted to decrease over trials, a finding reported by several investigators.

Goodness-of-Fit Results

We are now in a position to analyze the paired-associate experiment described earlier.

Parameter estimates for the TDF models were obtained by applying the chi-square minimization method described by Atkinson, Bower, and Crothers (1965). The data used in parameter estimation were the sequences of successes and errors from trials 2 through 5 and trials 6 through 9. The 16 possible combinations of correct responses (c) and errors (e) for a four-trial block are listed in Table 1 together with the observed frequencies of each combination for the three experimental groups. Thus, the sequence consisting of four errors (eeee) on trials 2 through 5 was observed in 6 of 225 item protocols in group 9, in 30 out of 375 protocols in group 15, and in 55 out of the 525 protocols in group 21. The sequences for trials 6 to 9 are listed in Table 2. In all of the theoretical analyses g was set equal to 1/3, the reciprocal of the number of response alternatives.

The theoretical expressions for the probability of a four-trial sequence was obtained. Following the notation of Atkinson and Crothers (1964), let $O_{i,j,n}$ be the i^{th} four-tuple in Table 1 for group j

TABLE 1

OBSERVED AND PREDICTED FREQUENCIES FOR RESPONSE SEQUENCES FROM TRIALS 2 THROUGH 5

	9 Items				15 Items				21 Items			
Trial 2345	Obs.	TDF	Linear	One-element	Obs.	TDF	Linear	One-element	Obs.	TDF	Linear	One-element
cccc	83	77.2	59.0	88.4	98	90.7	39.9	103.7	97	107.5	45.4	112.6
ccce	3	4.2	9.5	1.3	10	6.7	17.8	3.8	11	9.0	24.2	6.8
ccec	10	8.0	15.2	3.0	13	11.1	23.9	6.6	14	13.7	31.5	10.3
ccee	4	3.7	2.4	2.7	10	9.2	10.7	7.6	12	14.5	16.8	13.5
cecc	18	17.2	25.7	10.4	25	22.7	33.1	17.3	35	27.3	42.2	23.0
cece	2	4.4	4.1	2.7	4	9.9	14.8	7.6	14	15.1	22.5	13.5
ceec	10	8.5	6.6	6.1	7	16.5	19.8	13.3	17	23.3	29.3	20.7
ceee	3	3.9	1.1	5.3	12	13.6	8.9	15.2	20	24.5	15.6	27.1
eccc	40	39.5	48.3	41.9	58	54.6	48.7	57.3	78	67.6	59.4	67.6
ecce	3	4.9	7.8	2.7	6	10.5	21.8	7.6	15	15.6	31.7	13.5
ecec	12	9.4	12.5	6.1	16	17.4	29.2	13.3	22	24.0	41.2	20.7
ecee	2	4.4	2.0	5.3	12	14.3	13.0	15.2	30	25.3	22.0	27.1
eecc	14	20.2	21.1	20.8	31	35.4	40.5	34.6	47	47.6	55.2	46.0
eece	2	5.1	3.4	5.3	11	15.5	18.1	15.2	16	26.5	29.5	27.1
eeec	13	9.9	5.4	12.2	32	25.7	24.2	26.5	42	40.6	38.3	41.4
eeee	6	4.6	0.9	10.7	30	21.2	10.8	30.3	55	42.8	20.4	54.1
χ^2		11.0	73.5	42.5		21.7	173.2	30.3		17.0	180.5	21.8

TABLE 2

OBSERVED AND PREDICTED FREQUENCIES FOR RESPONSE SEQUENCES FROM TRIALS 6 THROUGH 9

Trial 6789	9 ITEMS Obs.	TDF	Linear	One-element	15 Items Obs.	TDF	Linear	One-element	21 Items Obs.	TDF	Linear	One-element
cccc	205	197.2	177.7	192.2	271	260.3	156.3	263.9	319	317.1	178.1	309.7
ccce	0	1.1	5.3	0.3	6	3.3	26.1	1.6	8	5.2	39.5	3.5
ccec	0	2.6	7.9	0.7	8	6.6	32.8	2.7	13	9.2	48.4	5.4
ccee	0	0.3	0.2	0.6	2	2.6	5.5	3.1	4	6.1	10.7	7.1
cecc	12	6.4	5.0	2.5	13	14.4	41.6	7.1	27	19.2	59.8	12.0
cece	0	0.5	0.4	0.6	1	3.1	6.9	3.1	6	6.8	13.3	7.1
ceec	1	1.2	0.5	1.5	2	6.2	8.7	5.4	11	12.1	16.3	10.8
ceee	0	0.2	0.0	1.3	5	2.4	1.5	6.2	10	8.0	3.6	14.1
eccc	13	15.4	18.3	10.1	24	33.7	53.5	23.5	55	45.8	74.8	35.3
ecce	0	0.6	0.5	0.6	2	3.6	8.9	3.1	10	7.5	16.6	7.1
ecec	0	1.5	0.8	1.5	11	7.2	11.2	5.4	5	13.2	20.3	10.8
ecee	0	0.2	0.0	1.3	1	2.8	1.9	6.2	3	8.8	4.5	14.1
eecc	1	3.7	1.2	5.0	15	15.8	14.2	14.2	17	27.4	25.1	24.0
eece	0	0.3	0.0	1.3	5	3.4	2.4	6.2	7	9.8	5.6	14.1
eeec	0	0.7	0.1	2.9	5	6.8	3.0	10.9	11	17.3	6.8	21.6
eeee	0	0.1	0.0	2.6	4	2.7	0.5	12.4	19	11.5	1.5	28.3
χ^2		15.8	25.5	21.3		18.9	210.0	52.0		31.2	428.9	76.0

($j = 9, 15, 21$) where the sequence begins at trial n. Let $\hat{N}(O_{i,j,n})$ be the observed frequency of this four-tuple, and let $Pr(O_{i,j,n};p)$ be the predicted probability for a particular choice of the parameters p of the model. The expected frequency may be obtained by taking the product of $Pr(O_{i,j,n};p)$ with T, the total number of item protocols in group j. We then define the function

$$\chi^2_{i,j,n} = \frac{[N(O_{i,j,n};p) - \hat{N}(O_{i,j,n})]^2}{N(O_{i,j,n};p)} . \qquad (22)$$

A measure of the discrepancy between a model and the data from group j is found by summing Eq. 22 over the sixteen possible sequences for both of the four-trial blocks; i.e.,

$$\chi^2_j = \sum_{i=1}^{16} \chi^2_{i,j,2} + \sum_{i=1}^{16} \chi^2_{i,j,6} . \qquad (23)$$

Equation 23 was also used to obtain estimates of c and θ for the one-element and linear models, respectively, for each of the three experimental groups (these models are described in the book by Atkinson, Bower, and Crothers).

The TDF formulation takes list length into account in the structure of the model, and so presumably the parameters a, b, and f should remain invariant over the three experimental groups. Thus, the estimation procedure was carried out simultaneously over all three groups, so that parameters a, b, and f were found that minimized the function

$$\chi^2 = \chi^2_9 + \chi^2_{15} + \chi^2_{21} \qquad (24)$$

where the χ^2_j are defined in Eq. 23. The minimization was carried out by using a digital computer to search a grid on the parameter space, yielding parameter values accurate to three decimal places.

The χ^2 value obtained by minimizing Eq. 24 does not have a chi-square distribution, since the frequencies in the two 4-trial sets are not independent. However, if one interprets the value obtained from this procedure as a true χ^2, it can be shown that in general the statistical test will be conservative; i.e., it will have a higher probability of rejecting the model than is implied by the confidence level (for a discussion of this problem, see Atkinson, Bower, and Crothers, 1965). In evaluating the minimum χ^2, each set of 16 sequences yields 15 degrees of freedom, since the predicted frequencies are constrained to add to the total number of protocols. Further, it is necessary to subtract one degree of freedom for each parameter estimate. Thus, there are 87 degrees of freedom over the three groups for the TDF model.

Tables 1 and 2 present the predicted frequencies of each response sequence for the TDF model using the minimum χ^2 parameter estimation procedure. Table 3 presents the minimum χ^2 values and the parameter estimates. For comparison purposes, the results for the one-element and linear models also are presented. It can be seen that the TDF model is a marked improvement over both the linear and the one-element models. In fact, the χ^2 of 115.5 (for 87 degrees of freedom) is remarkably low, considering that the parameters are simultaneously estimated for all three experimental groups. The theoretical curves drawn in Figs. 11 and 12 are those derived from the TDF model using the parameter values given in Table 3.

An interesting feature of the fit is that the estimate of the parameter b is about one-fourth as large as the estimate of a. To the extent that these values are accurate, the model predicts that the greatest increase

TABLE 3

Parameter Estimates and χ^2 Values

Model	Parameter	9 Items	15 Items	21 Items	χ^2 Values		
					Trials 2-5	Trials 6-9	Total
TDF	a	0.42	-	-			
	b	0.11	-	-	49.6	65.9	115.5
	f	0.19	-	-			
Linear	θ	0.32	0.17	0.15	427.2	664.4	1091.6
One-element	c	0.30	0.20	0.15	94.6	149.3	243.9

in the probability of recall from one trial to the next will occur if the number of intervening items is as small as possible (since each intervening item helps to return an item to state U where the probability of transition to state L is smallest). A paired-associate experiment reported by Greeno (1964) yielded results contradicting this prediction. Experimental items presented twice in succession on each trial took the same number of trials to reach criterion (i.e., twice the number of stimulus presentations) as control items presented once per trial, indicating that little or no learning took place during the second presentation on each trial, when an item would almost certainly be in state S.

It should be noted that the buffer model would not necessarily make the same prediction here. This is so because, as pointed out earlier, the parameters a and b of the TDF model provide only a rough approximation to the buffer-transfer process which takes place over an extended period of time. The approximation is convenient for the typical paired-associates experiment, but when items are repeated in juxtaposition more specificity is required. On the other hand, until a set of postulates is added concerning the successive presentation of items, one cannot say precisely what the buffer model will predict. Nevertheless, it seems likely that a buffer model would not predict that the maximum advantage would be gained by repeating an item twice in succession. In order to give more meaning to this statement, let us see what possible postulates could be appended to the buffer schema in light of the paired-associate analyses just presented.

Suggested Postulates Concerning Repeated Items

The buffer model has not yet been made applicable to situations where an item is presented more than once. For example, we have not considered the problem of what takes place when an item currently in the buffer is again presented. Several possibilities exist: (a) the incoming item could be shunted aside and the buffer left untouched, (b) the incoming item could occupy position r in the buffer and the old copy of that item could be the item bumped out, or (c) the incoming item could take the r^{th} position in the buffer and the item lost could be chosen by the κ_j function, thereby making it possible for an item to be represented several times in the buffer. Further questions now arise: if an item can be represented more than once in the buffer, does the probability of transfer to LTS proceed independently for each copy; or in the case of the strength model, is the strength built up as a function of the total time spent by both copies in the buffer? Similarly, several possibilities exist for other contingencies that can occur when an item is repeated. For example, if an item is presented which is not in the buffer but is in LTS, does the item get shunted aside and miss the buffer if its long-term copy is retrieved, or does the item get placed in the buffer regardless? Picking among these alternatives requires further experimentation, and is beyond the scope of this paper.

There is, however, one area in which the range of alternatives may be narrowed; namely with regard to retrieval schemes applicable to learning experiments. In our earlier discussion of short-term memory experiments it was necessary to postulate a retrieval process that permitted less than perfect recall for items in LTS. Obviously, for most learning experiments the subject will in time learn to perform perfectly; thus the retrieval process will have to be capable of generating perfect recall as the number of trials increases. One method of defining the retrieval function that would eventually permit perfect retrieval lets the probability of retrieving the i^{th} item depend not only on the relative strength of the item, but also on its absolute strength.* With an assumption of this nature, the probability of recall can go to unity with repeated presentations even though the retrieval process generates imperfect performance on early trials. In our initial discussion of the strength model a retrieval process was not defined, and the reason was that we wanted it to have the property just mentioned. In the next section a retrieval function of this kind will be appended to the strength model and applied to experiments on free verbal recall.

There are other considerations which also lead to a retrieval scheme that can undergo change from trial to trial. Consider, for example, an experiment by Tulving (1962) on free verbal recall. A list of 16 words was read in a random order over and over again until the subject had learned all the words in the list. After each reading of the list the subject would write down all the words he could remember. Each reading of the list was in a new random order; nevertheless the subjects tended to organize their recall in a similar fashion from trial to trial. This clearly contradicts the hypothesis that the subject searches through memory in a random fashion after each reading. The very first recall of the list could

*This notion will be generalized to multiple-copy models in a later section.

be a random search process of the type described earlier in this paper, but later recalls are clearly not a simple reiterating of this random search. For this reason another feature must be added to the retrieval process in experiments where items are repeated: namely, the items may be restructured (or rearranged) in LTS from trial to trial in such a way as to facilitate recall. Another way of saying this is that the retrieval process changes from trial to trial. For example, a subject might start out by searching LTS randomly with replacement. On later trials, however, the subject might restructure his LTS alphabetically, and now make an ordered alphabetic search without replacement. Further speculation on this point is beyond the scope of this paper. For now it should be noted that changes in the retrieval process from trial to trial are likely to be a very important feature of experiments with repeated items.

FREE VERBAL RECALL

The typical free verbal recall experiment involves reading a list of high frequency English words to the subject (Deese and Kaufman, 1957; Murdock, 1962). Following the reading the subject is required to recall as many of the words from the list as possible. Quite often list length has been a variable, and occasionally the time per item has been varied. Deese and Kaufman, for example, used lists of 10 and 32 items at one second per item. Murdock ran groups of 10, 15, and 20 items at two seconds per item, and groups of 20, 30, and 40 items at one second per item. The results are typically presented in the form of serial position curves: the probability of recall plotted against the item's position in the list. Examples of such curves have already been presented in Fig. 4.

It should be clear that this experimental situation can be analyzed within the framework of the buffer model. As the list is read to the subject, each item is postulated to enter the buffer and leave it in the usual fashion; and transfer to LTS is assumed to occur while the item is in the buffer. The type of retrieval scheme that must be postulated will be, in general, quite similar to the search processes already presented. However, there is one important difference. At the end of each trial the subject makes multiple responses (he reports out many different items) and the effect of these responses upon other items in memory has not previously been discussed. This problem is particularly acute in the case of items in the buffer, since it is a virtual certainty that making a response will disturb other items in the buffer. This statement is particularly relevant if one holds the kind of view proposed by Broadbent (1963) that the buffer acts as the input-output channel for the subject's interactions with the environment. In fact, Waugh and Norman (1965) have proposed that each response output has the same disrupting tendency upon other items in the buffer as the arrival of a new item.

On the other hand, it is not clear whether an emitted response disrupts items in LTS. At the very least, the act of recalling an item from LTS could be expected to raise that item's strength in LTS, or to increase the number of copies of that item in LTS. This paper is not the place for further speculations of this sort. The approach that will be followed here will be to assume that the retrieval of an item from LTS has no effect upon the store. Furthermore, the studies to be considered next incorporate an experimental procedure to clear out the buffer before the recall responses are requested, hence eliminating the need to examine effects related to the buffer.

FREE VERBAL RECALL EXPERIMENTS

Within the framework of the free verbal recall task described above, several experiments have used an arithmetic task interpolated between the end of the list and the test in order to eliminate recency effects. In an experiment by Postman and Phillips (1965) the interpolated task was counting backwards by three's and four's, a procedure originated by Peterson and Peterson (1959). In an unpublished experiment by Shiffrin the interpolated task consisted of serial addition; this experiment will now be presented in some detail.

Stimulus items were common English words. Lists of 6, 11, and 17 words were presented to the subjects at rates of either one or two seconds per word. Four conditions were run: (1) no interpolated arithmetic and immediate recall of the list; (2) 45 seconds of interpolated arithmetic and then recall; (3) no interpolated arithmetic, but a 45-second wait before recall; (4) 45 seconds of interpolated arithmetic, followed by a 45-second wait, followed by recall. In a two-hour session each subject was run twice under each of the conditions (rates of presentation and list length). Thus, 48 lists were given in a randomly mixed order. The only conditions of interest for this paper are those using interpolated arithmetic. The stimulus items were presented sequentially via a slide projector, but the arithmetic task was conducted aurally in the following manner: the slide following the last slide in the list presented a three-digit number and was removed. The experimenter then read a list of random digits from the set 1 to 9, one every three seconds. The subject was required to cumulatively add these to the original three-digit number, and report the total before recalling the words of the list. The fact that the 50 subjects were run in groups of about 12 each, plus the large number of different experimental conditions, tended to make the data somewhat variable, but for the rough analysis that will be presented here, they will be adequate. The data is shown in Fig. 13. For this experiment it is important to remember that the first item presented (the oldest) is labeled number 1 and is graphed to the far left.* Thus the upswing to the left represents a primacy effect; the recency effect, which would be to the right, has been eliminated.

These results are supported by the experiment of Postman and Phillips (1965). In that experiment the intervening task was counting backwards by three's or four's. In the condition of interest, the intervening task took 30 seconds. A control group had no intervening task. Three list lengths were used: 10, 20, and 30. The presentation time per item was always one second. Figure 14 shows the serial position curves for the control group and the arithmetic group.

The data, viewed from the vantage of the buffer model, make it clear that the arithmetic manipulation has achieved the effect of eliminating recall from the buffer. Thus, the primacy effect remains unchanged (because, for all but very short lists, the first items presented are recalled solely from LTS), but the last items presented are removed from the buffer by the intervening arithmetic and therefore can be retrieved only from LTS.

An explanation need be given here for the level asymptote that extends to the right-hand side of the graphs. The buffer model as stated in Models

---

*This is reversed from the numbering scheme used to describe the Phillips and Atkinson study.

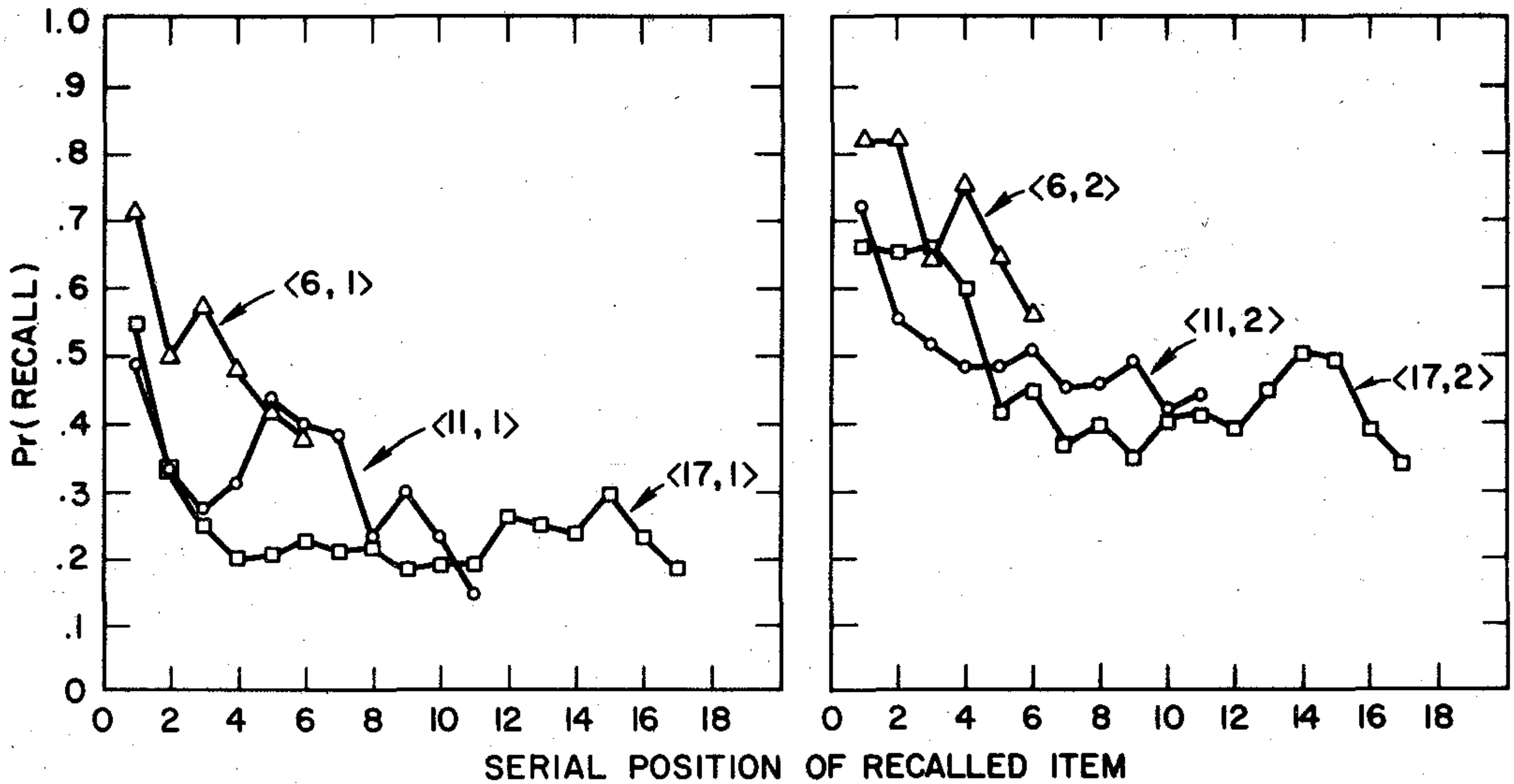

Fig. 13. Serial position curves for a free verbal recall task with interpolated arithmetic.

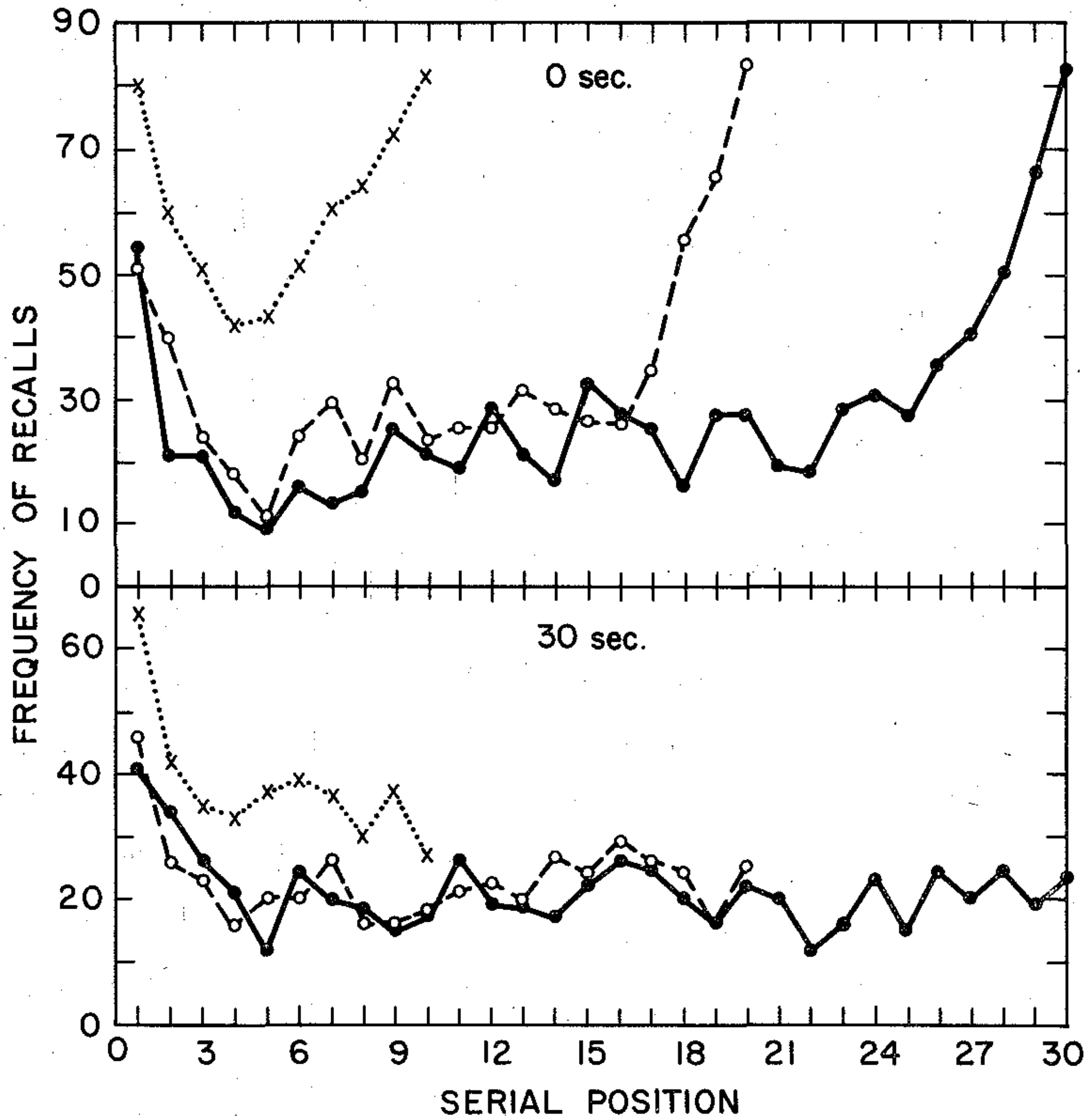

Fig. 14. Serial position curves for a free verbal recall task with interpolated arithmetic (after Postman and Phillips, 1965).

I, II, and III would predict that the probability of recall would go to zero for the last item input since that item could not be in LTS. That formulation, however, assumed that the test occurs immediately following presentation of the list. The assumption we will make concerning intervening arithmetic is that it clears the buffer in the same fashion and at the same rate as new incoming stimulus items.* Thus the last item presented could be expected to stay in the buffer for the same mean time as any other item which is input to a full buffer. This assumption will be formally stated in the theory to follow.

It should be noted that in Shiffrin's experiment the subjects did not know when a list would end. For this reason the observed drop in probability of recall from list length 6 to list length 17 cannot be explained by changes in the subjects' motivation from one list length to another. Furthermore, the fact that the subject does not know when the list will end is an indication that the δ parameter should be quite small. Hence, we shall let $\delta \to 0$, which means that we have one less parameter to estimate.

The model to be applied here is essentially the strength model discussed earlier with a few minor changes to accommodate the new experimental situation. As noted earlier the intervening arithmetic task is assumed to knock out items from the buffer at the same rate and in the same manner as additional new items. Thus the quantities $\omega_{ij}^{(d)}$ and $S(d)$ presented in Eqs. 14 and 16 must be modified to take this extra factor into consideration. First of all, $\omega_{ij}^{(d)}$ is no longer cut off at the end of the list proper as it was earlier. It is therefore defined for all j. (For all

*For evidence on this point, see Waugh and Norman (1965).

practical purposes this is true: for large j, $\omega_{ij}^{(d)}$ will be essentially zero and it is not important to consider the cutoff which occurs at the end of the intervening arithmetic.) Hence

$$\omega_{ij}^{(d)} = \begin{cases} \xi_{r,j} & , \text{ if } i \leq d-r+1 \\ \xi_{d-i+1,j-i+d+1-r} & , \text{ if } i > d-r+1 \text{ and } j > i-d+r-1 \\ 0 & , \text{ if } i > d-r+1 \text{ and } j \leq i-d+r-1. \end{cases} \tag{25}$$

Secondly, $S(d)$ represents the total strength in LTS which is now greater than before (see Eq. 16) because some items are in the buffer longer. The new value for $S(d)$ is as follows:

$$S(d) = \left\{ r(d-r)\theta_r + \left[\sum_{i=1}^{r} (i\theta_i) \right] + r(r-1)\theta_r \right\} t\gamma \tag{26}$$

where the last term in the brackets denotes the mean extra time items stay in the buffer. This means that $S(d)$ is now an expectation rather than a fixed value, but the variance of the last term in the brackets is quite small compared to the magnitude of $S(d)$ so that the approximation is fairly accurate. Thirdly, the probability of a hit is the same as before:

$$h_{ij}^{(d)} = \frac{\lambda_{ij}^{(d)}}{S(d)} .$$

It is now time to propose a retrieval scheme to apply to the present experiment. The first requirement this scheme should satisfy is that the probability of retrieval depends at least in part upon the <u>absolute</u> strength of an item in LTS. The postulate that will be used here is as follows: if a search of LTS is made and the i^{th} item is found, then the probability that the i^{th} item will be correctly reported is

$$1 - \exp\left[- \lambda_{ij}^{(d)}\right] .$$

For this equation, the probability of recall will go to 1 as $\lambda_{ij}^{(d)}$ becomes large, and will be zero for $\lambda_{ij}^{(d)} = 0$.

The final retrieval postulate holds that R searches are made into LTS, and on each search the probability of picking the i^{th} item is $h_{ij}^{(d)}$. Each time the i^{th} item is picked the probability that the subject is capable of reporting it is $1 - \exp\left[- \lambda_{ij}^{(d)}\right]$. Thus

$$\rho_{ij}^{(d)} = 1 - \left\{1 - h_{ij}^{(d)}\left[1 - \exp(- \lambda_{ij}^{(d)})\right]\right\}^{R} ,$$

and, from Eq. 17,

$$\Pr\left[C_i^{(d)}\right] = \sum_j \left[\omega_{ij}^{(d)}\right]\left[\rho_{ij}^{(d)}\right] ,$$

since it is assumed that the guessing probability is zero.

It has already been stated that we will set $\kappa_i = 1/r$ for all i; that is, δ is assumed to be arbitrarily close to zero. Further, to simplify the analysis, we will assume that all of the θ_j's are equal. This assumption means that the primacy effect is not due to a faster rate of transfer of the early items in the list, but due solely to the longer time spent by these items in the buffer. A fuller discussion of this problem will come later, but it is obvious that the assumptions concerning the θ_j's and the assumptions concerning retrieval are interrelated; it should be kept in mind that a retrieval function which works well given the equal θ_j assumption may be quite different from the best retrieval function for an unequal θ_j assumption.

Under the above simplifying assumptions, the mathematics of this model becomes quite simple. The results are as follows:

$$\omega_{ij}^{(d)} = \begin{cases} \frac{1}{r}\left(\frac{r-1}{r}\right)^{j-1} & \text{, for } i \leq d - r + 1 \\ \frac{1}{r}\left(\frac{r-1}{r}\right)^{j-i+d-r} & \text{, for } i > d - r + 1 \text{ and } j > i - d + r - 1 \\ 0 & \text{, for } i > d - r + 1 \text{ and } j \leq i - d + r - 1 \end{cases}$$

$$\lambda_{ij}^{(d)} = jt\gamma$$

$$S(d) = [dr + \tfrac{1}{2}r(r+1)]t\gamma$$

$$h_{ij}^{(d)} = \frac{j}{dr + \frac{1}{2}r(r+1)}$$

$$\rho_{ij}^{(d)} = 1 - \left\{1 - h_{ij}^{(d)}\left[1 - \exp(-\lambda_{ij}^{(d)})\right]\right\}^R$$

and

$$\Pr\left[C_i^{(d)}\right] = \sum_j \left[\omega_{ij}^{(d)}\right]\left[\rho_{ij}^{(d)}\right] .$$

Thus we have the probability of reporting item i as a function of three parameters: r, γ, and R. The parameter r will be estimated again by independent means; in most of the serial position curves shown, the primacy effect extends over three or four items. Hence r is set equal to 4. The number of searches, R, also must have certain restrictions placed upon it. For example, although the mean number of items reported out per list is generally quite small, occasionally subjects will report a very large number of items. Since the number of items reported cannot be greater than the number of searches made, the latter number must be fairly large. We therefore set R equal to 30; this value was selected arbitrarily but as we shall see, it yields good fits. Finally, the parameter γ was estimated on the basis of a best fit to the 17-item list in the Shiffrin experiment. The estimate of γ, .05, was then used to calculate theoretical serial position curves for all the conditions in the Shiffrin study and the first portions of the longer Murdock curves. It should be clear that for Murdock's 30 and 40 word lists, performance on the middle items is that which would be found even if arithmetic was given at the end, since there is very little likelihood that the first 15 or so items are still in the buffer at the finish of the list. The results are shown in Fig. 15, where the observed points are the same as the ones presented in Figs. 4 and 13.*

The fitting procedure used here is quite crude. Several assumptions were made solely to simplify the mathematics; two of the three parameters were set somewhat arbitrarily, and the final parameter was picked on the basis of a fit to only a single curve. Nevertheless, the fit (which is surely not optimal) provides a rather good description of the data. Table 4 gives the predicted and observed values for the first point in the list and the asymptote for each of the lists considered. The asymptotic value was obtained by averaging all points beyond list position three. The points for Murdock's 30 and 40 list lengths were recovered from Fig. 15b, and may be slightly inaccurate. It can be seen that, whatever the

---

*Postman's curves were not received in time to calculate theoretical curves for them but it can be seen that they fall approximately where they would be expected to lie on the basis of our fits to similarly sized lists.

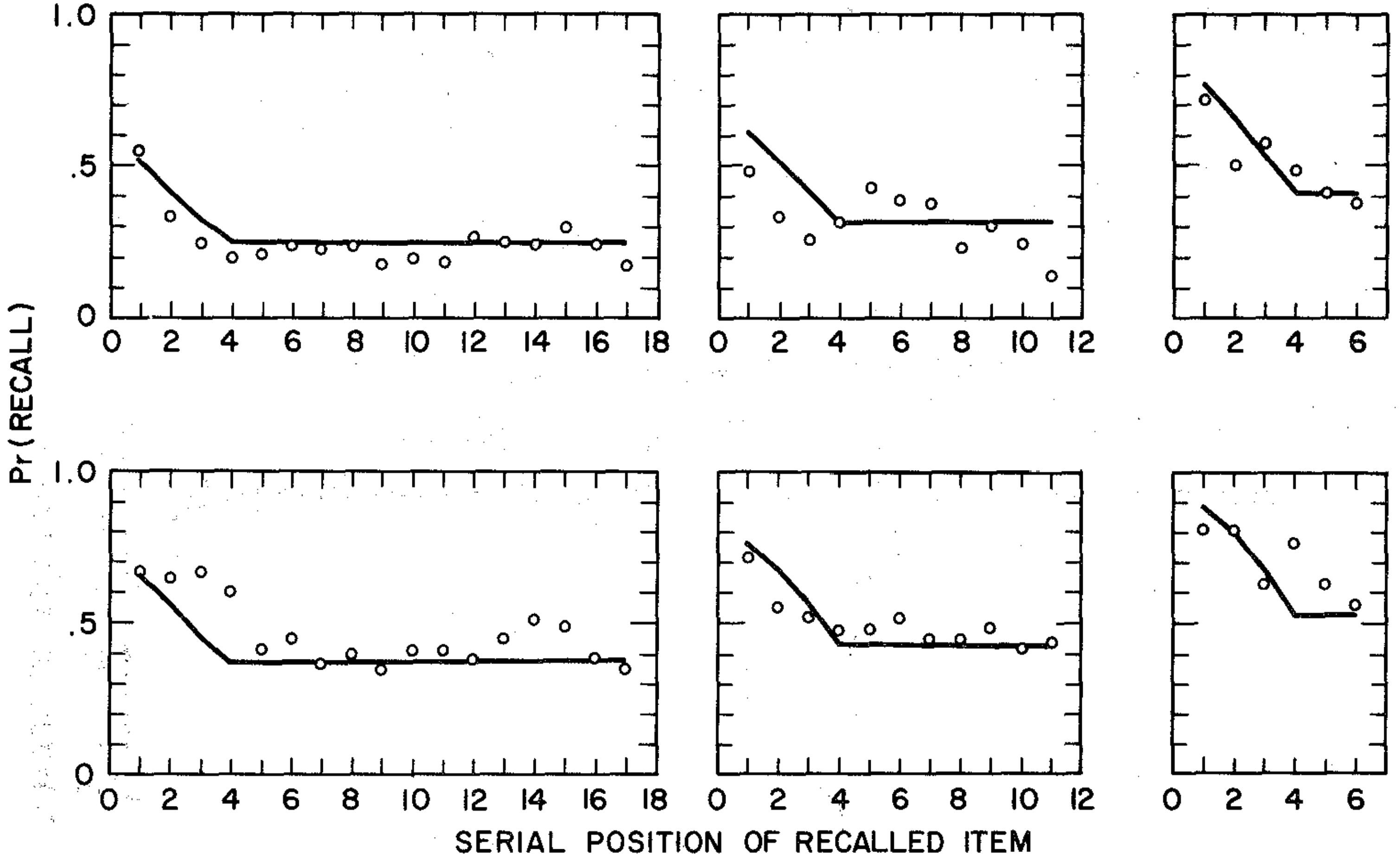

Fig. 15a. Observed and predicted values for the Shiffrin data (parameters: $r = 4$, $\gamma = .05$, $R = 30$, $\theta = 1$, $\delta \to 0$).

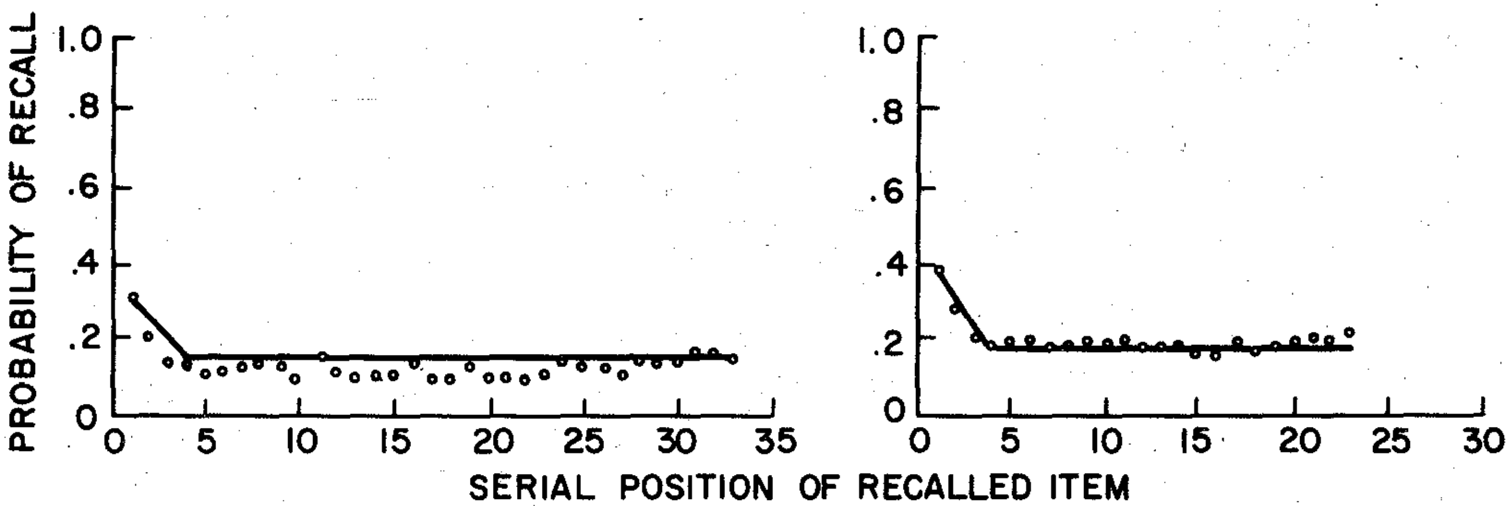

Fig. 15b. Observed and predicted values for the Murdock data (parameters: $r = 4$, $\gamma = .05$, $R = 30$, $\theta = 1$, $\delta \rightarrow 0$).

TABLE 4

Fit of the Strength Model to the Data of Shiffrin and Murdock (the condition is specified by the triple: experimenter, list length, and exposure time)

	First Position		Asymptote	
Condition	Observed	Expected	Observed	Expected
S-6-1	.72	.77	.42	.42
S-6-2	.82	.89	.61	.53
S-11-1	.48	.62	.38	.32
S-11-2	.73	.77	.45	.43
S-17-1	.55	.51	.24	.25
S-17-2	.67	.66	.42	.36
M-30-1	.39	.37	.19	.18
M-40-1	.30	.30	.13	.14

inadequacies of the fitting procedure, the results are quite good and the viability of two principal features of the model has been demonstrated. First, the assumption that the storage process is a function of the time spent in the buffer has proved to be quite reasonable in fitting lists in which the presentation time per item was varied. Secondly, while the precise retrieval scheme used undoubtedly depends upon the assumption made concerning the θ_j's, the assumption that the retrieval from LTS depends not only on relative strength but also on absolute strength has proved to be workable. A generalization of the model and a further discussion of retrieval schemes dealing with this question will be presented in the next section.

SOME GENERALIZATIONS

STRENGTH VS. MULTIPLE-COPIES

Two proposals were made in the first part of this paper concerning what is stored in LTS: strength, or multiple copies. A model embodying the first proposal has already been presented. We would now like to show that the multiple copy proposal is an exact counterpart of the strength notion. First recall Model I where in each unit of time an item had a probability θ_j of being copied in LTS, but once in LTS no additional copies could be made. The multiple-copy correlate of this would let the item be copied in LTS during one unit of time with probability θ_j, but more than one copy could be made in successive units of time. Thus if the items were presented at a one-second rate and item i stayed in the buffer for ten seconds, then the number of copies made would be integrally distributed with a minimum of 0 copies to a maximum of 10. What would happen, however, if the items were presented at two seconds per item? Can one copy be made each second of the item's stay in the buffer or can one copy be made during each two-second interval? Considerations like these suggest that a more general conception of the multiple-copy notion is that in each small unit of time one copy can be made with some small probability.

This statement, however, is no more than a definition of the Poisson distribution. For this reason the assumption is made that the number of copies made of item i is a Poisson function of the weighted time that the i^{th} item spends in the buffer. In the terminology already introduced, $\mu_{ij}^{(d)}$ is the weighted time spent in the buffer by the i^{th} item in a list of length d, given that the i^{th} item stayed in the buffer for j units of time ($\mu_{ij}^{(d)}$ is defined in Eq. 15). Thus the probability that k copies are made of the i^{th} item in a list of length d, given that this item stayed in the buffer j units of time, is:

$$\frac{\left[\gamma\mu_{ij}^{(d)}\right]^k}{k!} \exp\left[-\gamma\mu_{ij}^{(d)}\right]$$

where γ is the same rate parameter introduced earlier.

This process is now an exact counterpart, though discontinuous, of the strength process. If the weighted time an item spends in the buffer is doubled, the strength is doubled and alternately, so too is the expected number of copies. Similarly, just as the probability of picking item i in one search is the ratio of the strength of item i to the total strength, so the probability of picking item i in terms of the multiple-copy process is the ratio of the number of copies of item i to the total number of copies. The final indication of the similarity between the two approaches is the fact that the expected number of copies made of item i is $\gamma\mu_{ij}^{(d)}$, which is the same quantity that defines the strength process.

The reason for developing the strength process rather than the multiple-copy process can now be seen; the multiple-copy process is mathematically more complex, having an extra distribution, the Poisson. There is a reasonable alternative to both these processes, however, as will be seen in the next section.

<u>WHAT</u> <u>IS</u> <u>STORED?</u>

If an item is considered as an array of pieces of information, an alternative to the above schemes suggests itself. For example, the multiple-copy proposal may be set forth in the following manner. Suppose item i consists of bits (in the loose sense) of information. It may then be assumed that each copy is a random sample of m of these bits. Each of these partial copies, of course, may overlap others that have already been stored. For this reason, the amount of new information contributed by each new copy is a decreasing function. Now in the multiple-copy scheme defined above, a search into LTS is made by picking a single copy; this means that the probability of picking a copy of the i^{th} item is the ratio of the number of copies of the i^{th} item to the total number of copies in LTS. The information model, on the other hand, could be postulated to act as follows: what is stored in LTS is bits of information rather than copies; these bits are stored no more than once each. A search into LTS is then made by picking randomly one bit of information from the store. The probability of choosing a bit of information relevant to item i would then be the ratio of the number of stored bits making up item i to the total number of stored bits.

This "information" model has a different mathematical form than the earlier models. For example, if each copy contains a proportion p of the total number of bits making up an item, then the proportion of bits left to be stored after n copies have been made is $(1-p)^n$. Thus the proportion already stored is $1-(1-p)^n$. This can be rewritten $1 - \exp[n \log(1 - p)]$. Consider n to be the mean number of copies made in j units of time. Since the Poisson mean is a linear function of the weighted time the item spends in the buffer, $n = a\mu_{ij}^{(d)}$. Now let $a[\log(1-p)] = -\gamma$ and we can rewrite the proportion of bits already stored as $1 - \exp[-\gamma\mu_{ij}^{(d)}] = 1 - \exp[-\lambda_{ij}^{(d)}]$, which is the expression used earlier in the strength model for the probability of a recall, given that item i is picked. In terms of these remarks it is now clear that one interpretation of our earlier assumption is that the probability of recall is a direct function of the proportion of information stored about the item in question. This information model, remember, differs from the earlier one not in the probability that an item will be recalled once it is picked, but in the probability of picking the item in the first place. To illustrate this point, note that $h_{ij}^{(d)}$ for the strength model is

$$\frac{\gamma\mu_{ij}^{(d)}}{\sum_{i=1}^{d}\gamma\mu_{ij}^{(d)}},$$

whereas, for the information model $h_{ij}^{(d)}$ is

$$\frac{1 - \exp[-\gamma\mu_{ij}^{(d)}]}{\sum_{i=1}^{d}\left\{1 - \exp[-\gamma\mu_{ij}^{(d)}]\right\}}.$$

While still considering the information model, we will examine a retrieval assumption that has been mentioned several times without explanation. The assumption holds that an item can be picked during a search of LTS, but not necessarily reported. This notion is given support if one imagines that a small portion of the information making up any item can be picked on a single search. On any one search this information may be insufficient to actually report the correct answer with assurance. On the other hand the idea of a small portion of information being available gives a natural explanation for the difference between recall and recognition measures of retention: the smaller the choice set the subject is given, the more likely that his partial information will be enough to allow him to choose the correct answer.

Before the information model can be further elaborated, it will be necessary to specify the function relating the number of information bits to the probability of recall. This question once again returns us to the problem of the retrieval process. The next section will consider the problem in a general fashion and examine some of the assumptions which have been used in earlier parts of the paper.

THE RETRIEVAL PROCESS

In the course of the paper two retrieval processes have been suggested: an active disruption of LTS caused by the ongoing search, and an imperfect search in which items, about which some information is present in LTS, are not reported. The first of these is conceptually clear and does not need additional discussion here. The second process, however, requires clarification.

The first problem to consider is how successive lists are kept separate from each other by the subject. In free recall, for example, different lists of words are presented from trial to trial, and the subject is required to output all the items he can recall after each list. The items in each list supposedly are copied in LTS, but in our analysis the subject searches only through the items of the very last list. It does not strike the authors as particularly desirable to assume that LTS is also nothing more than a buffer which is wiped clean after each trial. In addition to the complexities that this would add to the model, this view gives no easy explanation of insertions in recall of items from previous lists. Rather it is our view that a random search process is a fictional ideal which is only approximated by any given subject. The subject undoubtedly makes a non-random search of LTS, but along a dimension unknown in any one case to the experimenter. The most likely dimension is a temporal one; thus the subjects would search among those bits of information which tell him how long ago the item was presented. Furthermore, the subject would have to make a selective search along the temporal dimension in order to search only through the most recent items, and this observation would suggest that LTS is arranged in a fashion akin to an efficient cross-indexing system. Various such systems could probably be proposed in terms of the information input characterizing each item, but this will not be done here. The notion that the subject is always making ordered searches of memory along one or several dimension(s) is similar to the proposals made earlier concerning changes in the retrieval process over repeated trials. Further consideration along these lines is unfortunately beyond the scope of this paper. In any event, the earlier assumptions regarding random searches should be taken as an approximation which may be accurate, possibly, only on the very first trial in experiments with repeated items.

There is one other feature of the retrieval process that requires some elaboration; namely, the assumptions regarding the probability of correctly recalling the i^{th} item, given that information relevant to it is found in a search of LTS. The following proposal is made: when an item is picked a portion p of the total stored information on that item becomes available for consideration. This proportion p determines the independence of successive searches for an item. Thus if $p = 1$, all of the stored information about item i becomes available the first time item i is picked. If item i is not reported after this first pick then it will not be reported on any successive pick. On the other hand, if p approaches 0 successive picks will be almost independent of each other and the probability of recalling the item will not change from pick to pick. This second assumption is the one used in the strength model applied to the free verbal recall data, where the probability of retrieval was

$$1 - \left\{1 - h_{ij}^{(d)}[1 - \exp(-\lambda_{ij}^{(d)})]\right\}^R$$

if R picks, or searches, were made. If the first assumption was used, however, the probability of retrieval would be

$$1 - (1 - h_{ij}^{(d)})^R - [1 - (1 - h_{ij}^{(d)})^R][\exp(-\lambda_{ij}^{(d)})] .$$

The last problem to consider is when to terminate the search process. Many possibilities come to mind: stop after R picks; stop only after finding item i; stop after the response time runs out; stop after k

successive searches uncover items already previously picked. It seems likely that the stopping rule would be highly dependent on the experimental situation; the amount of time given for responding, the motivating instructions given the subject, the rewards for correct and incorrect answers, and so on. These same comments apply to a destructive search, where each search disrupts LTS in some manner.

CONCLUDING REMARKS

The similarities of the model presented here to other theories of memory should be briefly mentioned. Interference theory is represented in our model in three separate processes: the buffer, in which succeeding items knock out previous items; the destructive search process, where items in LTS can be modified by the search operation; and the imperfect retrieval process, which can produce interference-type effects. Decay theory, on the other hand, is not represented in the model as stated. The evidence for a decay process accumulated by Brown, Conrad and Peterson, among others, is not necessarily explainable by the model in its present form. Nevertheless, there is no reason why a decay process cannot be added to the buffer postulates. If this were done it would be assumed that rehearsal or attention is the mechanism by which a certain number, r, of items may be kept at one time in the buffer with none decaying. When another item enters, however, the buffer becomes overloaded and the rehearsal or attention factor cannot keep all the items from decaying. One item then decays and the buffer returns to its equilibrium state. A theory of this sort would incorporate the decay notion into the buffer postulates without changing the present form of the model.

One final area of research which has not been mentioned explicitly is the "chunk" hypothesis proposed by Miller (1956) and others. The chunk hypothesis generally takes two forms. The first, the reorganizing of material into successive chunks; and the second, chunk constancy, referring to a constancy in the rate of transmission of information over many experiments. Without going into details it can be said that the chunk hypothesis is related to the information structure in the buffer, and the organization of this information in LTS. Although this paper does not make explicit use of information-theoretic concepts, nevertheless they underly much of the development of the model. For example, the hypothesis that the buffer is of constant size in terms of information content, and the proposals that the search scheme changes and LTS is reorganized from trial to trial, are related to the chunk hypothesis.

The model in this paper was not applied to several areas where it might prove fruitful. For example, latency data can be given a natural interpretation in terms of the processing time required before outputting a response. The assumption would be that an item in the buffer at the time of test would have a latency distributed with a mean which was quite small, whereas any other item would have a latency determined by the search time. Thus, the latencies should be smallest for the most recent items and longest for the oldest items, irrespective of the serial position curve. This prediction has been borne out in a recent study by Atkinson, Hansen, and Bernbach (1964).

There are other areas in which the model would be applicable with the addition of a few specific hypotheses. Confidence ratings are an example that has already been mentioned. Another example is prediction of error types and intrusions, such as those examined by Conrad (1964). Predictions of this sort would require further delineation of the retrieval process, just as would confidence ratings.

Finally, it should be pointed out that of all the assumptions and variations which have been introduced, three are crucial to the theory. First is the set of buffer assumptions; i.e., constant size, push-down list, and so on. Second is the assumption that items can be in the buffer and LTS simultaneously. Third is what was called the retrieval process--the hypothesis that the decrement in recall caused by increasing the list length occurs as the result of an imperfect search of LTS at the time of test. Within this framework, we feel that a number of the results in memory and learning can be described in quantitative detail.

References

Atkinson, R. C., G. H. Bower, and E. J. Crothers. *An introduction to mathematical learning theory*. New York: Wiley, 1965.

Atkinson, R. C. and E. J. Crothers. A comparison of paired-associate learning models having different acquisition and retention axioms. *J. math. Psychol.*, 1964, *1*, 285-315.

Atkinson, R. C., D. N. Hansen, and H. A. Bernbach. Short-term memory with young children. *Psychon. Sci.*, 1964, *1*, 255-256.

Bower, G. H., Notes on a descriptive theory of memory. Paper read at Second Conference on Learning, Remembering, and Forgetting, September 1964, Princeton, New Jersey.

Broadbent, D. E. A mechanical model for human attention and immediate memory. *Psychol. Rev.*, 1957, *64*, 205-215.

Broadbent, D. E. *Perception and Communication*. New York: Pergamon, 1958.

Broadbent, D. E. Flow of information within the organism. *J. verb. Learn. verb. Behav.*, 1963, *4*, 34-39.

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

Brown, J. Short-term memory. *Brit. Med. Bull.*, 1964, *20(1)*, 8-11.

Brown, S. C. and W. F. Battig. Second-list paired-associate facilitation produced by addition of previously-learned first-list pairs. *J. verb. Learn. verb. Behav.*, 1966, in press.

Brown, R. and D. McNeill. The "Tip of the tongue" phenomena. *J. verb. Learn. verb. Behav.*, 1966, in press.

Calfee, R. C. and R. C. Atkinson. Paired-associate models and the effects of list length. J. math. Psychol., 1965, 2, 254-265.

Conrad, R. Acoustic confusions in immediate memory. Brit. J. Psychol., 1964, 55, 1, 75-84.

Conrad, R. and B. A. Hille. The decay theory of immediate memory and paced recall. Canad. J. Psychol., 1958, 12, 1-6.

DeFinetti, B. Methods for discriminating levels of partial knowledge concerning a test item. Brit. J. math. statist. Psychol., 1965, 18, 87-124.

Deese, J. and R. A. Kaufman. Serial effects in recall of unorganized and sequentially organized verbal material. J. exp. Psychol., 1957, 54, 180-187.

Egan, J. P. Recognition memory and the operating characteristic. Tech. Rep. Contr. No. AF 19(604)-1962, Indiana University, 1958.

Estes, W. K. Learning and memory. Paper read at California Statewide Lecture Series sponsored by National Aeronautics and Space Administration, 1965.

Estes, W. K. and H. A. Taylor, A detection method and probabilistic models for assessing information processing from brief visual displays. Proc. Natl. Ac. Sci., 1964, 52, 2, 446-454.

Feigenbaum, E. A. and H. A. Simon. A theory of the serial position effect. Brit. J. Psychol., 1962, 53, 307-320.

Greeno, J. G. Paired-associate learning with massed and distributed repetition of items. J. exp. Psychol., 1964, 67, 286-295.

Mackworth, J. F. The relation between the visual image and post-perceptual immediate memory. J. verb. Learn. verb. Behav., 1963, 2, 113-119.

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

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev., 1956, 63, 81-97.

Murdock, B. B., Jr. The serial position effect of free recall. J. exp. Psychol., 1962, 64, 482-488.

Peterson, L. R. Immediate memory: Data and theory. In C. N. Cofer (Ed.) Verbal learning and behavior: Problems and processes. New York: McGraw-Hill, 1963.

Peterson, L.R., and Margaret J. Peterson. Short-term retention of individual verbal items. J. exp. Psychol., 1959, 58, 193-198.

Phillips, J. L. and R. C. Atkinson. The effects of display size on short-term memory. Tech. Rep. 78, Institute for Mathematical Studies in the Social Sciences, Stanford Univer., 1965.

Postman, L. and Laura W. Phillips. Short-term temporal changes in free recall. Quart. J. exp. Psychol., 1965, 17, 132-138.

Sperling, G. The information available in brief visual presentations. Psychol. Monogr., 1960, 74, Whole No. 498.

Sternberg, S. Retrieval from recent memory: some reaction-time experiments and a search theory, 1963, unpublished manuscript.

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

Waugh, Nancy C. and D. A. Norman. Primary memory. Psychol. Rev., 1965, 72, 89-104.

Yntema, D. B. and F. P. Trask. Recall as a search process. J. verb. Learn. verb. Behav., 1963, 2, 65-74.