Technical reports authored by Richard M. Shiffrin while a graduate student at Stanford University. These six reports are part of a series of 323 reports published from 1955 to 1992 by the Institute for Mathematical Studies in the Social Sciences. The complete set of reports can be found at http://suppes-corpus.stanford.edu/imsss.html.

This volume was assembled by R. C. Atkinson to honor Shiffrin on the occasion of "A Festschrift for Richard M. Shiffrin" held at Indiana University, May 7-9, 2012.

#79 Mathematical Models for Memory and Learning, 1965

#96 Multi-Process Models for Memory with Applications to a Continuous Presentation Task, 1966

#107 Some Two-Process Models for Memory, 1966

#110 Human Memory: A Proposed System and its Control Processes, 1967

#127 Some Speculations on Storage and Retrieval Processes in Long-Term Memory, 1968

#137 Search and Retrieval Processes in Long Term Memory, 1968

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

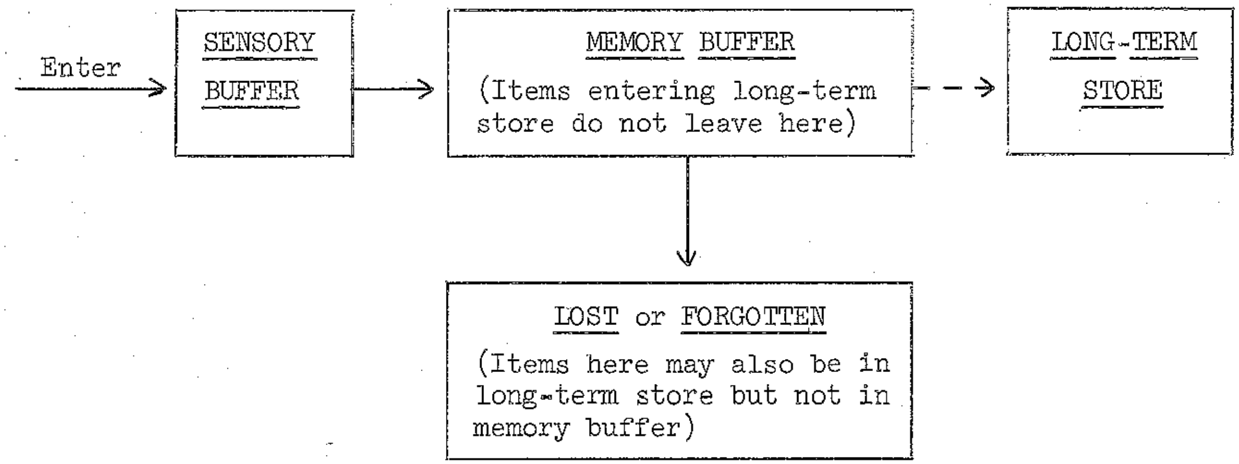


Fig. 1. Flow chart for the general system.

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.

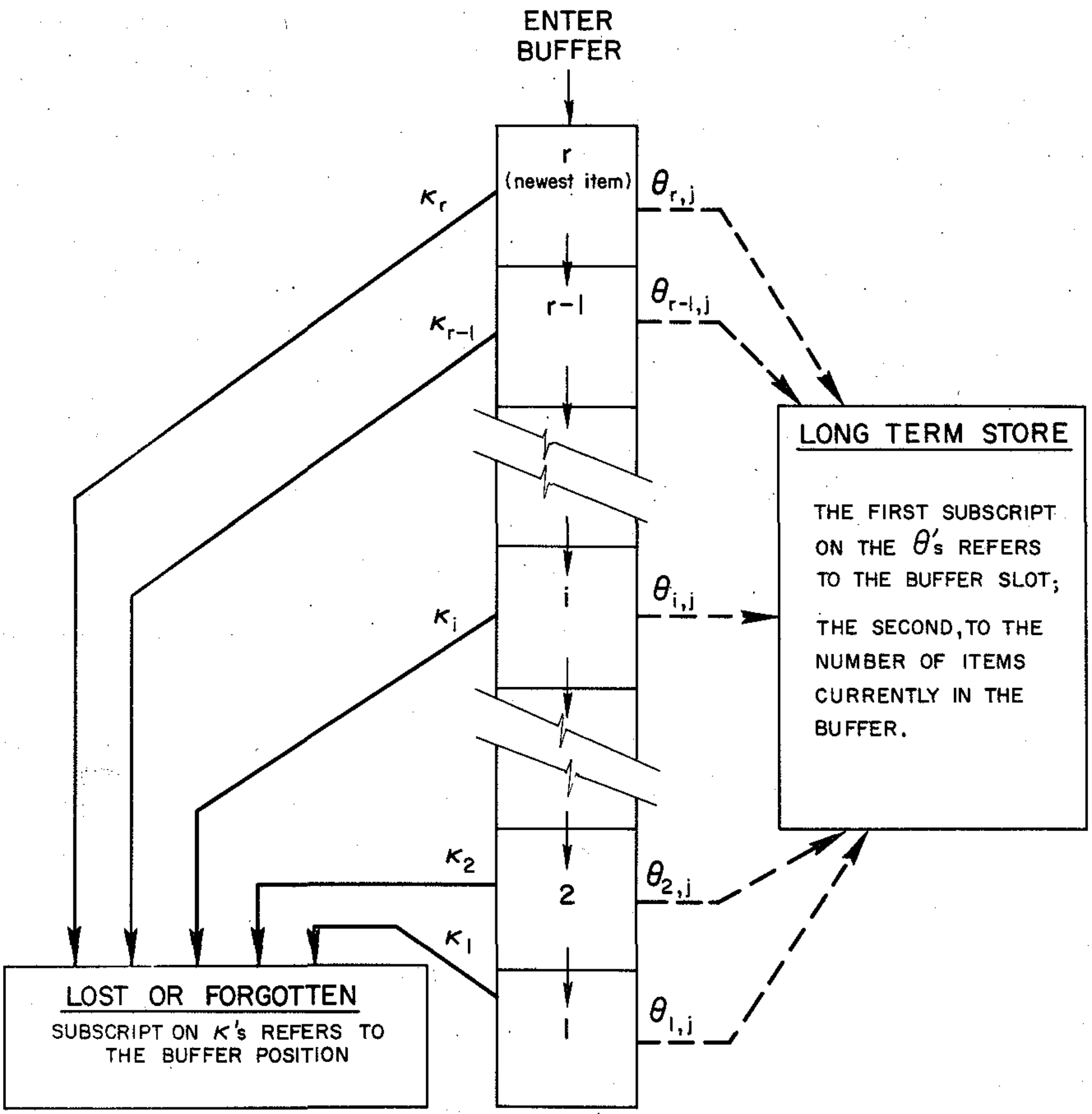


Fig. 2. Flow chart for the memory buffer.

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 $\delta \to 0$

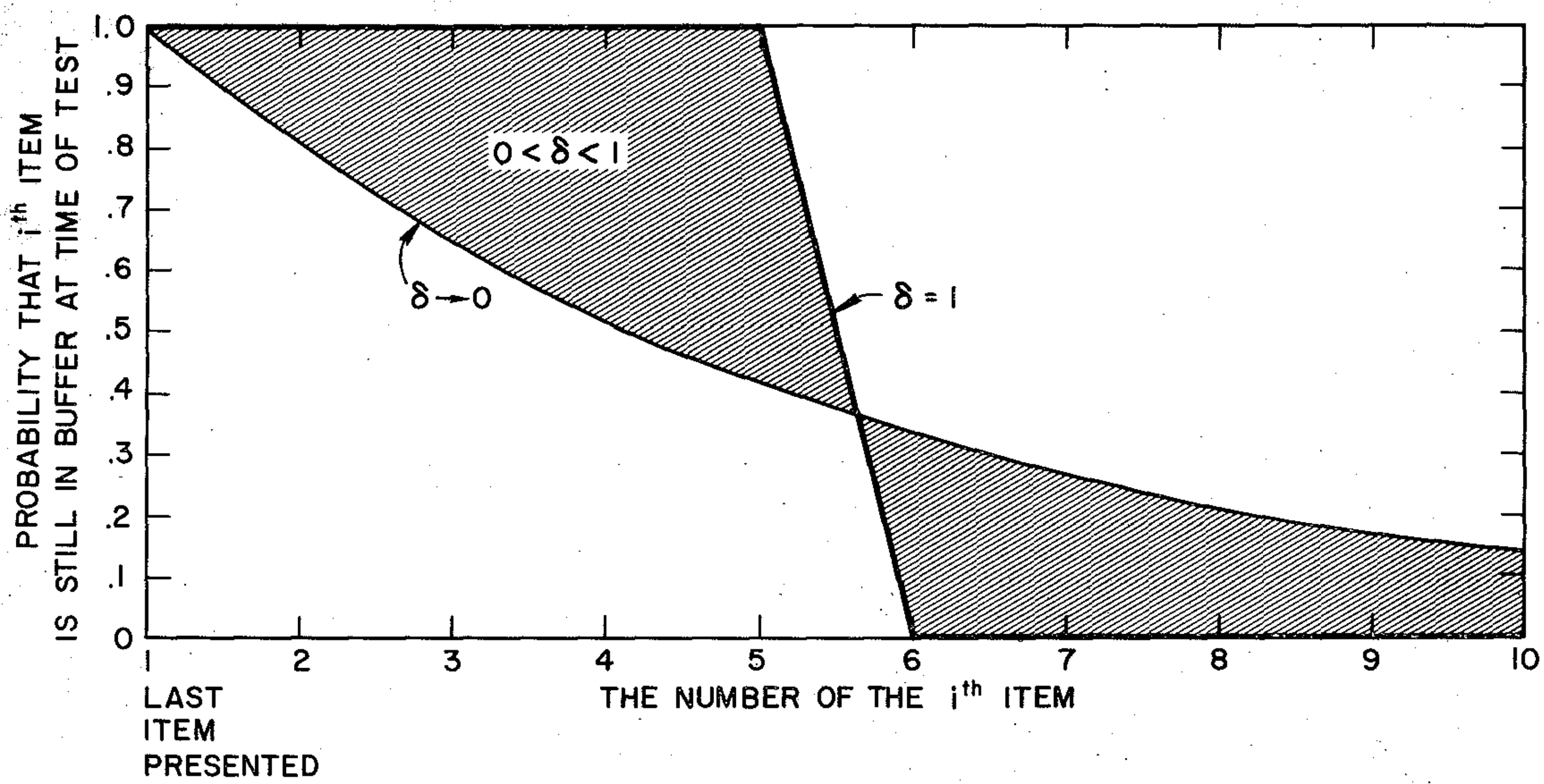


Fig. 3. Recency curves as a function of δ (the functions are computed for $r = 5$).

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. RETRIEVAL OF ITEMS FROM MEMORY

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

2) Retrieval from the lost state. 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) Retrieval from LTS. 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.

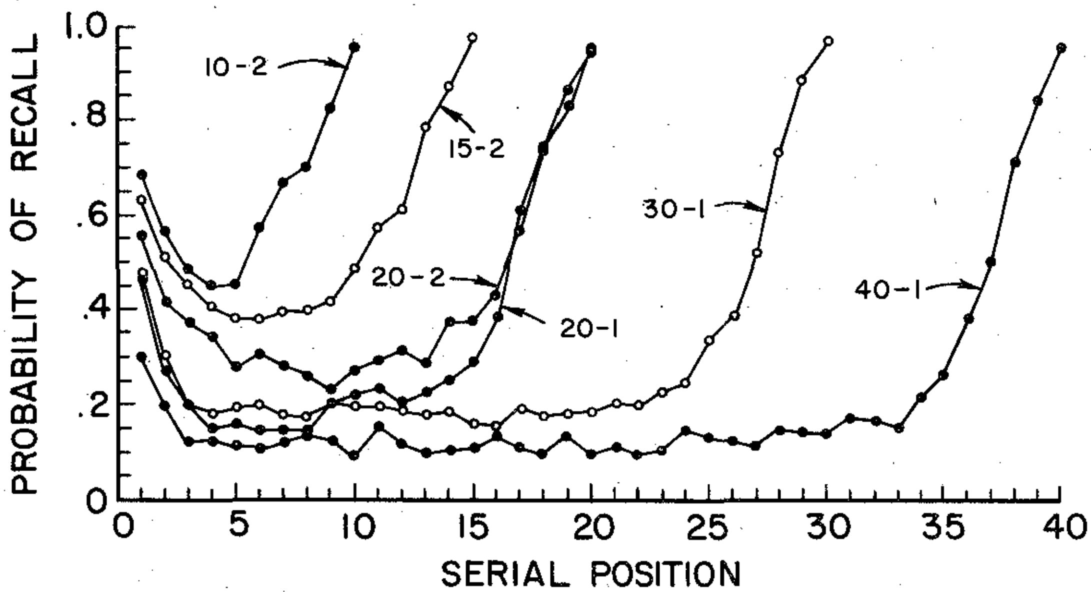


Fig. 4. Serial position curves for free verbal recall (after Murdock, 1962).

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

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

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

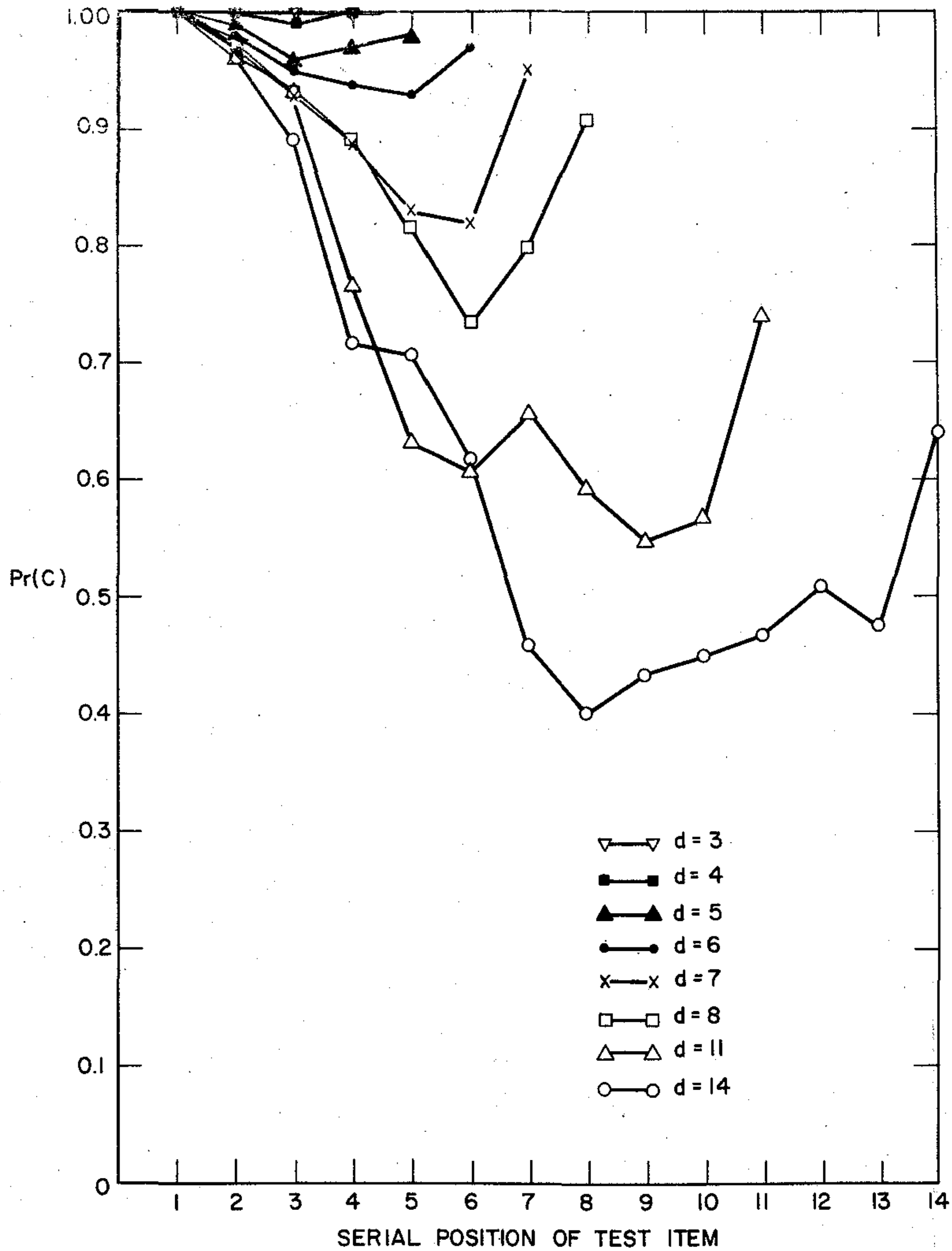


Figure 5. Proportion of items correctly recalled at each serial position for the various display sizes.

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:

$$\varphi_{2,j} = \kappa_2 + (1-\frac{\theta}{r})[\kappa_1\varphi_{1,j-1} + (\kappa_3 + \kappa_4 + \dots + \kappa_r)\varphi_{2,j-1}]$$

$$\varphi_{3,j} = \kappa_3 + (1-\frac{\theta}{r})[\kappa_1 + \kappa_2)\varphi_{2,j-1} + (\kappa_4 + \kappa_5 + \dots + \kappa_r)\varphi_{3,j-1}]$$

$$\vdots$$

$$\varphi_{i,j} = \kappa_i + (1-\frac{\theta}{r})[\kappa_1 + \kappa_2 + \dots + \kappa_{i-1})\varphi_{i-1,j-1} + (\kappa_{i+1} + \kappa_{i+2} + \dots + \kappa_r)\varphi_{i,j-1}]$$

$$\vdots$$

$$\varphi_{r-1,j} = \kappa_{r-1} + (1-\frac{\theta}{r})[(\kappa_1 + \kappa_2 + \dots + \kappa_{r-2})\varphi_{r-2,j-1} + \kappa_r\varphi_{r-1,j-1}] \qquad (2)$$

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

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:

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

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:

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

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 χ^2's for each list length. The result is a χ^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

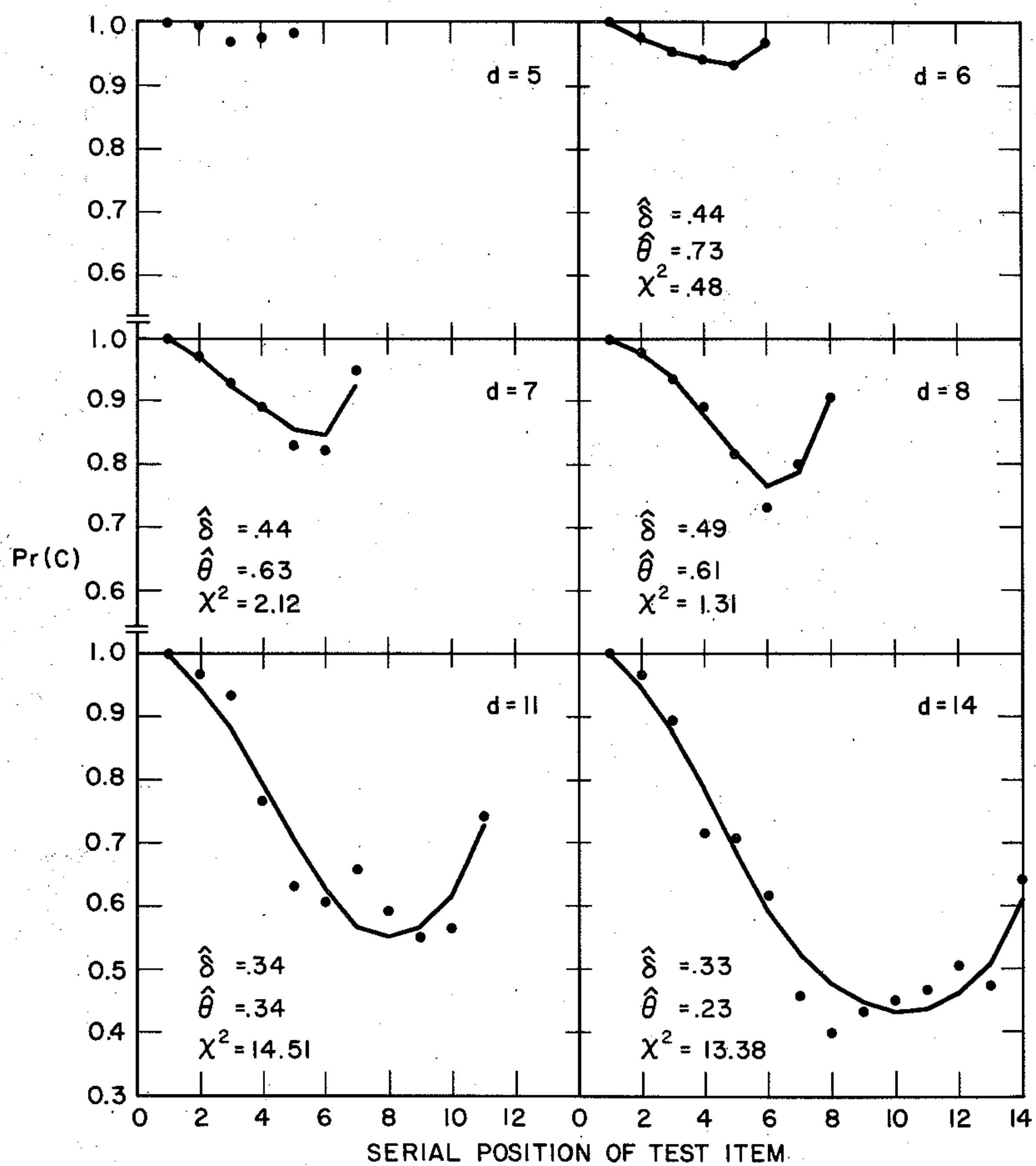


Fig. 6. Goodness-of-fit results for Model I ($\Sigma\chi^2(d)$ = 31.8 on 36 degrees of freedom).

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.

<u>Mathematical Development of Model II</u>

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:

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

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

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

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

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

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

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

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

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

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

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

We next define

$$\chi^2 = \chi^2(6) + \chi^2(7) + \chi^2(8) + \chi^2(11) + \chi^2(14) \tag{11}$$

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.

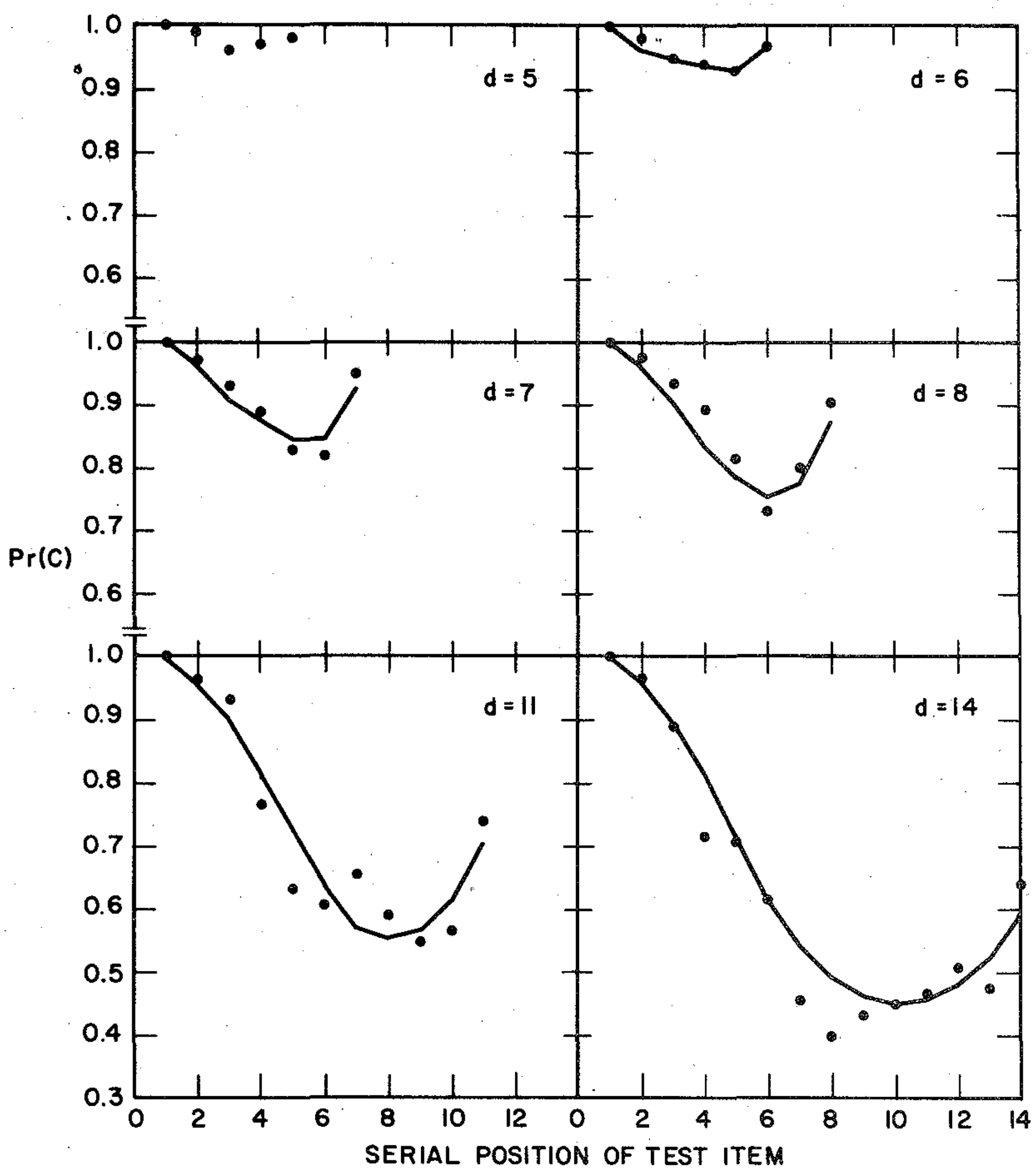


Fig. 7. Goodness-of-fit results for Model II_2 (parameter values: δ = .39, θ = .72, r = 5, R = 3.15, X^2 = 46.2 on 43 degrees of freedom).

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

$$\begin{aligned}\rho_i^{(d)} &= h_i^{(d)} + [1 - h_i^{(d)}]\,(1 - R')h_i^{(d)} + \left\{[1 - h_i^{(d)}](1 - R')\right\}^2 h_i^{(d)} + \ldots \\ &= h_i^{(d)}\left\{\sum_{j=0}^{\infty}[1 - h_i^{(d)}]^j (1 - R')^j\right\} \qquad (12) \\ &= \frac{h_i^{(d)}}{1 - [1 - h_i^{(d)}](1 - R')} .\end{aligned}$$

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.

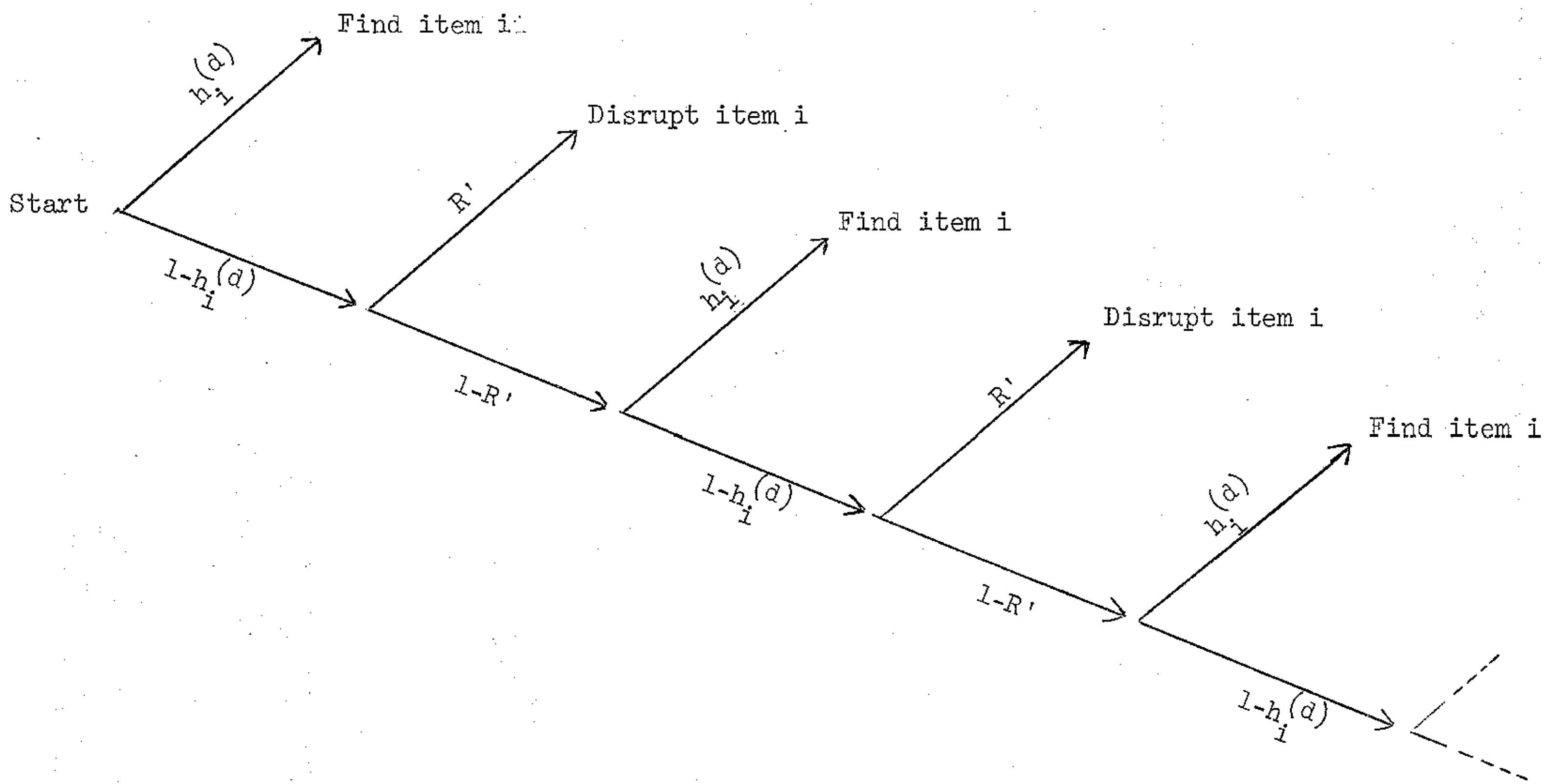


Figure 8. Retrieval process for Model III

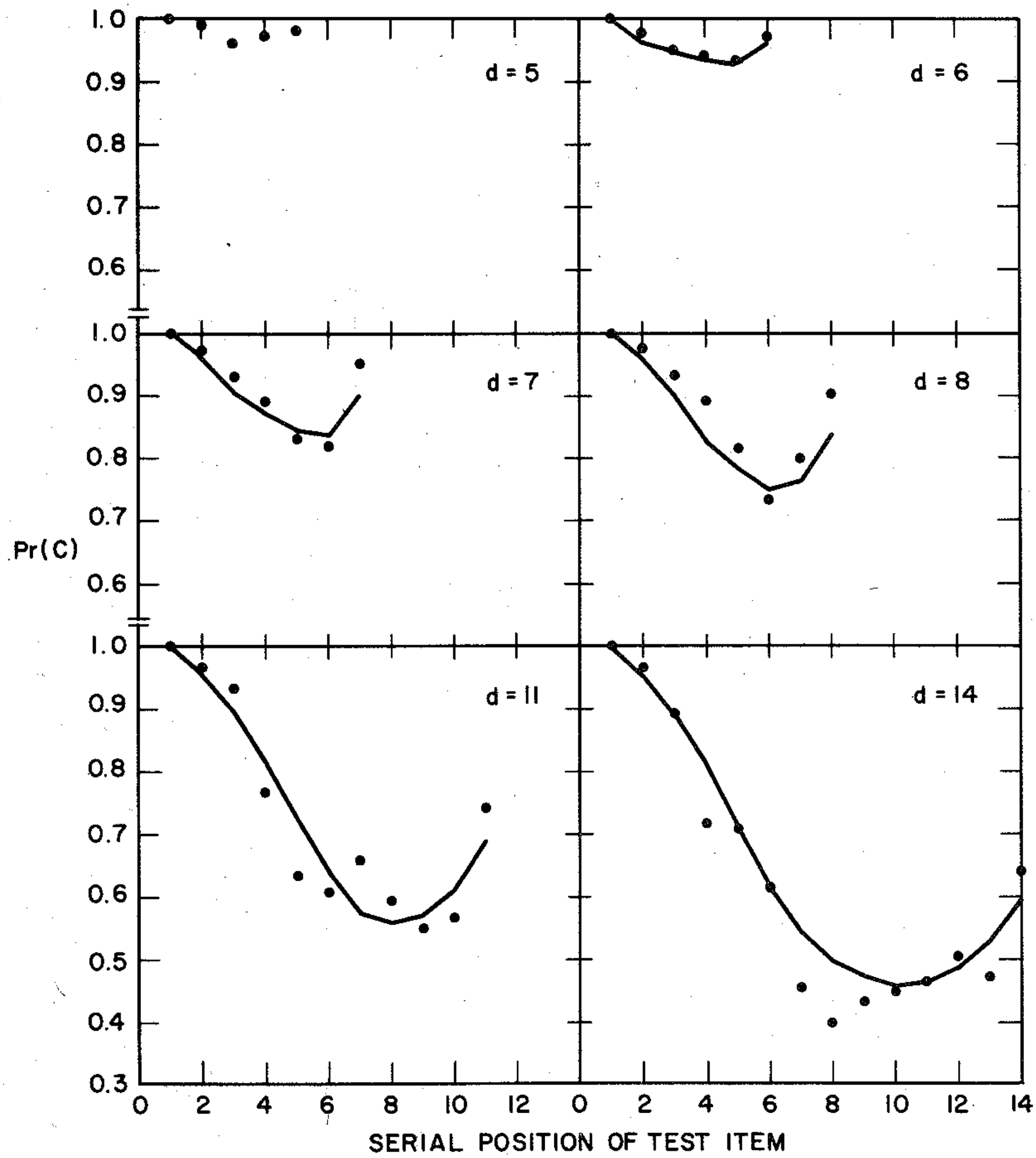


Fig. 9. Goodness-of-fit results for Model III (parameter values: $\delta = .38$, $\theta = .80$, $r = 5$, $R' = .25$, $\chi^2 = 55.0$ on 43 degrees of freedom).

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

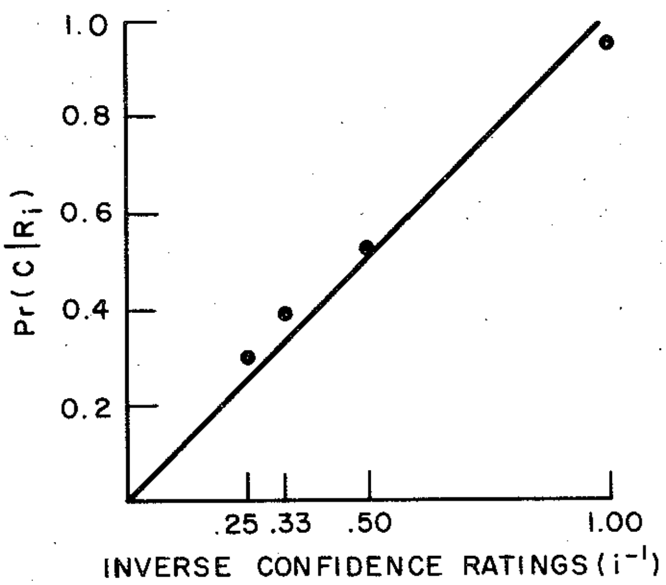


Fig. 10. Probability of a correct recall vs. reciprocal of the confidence ratings.

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:

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

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.

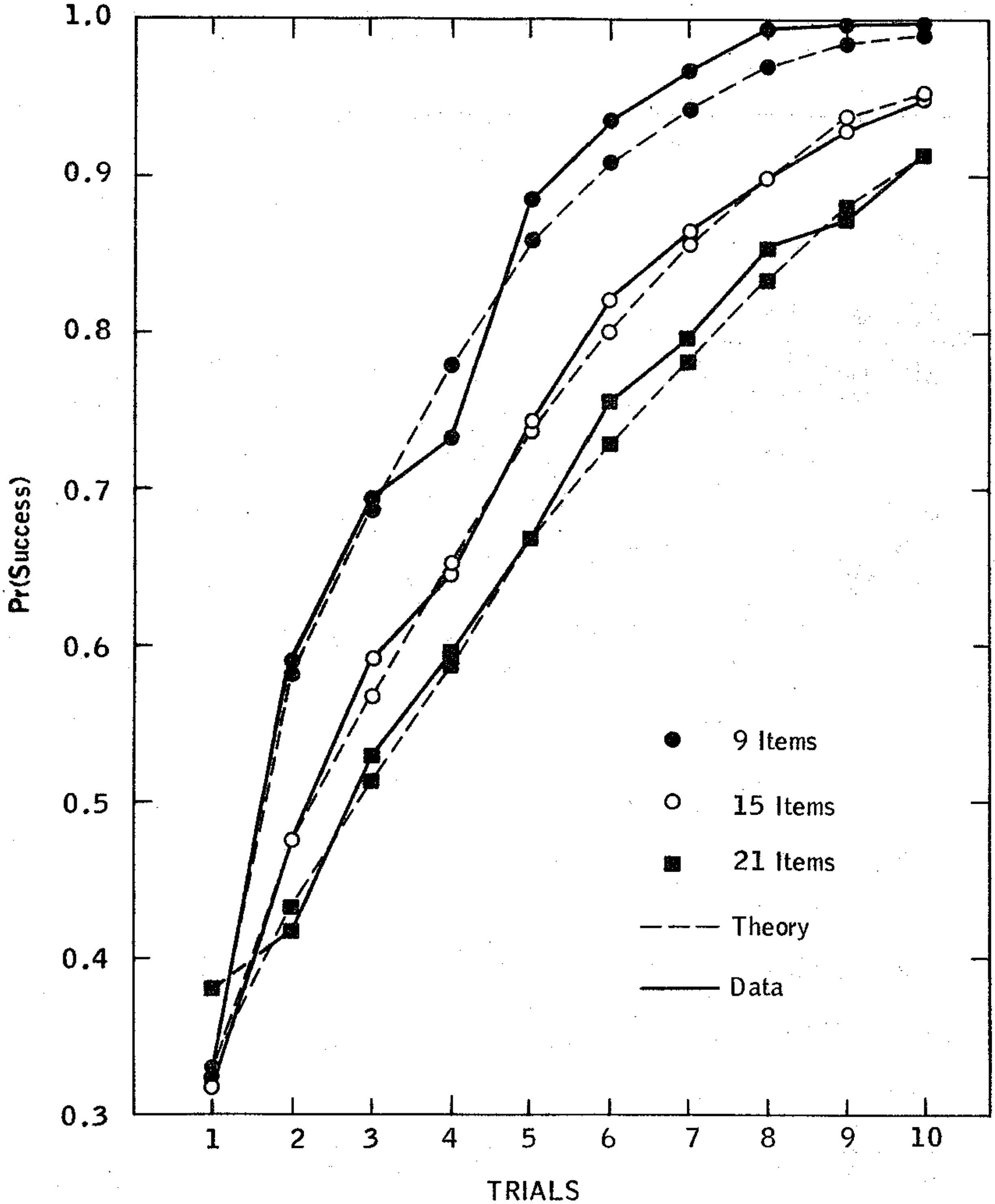


Fig. 11. Average probability of a success on trial n for three groups with different list lengths. See text for description of theoretical curves.

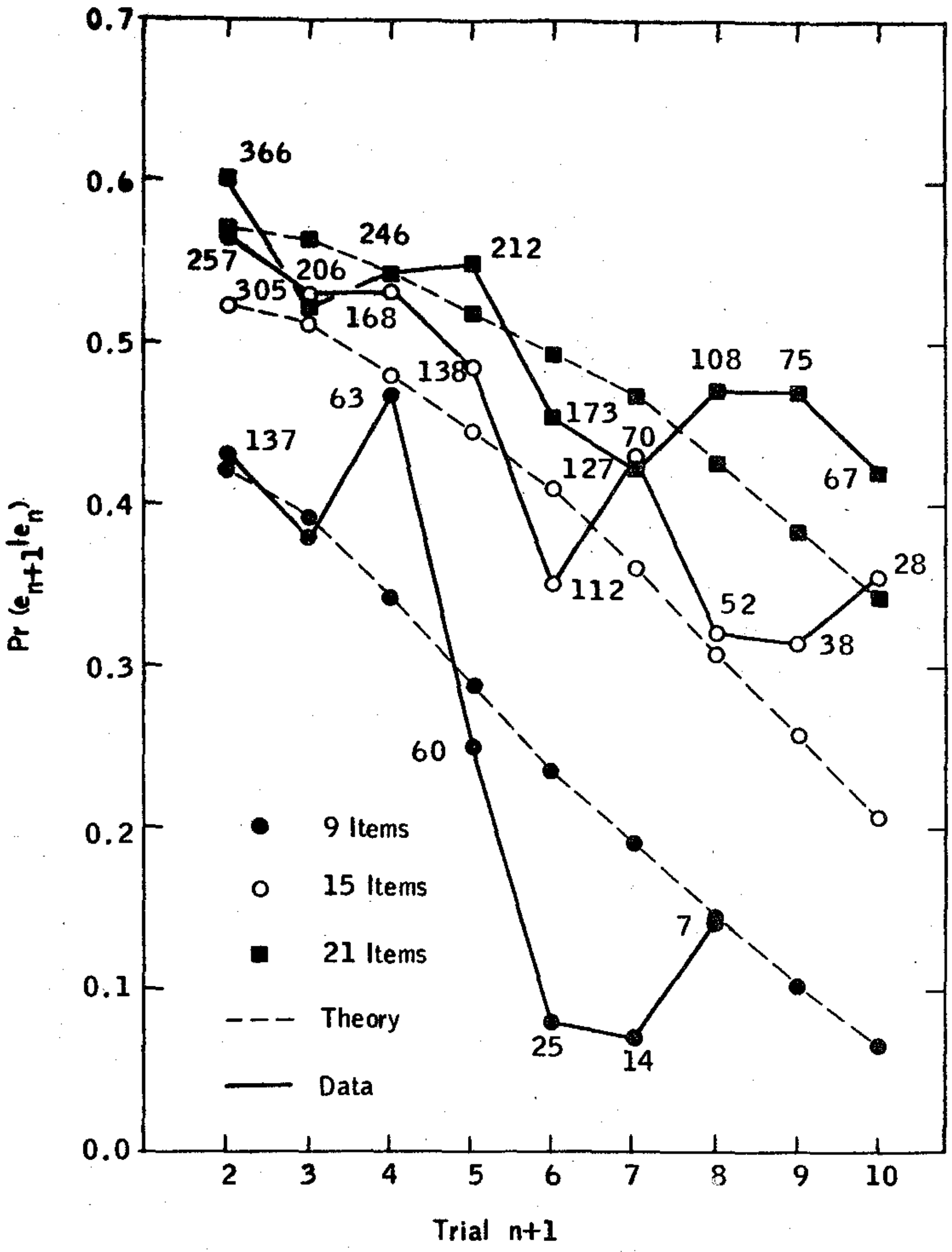


Fig. 12. Average probability of an error on trial n+1, given an error on trial n for three groups with different list lengths.

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				15 Items				21 Items			
	Obs.	TDF	Linear	One-element	Obs.	TDF	Linear	One-element	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

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

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.,

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

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

$$\chi^2 = \chi^2_9 + \chi^2_{15} + \chi^2_{21} \tag{24}$$

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.

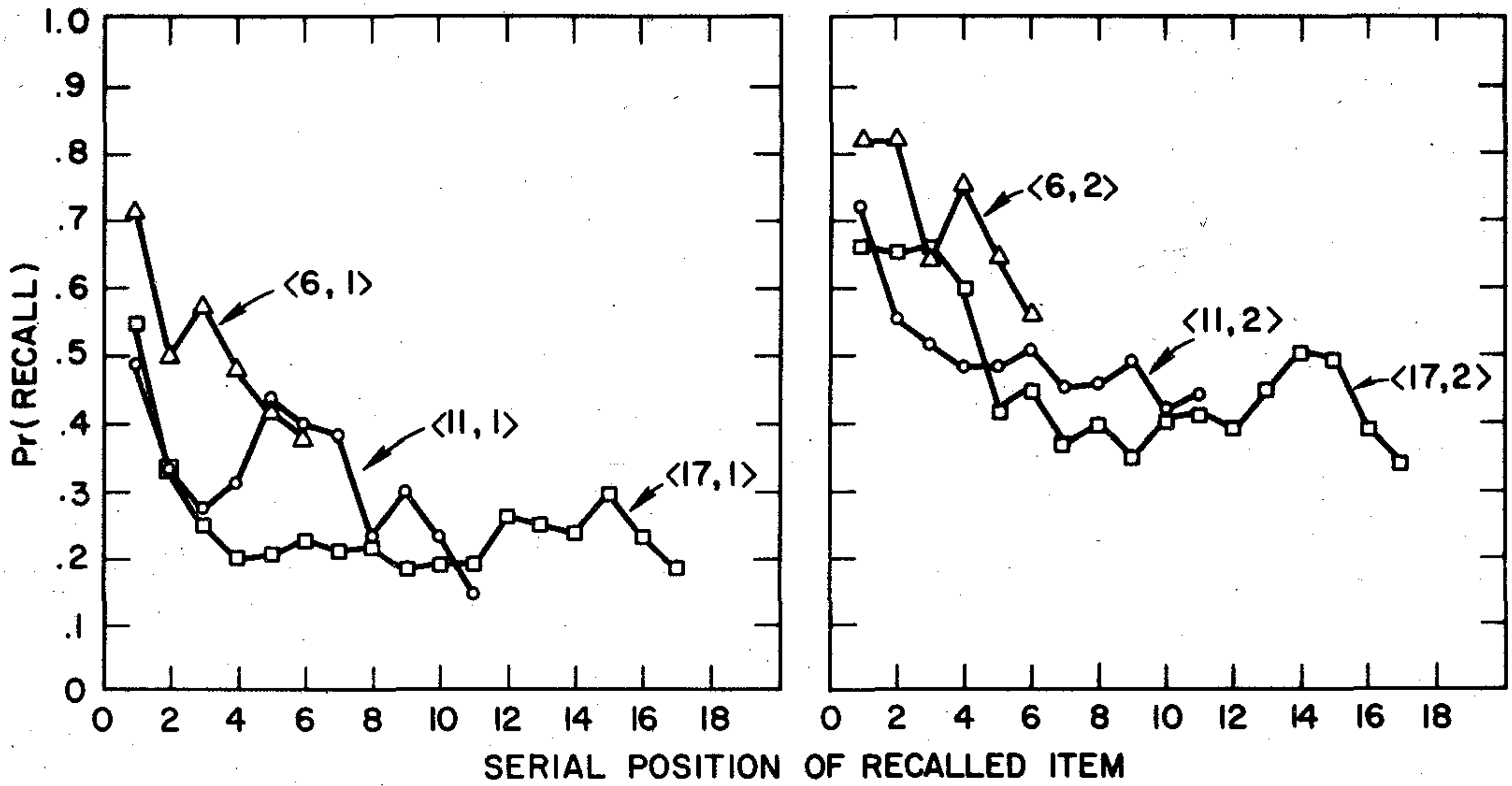


Fig. 13. Serial position curves for a free verbal recall task with interpolated arithmetic.

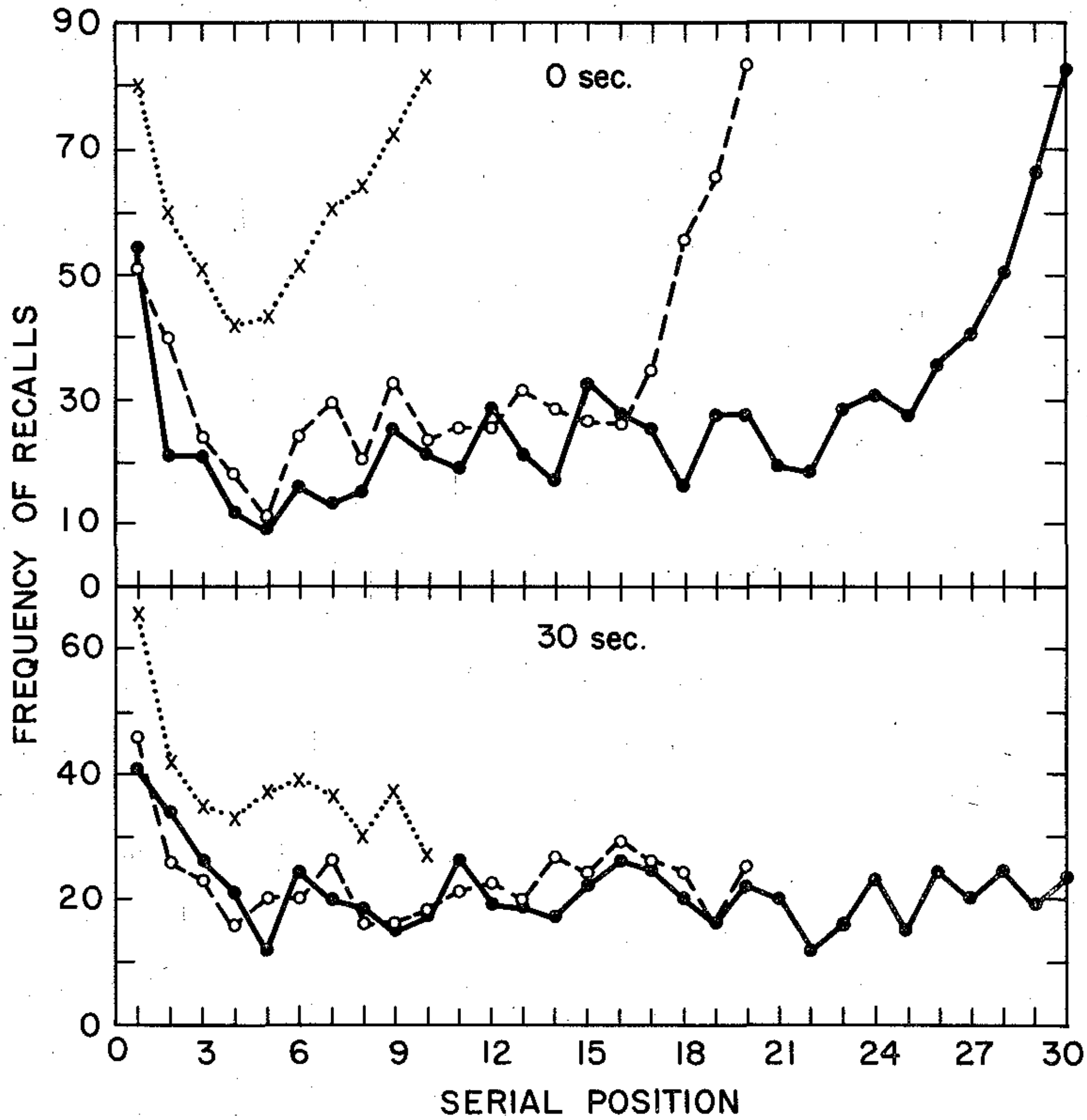


Fig. 14. Serial position curves for a free verbal recall task with interpolated arithmetic (after Postman and Phillips, 1965).

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.

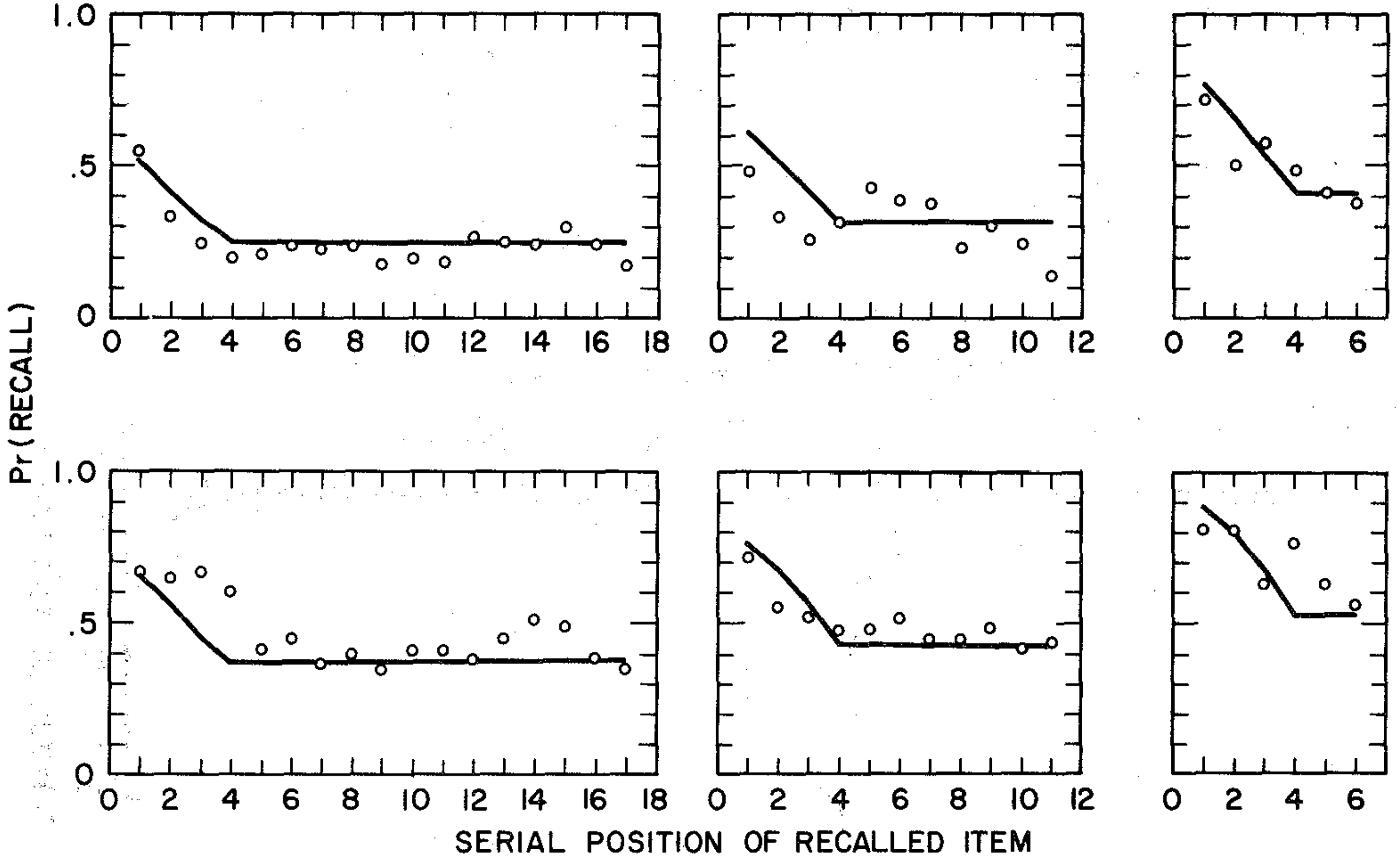


Fig. 15a. Observed and predicted values for the Shiffrin data (parameters: $r = 4$, $\gamma = .05$, $R = 30$, $\theta = 1$, $\delta \to 0$).

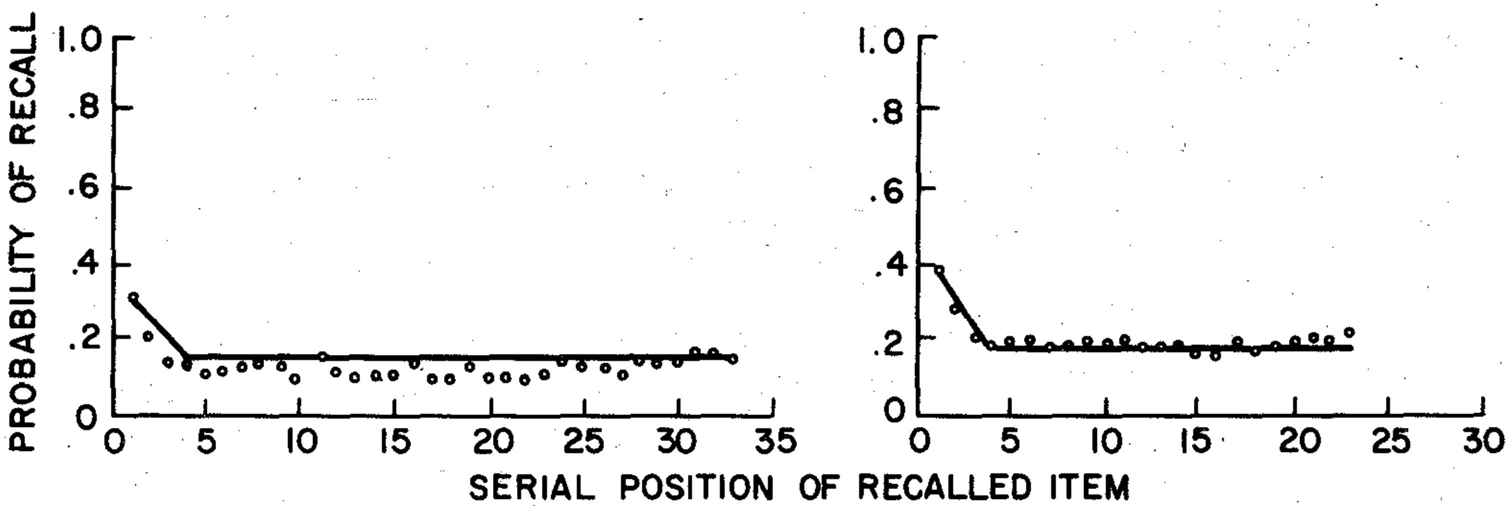


Fig. 15b. Observed and predicted values for the Murdock data (parameters: $r = 4$, $\gamma = .05$, $R = 30$, $\theta = 1$, $\delta \to 0$).

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 IS 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.

MULTI-PROCESS MODELS FOR MEMORY WITH APPLICATIONS TO A CONTINUOUS PRESENTATION TASK

R. C. Atkinson, J. W. Brelsford, and R. M. Shiffrin

Stanford University

TECHNICAL REPORT NO. 96

April 13, 1966

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Multi-Process Models for Memory with Applications to a Continuous Presentation Task[1]

R. C. Atkinson, J. W. Brelsford, and R. M. Shiffrin

Stanford University

Abstract

A multi-process model for memory and learning is applied to the results of two complementary experiments. In Experiment I the subject was required to keep track of the randomly changing responses associated with a fixed set of stimuli. The task involved a lengthy and continuous sequence of trials, each trial consisting of a test on one of the stimuli followed by study on that same stimulus paired with a new response. The size of the stimulus set, s, took on the values 4, 6, and 8. Experiment II differed from Experiment I in that a large number of stimuli were used even though in any experimental condition the subject was required to remember only 4, 6, or 8 stimuli at one time. In both experiments the basic dependent variable was the probability of a correct response as a function of the number of intervening trials between study and test on a given stimulus-response pair (called the "lag"). The lag curves were all near 1.0 at lag 0 and monotonically decreased as the lag increased; the lag curves for the three conditions (s = 4, 6, and 8) decreased at different rates in Experiment I, whereas in Experiment II these curves were identical. Using four estimated parameters the model generated accurate predictions for the various response measures collected.

[1]This research was supported by the National Aeronautics and Space Administration, Grant No. NGR-05-020-036.

A quantitative model for human memory and learning has been proposed by Atkinson and Shiffrin (1965). Specific versions of the general model have been used to predict serial position curves obtained from free-verbal recall and paired-associate experiments. The variables which have been successfully handled include list length, presentation rate, and in a study by Phillips, Shiffrin, and Atkinson (1966), confidence ratings. These previous studies were all conducted with a discrete-trial procedure, i.e., the presentation of an entire list of items was followed by a single test. In the present study it was desired to test the model in a situation involving a continuous succession of study and test items. Additionally, the present study involved the manipulation of certain experimental variables that have logical relationships to model parameters. The specific experimental variable manipulated was the size of the stimulus set being remembered by a subject.

The task employed in the experiments to be described here involves a modification of the typical paired-associate procedure which makes it possible to study the memory process under conditions that are quite uniform and stable throughout the course of an experiment. This is the case because the task is continuous and each subject is run for 10 to 12 daily sessions.[2] In essence the task involves having the subject keep track of the randomly changing response members of s different stimuli. Each trial of the experiment is divided into a test period and a study period. During the test

[2]The task is similar to those used by Yntema and Mueser (1962), Brelsford, Keller, Shiffrin, and Atkinson (1966), and Katz (1966).

phase a stimulus is randomly selected from among the set of s stimuli and the subject tries to recall the response <u>last</u> associated with that stimulus. Following the test, the study phase of the trial occurs. During this phase, the stimulus used in the test phase of the trial is re-paired with a new response for study. Thus every trial is composed of a test and study period on the same stimulus. Following each trial a new stimulus is chosen randomly from the set of s stimuli and the next trial begins. The instructions to the subject require that on a test he is to give the response that was paired with the stimulus the last time it was presented for study.

The number of trials intervening between study and test on a given stimulus-response pair will be referred to as the "lag" for that item. Thus, if the test occurs immediately following the study period the lag is zero. If one trial intervenes (involving test and study on another stimulus), then the lag is 1; and so on. It should be clear that in this task the number of stimulus-response pairs that the subject is trying to remember at any given time is fixed throughout an experimental session. Each time a stimulus is tested it is immediately re-paired with a new response, keeping the size of the to-be-remembered stimulus set always equal to s. Of course, in order to start an experimental session, an initial series of trials must be given with the test phase omitted. The stimuli presented during these study trials are the ones used throughout the rest of the experimental session. In the present experiments there were three experimental conditions in which the size of the stimulus set, s, was either 4, 6, or 8. For each daily session, a subject was randomly assigned to one of these three conditions. The principal dependent variable is the probability of a correct response as a function of lag.

Model

The model assumes three memory states: a very short-lived memory system called the sensory buffer; a temporary memory state called the memory (or rehearsal) buffer; and a long-term storage state called LTS. In the discussion of the model which follows, reference is frequently made to the term "stimulus-response item." Items are postulated to enter and leave the two buffers at various times. At the outset, the question arises, what is an item? In terms of the present model an item will be defined as that amount of information that allows one to make a correct recall when a stimulus is presented for a test. The specification of the exact form of this information (i.e., whether it be acoustic rehearsal, visual imagery, or some type of mnemonic) is not within the scope of the present paper. Nevertheless, in view of the work of Conrad (1964), Wickelgren (1965), and others on auditory confusions in short-term memory, we would be satisfied with the view that items in the memory buffer are acoustic mnemonics and are kept there via rehearsal, at least for experiments of a verbal character.

The Sensory Buffer

It is assumed that all external stimulation coming into the system enters the sensory buffer, resides there for a short time (perhaps on the order of a few seconds), decays and is lost.[3] In the context of the present experiment it will be assumed that every item enters the sensory buffer. Furthermore, it will be assumed that a test follows the preceding study period closely enough in time so that an item will always be recalled

---

[3]We imagine that the form of the decay is roughly representable by the results from the Peterson and Peterson (1959) experiment on the decay of a consonant trigram in the absence of rehearsal.

correctly if it is tested immediately following its entry into the buffer. Therefore, since every item enters the sensory buffer, the probability of a correct recall at lag 0 will be unity. For lags greater than zero, items will have decayed, and the sensory buffer will have no further significance. For this reason, in the remainder of this paper, the term buffer when used by itself will refer to the memory buffer.

<u>The Memory Buffer</u>

The memory buffer is postulated to have a limited and constant capacity for homogeneous items. It may be viewed as a state containing those items which have been selected from the sensory buffer for repeated rehearsal. Once the memory buffer is filled, each new item which enters causes one of the items currently in the buffer to be lost. It is assumed that the series of study items at the start of each experimental session fills the buffer and that the buffer stays filled thereafter. The size of the buffer, r (defined as the number of items which can be held simultaneously), depends upon the nature of the items and thus must be estimated for each experiment. It is assumed that a correct response is given with probability one if an item is in the buffer at the time it is tested.

We have already said that every item enters the sensory buffer and that items are selected from there to be entered into the memory buffer. Assume that at the time items enter the sensory buffer they are examined. These items fall into one of two categories. They may be items which are already in the buffer, i.e., their stimulus member may already be in the buffer. Alternatively, their stimulus member may not currently be in the buffer. The former kind of item shall be referred to as an O-item ("old" item), and the latter kind as an N-item ("new" item).[4] When an O-item is presented for study, it enters the memory buffer with probability one; the corresponding item, which was previously in the buffer, is discarded. Thus an O-item may be said to replace itself in the buffer. When an N-item is presented for study it enters the buffer with probability α. The value of the parameter α may be related in some manner to the particular scheme that a subject is using to rehearse the items currently in the buffer. When an N-item enters (with probability α) some item currently in the buffer is lost. This loss is called the "knockout process" and will be described below. With probability $(1-\alpha)$ an N-item fails to enter the buffer. In this case the buffer remains unchanged, the item in question decays from the sensory buffer, and is permanently lost from memory. For reference, the memory system is diagrammed in Fig. 1.

The memory buffer is arranged as a push-down list. The newest item that enters the buffer is placed in slot r, and the item that has remained in the buffer the longest is in slot 1. If an O-item is presented it enters slot r and the corresponding item is lost (in effect, the stimulus moves from its current slot to slot r and the response is changed). Then the other items move down one slot if necessary, retaining their former order. When an N-item is presented for study and enters the buffer (with probability α) it is placed in the r^{th} slot. The item to be knocked out is chosen according to the following scheme: with probability κ_j the item currently in slot j is the particular item that is discarded, where

[4]The reader should keep in mind that O-items and N-items are theoretical constructs and do not refer to observable experimental events.

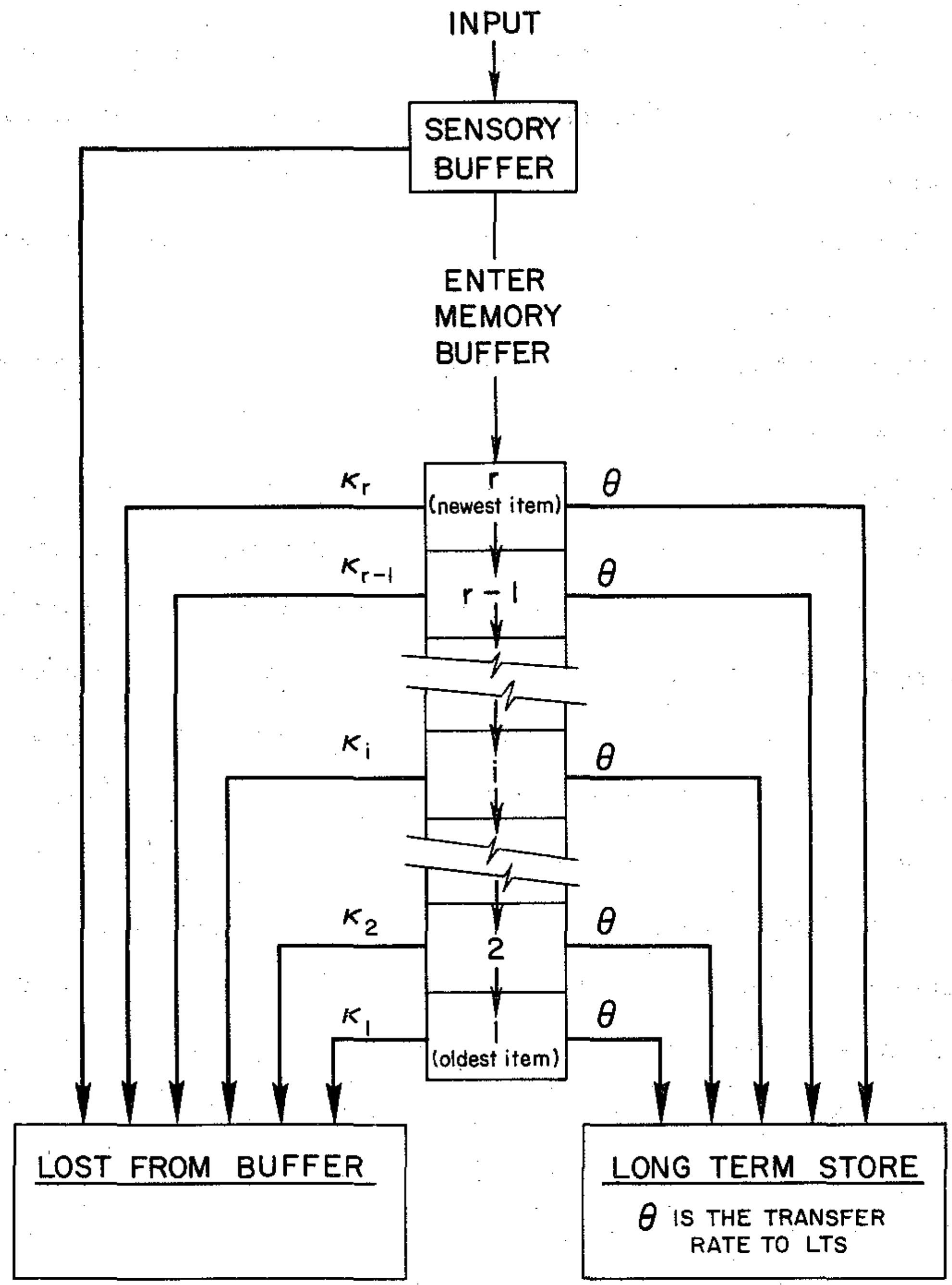


Figure 1. A flow chart characterizing inputs to the memory system.

$\kappa_1 + \kappa_2 + \cdots + \kappa_r = 1$. When the j^{th} item is discarded each item above the j^{th} slot moves down one, and the new item enters the r^{th} slot. Various schemes can be used to develop the κ_j's. The simplest is to let $\kappa_j = \frac{1}{r}$, in which case the item to be knocked out is chosen independently of the buffer position. However, in some experiments it has been necessary to postulate more general schemes which require that the longer the item has been in the buffer the greater its probability of being knocked out (Atkinson and Shiffrin, 1965).

Long-Term Storage

LTS is viewed as a memory state in which information accumulates for each item.[5] It is assumed that information about an item may enter LTS only during the period that an item resides in the buffer. We postulate that the status of an item in the buffer is in no way affected by transfer of information to LTS. Whereas recall from the buffer was assumed to be perfect, recall from LTS is not necessarily perfect and usually will not be. At the time of a test on an item, a subject gives the correct response if the item is in the sensory or memory buffer, but if the item is not in either of these buffers the subject searches LTS. This LTS search is called the retrieval process. Two features of the LTS retrieval process must be specified. First it is assumed that the likelihood of retrieving the correct response for a given item improves as the amount of information stored concerning that item increases. Second, the retrieval of an item gets worse the longer the item has been stored in LTS. This may simply mean that there is

[5]The term "information" is not used here in a technical sense. We use the term to refer to codes, mnemonics, images or anything else the subject might store that would be retrievable at the time of test.

some sort of decay in information as a function of the length of time information has been stored in LTS.

We shall specifically assume in this paper that information is transferred to LTS at a constant rate θ during the entire period in which an item resides in the buffer; θ is the transfer rate per trial. Thus, if an item remains in the buffer for exactly j trials (i.e., the j^{th} study item following the presentation of a given item causes it to be knocked out of the buffer), then that item accumulated an amount of information equal to $j\theta$. Next assume that each trial following the trial on which an item is knocked out of the buffer causes the information stored in LTS for that item to decrease by a constant proportion τ. Thus, if an item were knocked out of the buffer at trial j, and i trials intervened between the original study and the test on that item, the amount of information stored in LTS at the time of test would be $j\theta\tau^{i-j}$. We now want to specify the probability of a correct retrieval of an item from LTS. If the amount of information stored at the moment of test for an item is zero, then the probability of a correct retrieval should be at the guessing level. As the amount of information increases, the probability of a correct retrieval should increase toward unity. We define ρ_{ij} as the probability of a correct response from LTS of an item that had a lag of i trials between its study and test, and that resided in the buffer for exactly j trials. Considering the above specifications on the retrieval process,

$$\rho_{ij} = 1 - (1-g)\exp[-j\theta(\tau^{i-j})] \qquad (1)$$

where g is the guessing probability and in the present experiment is 1/26 since there were 26 response alternatives.

Lest the use of an exponential function seem entirely arbitrary, it should be noted that this function bears a close relation to the familiar linear model of learning theory. If we ignore for the moment the decay feature, then $\rho_{ij} = 1 - (1-g)\exp(-j\theta)$. It is easily seen that this is the linear model expression for the probability of a correct response after j reinforcements with parameter $e^{-\theta}$. Thus, the retrieval function ρ_{ij} can be viewed as a linear model with time in the buffer as the independent variable. To be sure, the decay process complicates matters, but the reason for choosing the exponential function becomes somewhat less arbitrary. A decay process is needed so that the probability of a correct retrieval from LTS will approach a chance level as the lag tends toward infinity.

<u>Derivation of Lag Curves</u>[6]

The basic dependent variable in the present experiment is the probability of a correct recall at the time of a test, given lag i. In order to derive this probability we need to know the length of time that an item resides in the memory buffer. Therefore, define

> β_j = probability that an item (i.e., a specific stimulus-response pair) resides in the buffer for exactly j trials, given that it is tested at a lag greater than j.

In the general case we must define another quantity in order to find β_j; namely

---

[6]The derivations are for the case where $r < s$. If $r \geq s$ a given item will always remain in the buffer until it is tested and consequently performance will be perfect at all lags.

β'_{ij} = probability that an item (i.e., a specific stimulus-response pair) currently in slot i resides in the buffer for exactly j more trials, given that it is tested at some point following this period.

Remember that r represents the number of slots in the buffer, and κ_j is the probability that the item in the j^{th} slot will be knocked out when an N-item enters. The probability of an N-item (one not currently in the buffer) being presented on a trial is $(s-r)/s$, where s is the number of stimuli used in a given experimental condition; likewise, the probability of an O-item being presented is r/s. We shall define β'_{ij} recursively. Note that an item's buffer position on a trial is either the same, or one less on the succeeding trial (if it is not knocked out of the buffer). We therefore obtain the following difference equations:

$$\beta'_{1,j} = \left(1 - \frac{s-r}{s-1}\alpha\kappa_1\right)\beta'_{1,j-1}$$

$$\beta'_{i,j} = \left\{\frac{r-i}{s-1} + \frac{s-r}{s-1}\left[(1-\alpha) + \alpha(\kappa_{i+1} + \kappa_{i+2} + \cdots + \kappa_r)\right]\right\}\beta'_{i,j-1}$$
$$+ \left\{\frac{i-1}{s-1} + \frac{s-r}{s-1}\alpha(\kappa_1 + \kappa_2 + \cdots + \kappa_{i-1})\right\}\beta'_{i-1,j-1}$$

$$\beta'_{r,j} = \left\{\frac{s-r}{s-1}(1-\alpha)\right\}\beta'_{r,j-1} + \left\{\frac{r-1}{s-1} + \frac{s-r}{s-1}\alpha(1-\kappa_r)\right\}\beta'_{r-1,j-1} \quad (2)$$

The initial conditions are $\beta'_{i,1} = \frac{s-r}{s-1}\alpha\kappa_i$. Recall that when an N-item is presented it will enter the memory buffer with probability α. Also, note that the denominator in the terms denoting the probabilities of N-items and O-items is $(s-1)$ rather than s. This is the case because β'_{ij} is a probability conditionalized upon the fact that we have yet to present the item in question for test. Now we can write:

$$\beta_j = \begin{cases} \frac{s - r}{s} (1-\alpha) & , \text{ for } j = 0 \\ \left\{1 - \frac{s - r}{s} (1-\alpha)\right\} \beta'_{r,j} & , \text{ for } j > 0 , \end{cases} \tag{3}$$

where β_0 is the probability that the item in question does not enter the memory buffer in the first place. It should be clear that the above difference equations can be solved by successive substitution, but such a process is lengthy and cumbersome. In practice, numerical solutions are easily obtained using a high-speed computer.

The probability of a correct response to an item tested at lag i can now be written in terms of the β_j's. Let "C_i" represent the occurrence of a correct response to an item tested at lag i. Then

$$\Pr(C_i) = \left[1 - \sum_{k=0}^{i} \beta_k\right] + \left[\sum_{k=0}^{i} \beta_k \rho_{ik}\right] . \tag{4}$$

The first bracketed term is the probability that the item is in the buffer at the time of test. The second bracket contains a sum of probabilities, each term representing the probability of a correct retrieval from LTS of an item which remained in the buffer for exactly k trials and was then lost.

Experiment I

The first experiment was carried out to determine whether reasonable predictions could be made assuming that the parameters of the model (r, α, θ, and τ) are independent of the number of stimuli the subject is trying to remember. Three experimental conditions were run: $s = 4$, 6, and 8.

Method

Subjects. The subjects were 9 students from Stanford University who received \$2 per experimental session. Each subject participated in approximately 10 sessions.

Apparatus. The experiment was conducted in the Computer-Based Learning Laboratory at Stanford University. The control functions were performed by computer programs running in a modified PDP-1 computer manufactured by the Digital Equipment Corporation, and under control of a time-sharing system. The subject was seated at a cathode-ray-tube display terminal; there were six terminals each located in a separate 7×8 ft. sound-shielded room. Stimuli were displayed on the face of the cathode ray tube (CRT); responses were made on an electric typewriter keyboard located immediately below the lower edge of the CRT.

Stimuli and responses. The stimuli were two-digit numbers randomly selected for each subject and session from the set of all two-digit numbers between 00 and 99. Once a set of stimuli was selected for a given session, it was used throughout the session. Responses were letters of the alphabet, thus fixing the guessing probability of a correct response at 1/26.

Procedure. For each session the subject was assigned to one of the three experimental conditions (i.e., s was set at either 4, 6, or 8). An attempt was made to assign subjects to each condition once in consecutive three-session blocks. Every session began with a series of study trials: one study trial for each stimulus to be used in the session. On a study trial the word "study" appeared on the upper face of the CRT. Beneath the word "study" one of the stimuli appeared along with a randomly-slected letter from the alphabet. Subjects were instructed to try to remember the association between the stimulus-response pairs. Each of these initial study trials lasted for 3 sec. with a 3-sec. intertrial interval. As soon as there had been an initial study trial for each stimulus to be used in the session, the session proper began.

Each subsequent trial involved a fixed series of events. (1) The word <u>test</u> appeared on the upper face of the CRT. Beneath the word <u>test</u> a randomly selected member of the stimulus set appeared. Subjects were instructed that when the word <u>test</u> and a stimulus appeared on the CRT, they were to respond with the last response that had been associated with that stimulus, guessing if necessary. This test portion of a trial lasted for 3 sec. (2) The CRT was blacked out for 2 sec. (3) The word <u>study</u> appeared on the upper face of the CRT for 3 sec. Below the word <u>study</u> a stimulus-response pair appeared. The stimulus was the same one used in the preceding test portion of the trial. The response was randomly selected from the letters of the alphabet, with the stipulation that it be different from the immediately preceding response assigned to that stimulus. (4) There was a 3-sec. intertrial interval before the next trial. Thus a complete trial (test plus study) took 11 sec. A subject was run for 220 such trials during each experimental session.

<u>Results</u>

In order to examine the data for habituation or learning-to-learn effects, the overall probability of a correct response for each stimulus condition (s = 4, 6, and 8) was plotted in consecutive 25-trial blocks. It was found that after a brief rise at the start of each daily session, the curves appeared to level off at three distinct values. Due to this brief initial warm-up effect, subsequent analyses will not include data from the first 25 trials of each session. Furthermore, the first session for each subject will not be used.

Figure 2 presents the probability of a correct response as a function of lag for each of the three stimulus set sizes examined. It can be seen that the smaller the stimulus set size, the better the overall performance. It is important to note that the theory presented in the earlier part of this paper predicts such a difference on the following basis: the larger the size of the stimulus set, the more often an N-item will be presented; and the more often N-items are presented, the more often items in the buffer will be knocked out. Recall that only N-items can knock items from the buffer; O-items merely replace themselves.

It can be seen that performance is almost perfect for lag 0 in all three conditions. This might be expected because lag 0 means that the item was tested immediately following its study. The curves drop sharply at first and slowly thereafter, but have not yet reached the chance level at lag 17, the largest lag plotted. The chance level should be 1/26 since there were 26 response alternatives.

It is of interest to examine the type of errors occurring at various lags in the three experimental conditions. There are two categories of errors that are of special interest to us. The first category is composed of errors which occur when the immediately preceding correct response to a stimulus is given, instead of the present correct response. The proportions of errors of this type were calculated for each lag and each condition. The proportions were found to be quite stable over lags with mean values of .065, .068, and .073 for the 4, 6, and 8 stimulus conditions, respectively. If the previously correct response to an item is randomly generated on any

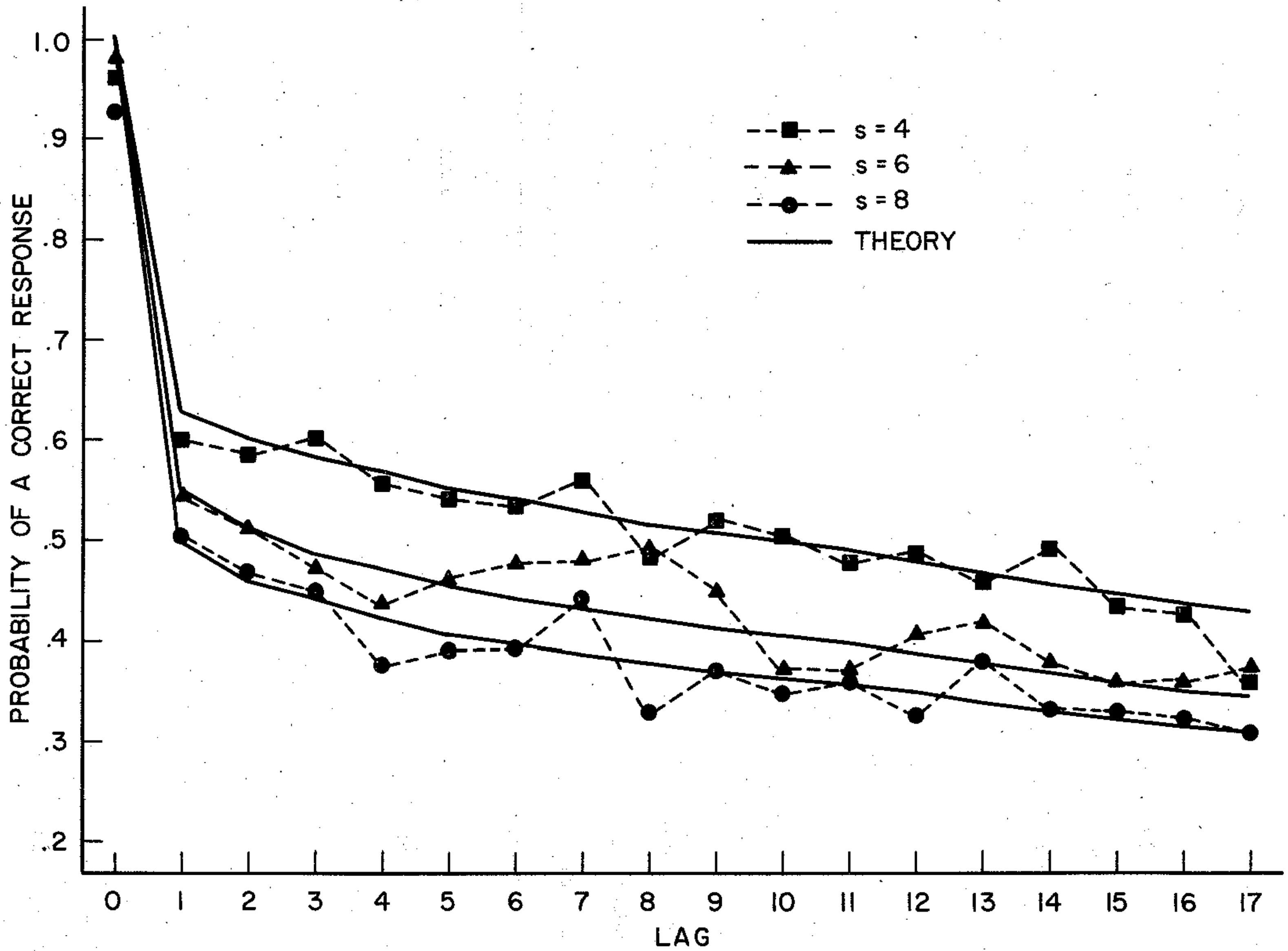


Figure 2. Observed and theoretical probabilities of a correct response as a function of lag (Experiment I).

given error, these values should not differ significantly from $1/25 = .04$. The mean proportion for this type of error was computed for each subject and each condition. In both the $s = 4$ and $s = 6$ conditions 7 of the 9 subjects had mean values above chance; in the $s = 8$ condition 8 of the 9 subjects were above chance. A second category of errors of interest to us is composed of those responses that are members of the current set of responses being remembered, but are not the correct response. The proportions of this type of error were calculated for each lag in each of the three experimental conditions. Again, the proportions were found to be quite stable over lags. The mean values were .23, .28, and .35 for the 4, 6, and 8 stimulus conditions, respectively; on the basis of chance these values would have to be bounded below .12, .20, and .28, respectively. No statistical tests were run, but again the values appear to be above those expected by chance. While a detailed examination of the implications of these conditional error results is not a purpose of this paper, it should be pointed out that this type of analysis may yield pertinent information regarding the nature of the LTS retrieval process.

There are two other lag curves that prove interesting. We shall call these the "all-same" and the "all-different" curves. In the all-same conditions, we compute the probability of a correct response as a function of the lag, when all of the intervening items between study and test involve the same stimulus. The model predicts that once the intervening stimulus enters the buffer, there will be no further chance of any other item being knocked out (i.e., once the intervening item enters the buffer, each succeeding presentation is an O-item). Hence, these curves should drop at a slower rate than the unconditional lag curves presented in Fig. 2. The all-same curves are plotted in Fig. 3. The points for lag 0 and lag 1 are, of course, the same as in the unconditional lag plots of Fig. 2. It can be seen that the curves indeed drop at a slower rate in this condition.

The all-different condition refers to the probability of a correct response as a function of lag, when the intervening items between study and test all involve different stimuli. For this reason the maximum lag which can be examined is one less than the size of the stimulus set. It should be clear that the all-different condition maximizes the expected number of intervening N-items at a given lag. This lag curve should therefore have a faster drop than the unconditional lag curves presented in Fig. 2. The data are shown in Fig. 4. While it is difficult to make a decision by inspection in this condition because the data are quite unstable, it does seem that the curves drop faster than the corresponding ones in Fig. 2. Note that here, also, the points for lag 0 and lag 1 are of necessity the same as in the previous conditions.

The results that have been presented to this point have been group data. It is of interest to see whether individual subjects perform in a fashion similar to the group curves. Table 1 presents the lag curves for the three experimental conditions for individual subjects. The lag curves have been collapsed into three-lag blocks to minimize variability. An examination of these individual curves indicates that all subjects, except for subject 8, appear to be performing in a manner very similar to the group data.

A final remark should be made regarding the number of observations taken at each point on these lag curves. Because of the random procedure used to select the stimuli from trial to trial, the number of observations going into successive points on the lag curves decrease geometrically. For the group

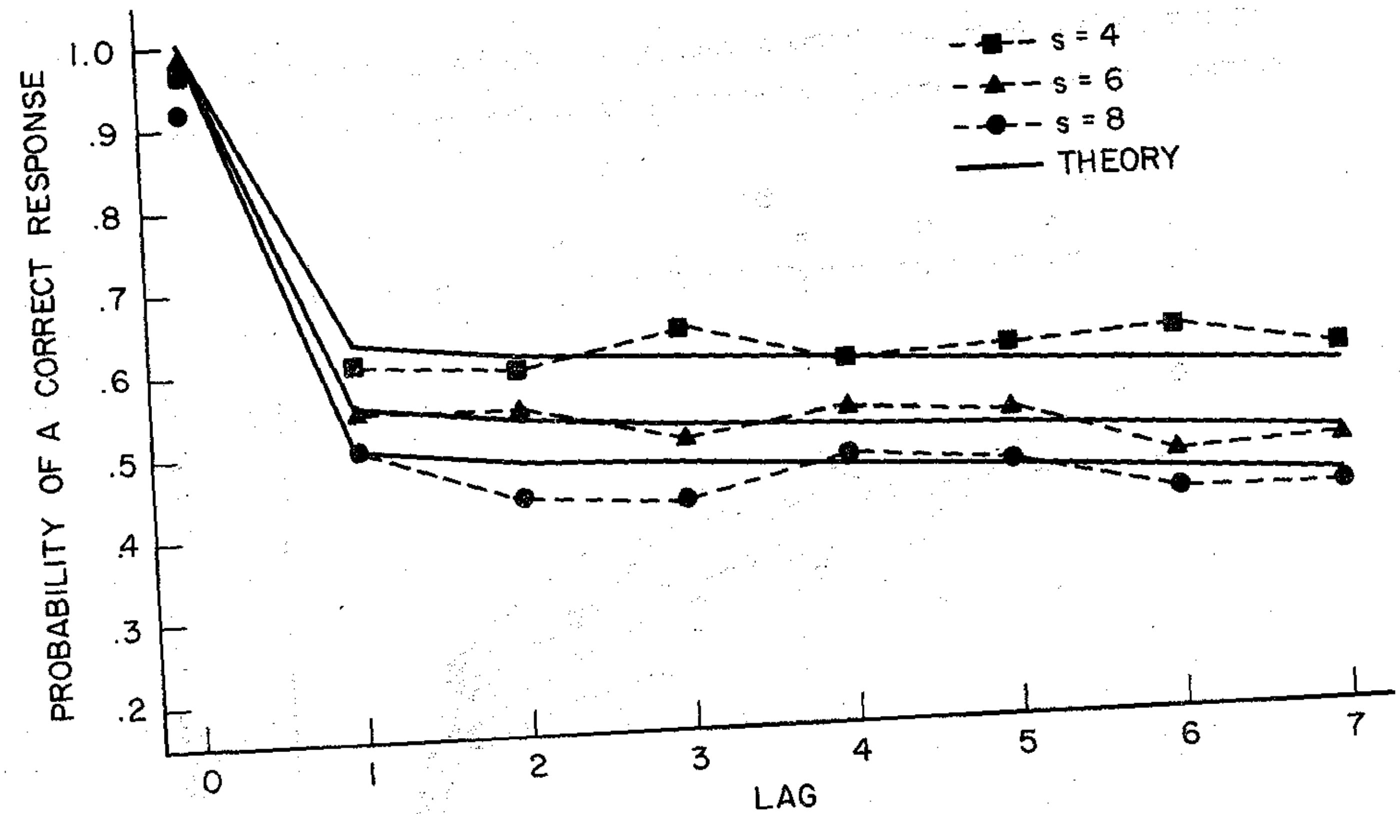


Figure 3. Observed and theoretical probabilities of a correct response as a function of lag for the "all-same" condition (Experiment I).

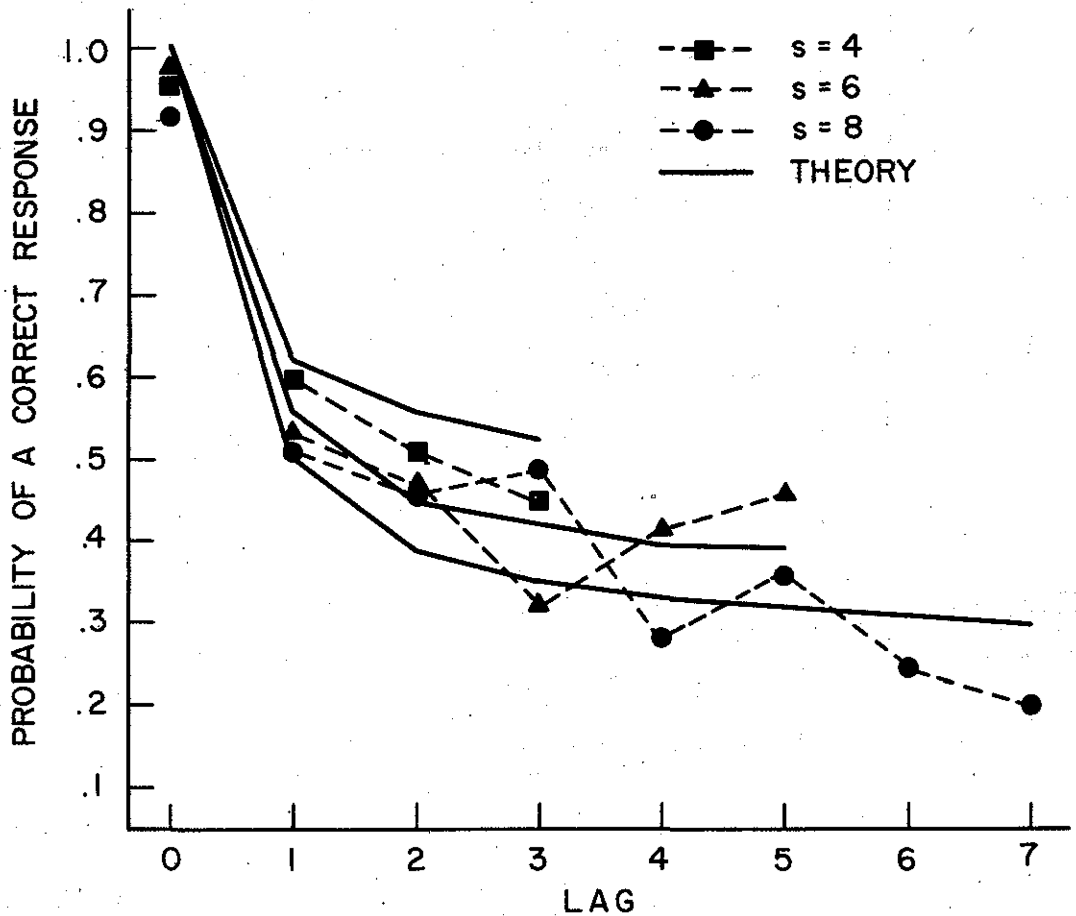


Figure 4. Observed and theoretical probabilities of a correct response as a function of lag for the "all-different" condition (Experiment I).

TABLE I

Observed and predicted probabilities of a correct response as a function of lag for individual subjects. The predicted values are in parentheses and are based on the parameter estimates that give the best fit for that subject; these estimates are presented in the bottom section of the table. The χ^2_I and χ^2_G are computed for each subject using the individually estimated parameters and the group parameters, respectively. Entries in the top section of the table should be read with a leading decimal point (Experiment I).

Experimental Condition	Lag	Subjects 1	2	3	4	5	6	7	8	9
s = 4	0-1-2	72 (70)	80 (85)	86 (86)	66 (67)	73 (75)	86 (84)	75 (73)	60 (60)	89 (84)
	3-4-5	51 (50)	74 (72)	76 (78)	52 (56)	62 (64)	65 (73)	50 (50)	61 (58)	70 (69)
	6-7-8	49 (48)	72 (69)	74 (76)	61 (52)	52 (58)	71 (68)	42 (44)	59 (56)	64 (63)
	9-10-11	50 (47)	62 (68)	72 (74)	59 (50)	49 (54)	67 (67)	41 (42)	60 (54)	66 (62)
	12-13-14	51 (46)	64 (66)	70 (73)	50 (49)	44 (49)	55 (66)	41 (40)	58 (52)	61 (61)
s = 6	0-1-2	63 (62)	82 (77)	78 (75)	56 (55)	56 (53)	72 (71)	61 (58)	49 (46)	68 (67)
	3-4-5	43 (42)	74 (66)	71 (69)	56 (50)	52 (46)	68 (63)	53 (49)	48 (42)	59 (58)
	6-7-8	41 (41)	72 (63)	65 (65)	45 (43)	41 (39)	63 (60)	35 (37)	40 (41)	53 (53)
	9-10-11	48 (40)	62 (61)	66 (62)	53 (42)	43 (37)	58 (59)	37 (35)	40 (40)	55 (52)
	12-13-14	51 (39)	64 (60)	60 (60)	54 (41)	45 (36)	60 (58)	38 (34)	44 (39)	54 (51)
s = 8	0-1-2	59 (57)	69 (73)	73 (70)	47 (49)	43 (45)	63 (63)	50 (51)	35 (38)	58 (57)
	3-4-5	40 (40)	67 (62)	63 (62)	46 (43)	47 (38)	60 (54)	45 (41)	56 (35)	52 (48)
	6-7-8	35 (36)	53 (60)	61 (60)	31 (39)	39 (31)	54 (52)	30 (32)	52 (32)	44 (45)
	9-10-11	30 (30)	56 (59)	58 (58)	41 (37)	45 (29)	53 (51)	31 (30)	44 (31)	45 (42)
	12-13-14	29 (29)	57 (58)	51 (57)	38 (36)	36 (28)	51 (50)	29 (29)	37 (30)	49 (41)
Parameter Estimates										
r		2	3	3	2	2	2	2	2	2
α		.56	.66	.59	.50	.50	.32	.65	.38	.31
θ		.70	1.22	2.09	.39	1.12	1.84	.62	1.37	.82
τ		.84	.95	.93	.89	.80	.82	.78	.99	.94
χ^2_I		23.6	29.2	24.8	31.2	38.2	26.4	31.2	67.2	13.6
χ^2_G		29.2	48.3	29.4	33.6	47.6	32.6	34.7	89.3	15.2

data there are over 1000 observations at lag 0 and slightly more than 100 at lag 17 for each of the three experimental conditions. Of course, the exact form of the distribution of data points varies as a function of the experimental condition, with more short lags occurring in the $s = 4$ condition and more long lags occurring in the $s = 8$ condition.

Model Predictions

In order to estimate parameters and evaluate the goodness-of-fit of the theory to the data, we define the following χ^2 function:

$$\chi^2 = \sum_i \left\{\frac{1}{N_i Pr(C_i)} + \frac{1}{N_i - N_i Pr(C_i)}\right\} \left\{N_i Pr(C_i) - O_i\right\}^2 \qquad (5)$$

where the sum is taken over all data points i which are being evaluated. The observed number of correct responses for the i^{th} point is denoted by O_i; N_i is the total number of responses for the i^{th} point; and $Pr(C_i)$ is the theoretical probability of a correct response which depends on r, α, θ, and τ. Thus $N_i Pr(C_i)$, the predicted number of correct responses for the i^{th} point, should be close to O_i if the theory is accurate.

We first analyze the lag curves displayed in Fig. 2. The set of parameter values r, α, θ, and τ that minimizes the above χ^2 function over the $3 \times 17 = 51$ data points in Fig. 2 will be taken to be the best fit of the model.[7] In order to minimize χ^2 we resorted to a numerical routine using a computer. The routine involved selecting tentative values for r, α, θ, and τ, computing the $Pr(C_i)$'s and the related χ^2, repeating the procedure with another set of parameter values, and continuing thusly until

---

[7] The lag 0 point in this and subsequent analyses is not included in the χ^2 since its predicted probability value is one.

the space of possible parameter values has been systematically explored. The parameter values yielding the smallest χ^2 are then used as the estimates. When enough points in the parameter space are scanned, the method yields a close approximation to the true minimum.[8]

The predictions for $\Pr(C_i)$ could be derived using Eqs. 3, but it was decided to set the $\kappa_i = 1/r$, in which case the equations simplify greatly. In a study by Phillips, Shiffrin and Atkinson (1966) it was found that the assumption $\kappa_i = 1/r$ was not tenable; in that experiment, however, there were strong reasons for expecting that the subject would tend to eliminate the oldest items from the buffer first. In the current experiment there is a continuous display of items and there seemed to be no compelling reason to believe that the subject would not discard items from the buffer in a random fashion. For this reason κ_i was set equal to $1/r$ for every buffer position. Under this assumption it is immaterial what position an item occupies in the buffer. Thus $\beta'_{i,k} = \beta'_{j,k}$ for all i and j; hence (as can be easily verified) every line of Eq. 3 can be rewritten as follows:

$$\beta'_{i,k} = \left\{1 - \frac{s-r}{s-1}\,\alpha\,\frac{1}{r}\right\} \beta'_{i,k-1} \;. \tag{6}$$

Let the term in brackets be denoted by $1 - X$. Then we have $\beta_0 = (1-\alpha)(s-r)/s$ which is the probability that the item will not enter the buffer, and

$$\beta_k = (1-\beta_0)X(1-X)^{k-1} \;. \tag{7}$$

[8]For a discussion of the minimum χ^2 method see Holland (1965) or Atkinson, Bower, and Crothers (1965).

It is easy to verify this equation if we note that X is the probability that an intervening item will enter the buffer and knock out the item of interest. For the item of interest to be knocked out of the buffer by exactly the k^{th} following item, it is necessary that the following conditions hold: (1) the item must enter the buffer in the first place; (2) the next $k - 1$ intervening items must not knock it out; (3) the k^{th} item must knock out the item of interest. These considerations lead directly to Eq. 7.

Given β_k we can calculate the predicted lag curves for each set of parameters considered using Eq. 4. The χ^2 procedure described earlier was applied simultaneously to all three curves displayed in Fig. 2 and the values of the parameters that gave the minimum χ^2 were as follows: $r = 2$, $\alpha = .39$, $\theta = .40$, and $\tau = .93$. The theoretical lag curves generated by these parameters are shown in Fig. 2. It can be seen that the observed data and the predictions from the model are in close agreement; the minimum χ^2 value is 43.67 based on 47 degrees of freedom ($17 \times 3 = 51$ data points minus four estimated parameters).[9] It should be emphasized that the three curves are fit simultaneously using the same parameter values, and the differences between the curves depend only on the value of s

[9] In this and all subsequent minimizations reported in this paper, r was permitted to take on only integer values. Better fits can be obtained by removing this constraint (e.g., in this case the minimum χ^2 is 40.36 when $r = 2.1$, $\alpha = .37$, $\theta = .44$, and $\tau = .91$), but we prefer to evaluate the model assuming r is fixed for all subjects.

used. The predicted probabilities of a correct response weighted and summed over all lag positions are .562, .469, and .426 for s equal to 4, 6, and 8, respectively; the observed values are .548, .472, and .421.

The estimated value of α indicates that only 39 percent of the N-items presented actually enter the buffer (remember that O-items always enter the buffer). At first glance this percentage may seem low, but a good deal of mental effort may be involved in keeping an item in the buffer via rehearsal, and the subject might be reluctant to discard an item which he has been rehearsing before it is tested. Actually, if there were no long-term storage, the subject's overall probability of a correct response would be independent of α. Thus it might be expected that α would be higher the greater the effectiveness of long-term storage in an experiment. The estimate of θ found does not have a readily discernable interpretation, but the value of $\tau = .93$ indicates that the decay in LTS is extremely slow. It is not necessary to assume that any actual decay occurs--several alternative processes are possible. For example, the subject could search LTS backwards along a temporal dimension, sometimes stopping the search before the information relevant to the tested item is found.[10]

Next we examine the lag curves for the all-same condition. As indicated earlier these curves should be less steep than the unconditional lag curves. This would be expected because, in the all-same conditions (where the intervening trials all involve the same stimulus), once an intervening item enters the buffer, every succeeding item will be an O-item and will

---

[10]The high value of τ might suggest that a reasonable fit could be obtained setting $\tau = 1$. When this was done, however, the minimum χ^2 was 62.74 with parameter estimates $r = 2$, $\alpha = .42$, $\theta = .24$.

replace itself. Indeed, if $\alpha = 1$ and there is no LTS storage, the all-same lag curves would be level from lag 1 onward. The model applies directly to this case. Define β_j^* as the probability that an item resides in the buffer for exactly j trials and is then knocked out, given that all the intervening trials involve the same stimulus. Then

$$\beta_j^* = \begin{cases} \frac{s - r}{s}(1 - \alpha) & , \text{ for } j = 0 \\ (1 - \beta_0^*)\left[\frac{s - r}{s - 1}(1 - \alpha)^{j-1}\frac{\alpha}{r}\right] & , \text{ for } j > 0 . \end{cases} \tag{8}$$

It can easily be seen that the β_j^* have the above form. For an item to be knocked out by the j^{th} succeeding item it is necessary that the following holds: (1) the item enters the buffer initially; (2) the following items must be new items and must not enter the buffer for $j - 1$ trials (clearly, if the first intermediate item is an N-item, then in the all-same condition each succeeding item has to be an N-item until one of the items enters the buffer); (3) the j^{th} following item enters the buffer and knocks out the item of interest. The predicted lag curves for the all-same condition may be calculated substituting β_j^* for β_j in Eq. 4. The parameters found in fitting the unconditional lag curves in Fig. 2 were used to generate predictions for the all-same condition, and the predicted lag curves are presented in Fig. 3. The fit is excellent as indicated by a χ^2 of 26.8 based on 21 degrees of freedom.

Next we turn to the lag curves for the all-different condition. Considerations similar to those presented in the discussion of the all-same data lead to the prediction that the all-different lag curves will be steeper than the unconditional lag curves. Unfortunately there were relatively few observations in this condition and the data is fairly unstable. Nevertheless we shall apply the model to these data in large part because the mathematical techniques involved are rather interesting. Define

β_j^{**} = the probability that an item will reside in the buffer for exactly j trials, given that the intervening stimuli are all different.

It can be quickly demonstrated that an attempt to develop the β_j^{**} equations directly does not succeed, primarily because the probability of presenting an N-item changes from trial to trial. The solution is to view the process an an <u>inhomogeneous</u> Markov chain with $r + 1$ states. The first state will correspond to the event that the item of interest is currently not in the buffer. The other r states will denote the conditions in which the item of interest is in the buffer and m ($m = 0$ to $r - 1$) of the remaining places in the buffer are filled with items that have already been presented in the sequence of all-different items. For the sake of simplicity we shall develop the process for the case where $r = 2$ since the all-different curves will be fit using the parameters estimated from the unconditional lag curves. It is easy to see how to generalize the method to larger values of r.

To start with, define $\overline{B}$ as the state in which the item of interest is not currently in the buffer. Define BA as the state where the item of interest is in the buffer and the other slot of the buffer is occupied by an item which has <u>already</u> been presented in the sequence of all-different items. Define $B\overline{A}$ as the state in which the item of interest is in the buffer and the other slot of the buffer is not occupied by an item which has already been presented in the sequence of all-different intervening items. Then the following matrix describes transitions from intervening trial k to intervening trial $k + 1$:

$$
\begin{array}{cc}
 & \text{Trial } k+1 \\
 & \begin{array}{ccc} \quad\bar{B}\quad & \quad BA\quad & \quad B\bar{A}\quad \end{array} \\
\begin{array}{c} \text{Trial} \\ k \end{array}
\begin{array}{c} \bar{B} \\ BA \\ B\bar{A} \end{array}
&
\begin{bmatrix}
1 & 0 & 0 \\
\frac{\alpha}{2} & 1 - \frac{\alpha}{2} & 0 \\
\frac{\alpha}{2}\left[\frac{s-(k+2)}{s-(k+1)}\right] & 1 - \left[1 - \frac{\alpha}{2}\right]\left[\frac{s-(k+2)}{s-(k+1)}\right] & [1-\alpha]\left[\frac{s-(k+2)}{s-(k+1)}\right]
\end{bmatrix}
\end{array} . \tag{9}
$$

The starting vector at $k = 0$ is as follows:

$$
\begin{array}{ccc} \bar{B} & BA & B\bar{A} \\ \end{array}
$$

$$
\left[\frac{s-2}{s}(1-\alpha) \quad 0 \quad 1 - \frac{s-2}{s}(1-\alpha)\right] .
$$

Let the probability of being in state $\bar{B}$ on intervening trial k be $p_k(\bar{B})$. Then

$$
\beta_j^{**} = \begin{cases} p_j(\bar{B}) - p_{j-1}(\bar{B}) & , \text{ for } j > 0 \\ p_0(\bar{B}) & , \text{ for } j = 0 \end{cases} \tag{10}
$$

where $p_0(\bar{B}) = (1-\alpha)(s-2)/s$.

In order to determine $p_k(\bar{B})$ we used a computer to multiply the starting vector by the transition matrix the appropriate number of times. This was done using the parameter values from the fit of the unconditional lag curves. The β_j^{**} were then computed and the lag curves generated as before. The predicted curves are shown in Fig. 4. Considering the lack of stability in the data, the fit is not too bad. The χ^2 was 64.8 based on 15 degrees of freedom.

The model is not explicit regarding the likelihood of the previously correct response being incorrectly emitted at the time of test. Nevertheless, the interpretation of the LTS retrieval process which postulates a temporal search of stored items suggests that the previously correct information may be accidentally found during retrieval, thus heightening the probability that the prior response will be given. A slight anomaly here is that in the data this probability appears to be independent of lag which might not be predicted from the preceding argument.

Similarly, the model does not make predictions concerning the probability that a response in the current response set will be given as an error. However, there will be overlap between the current response set and the items stored in the buffer; it does not seem unreasonable that subjects who cannot find the correct response in their search of the buffer and LTS might tend to guess by favoring a response currently in the buffer. The data indicate that this tendency is above the chance level. This suggests that our assumption of a guessing level of 1/26 could be slightly inaccurate. In future work it may prove necessary to postulate a changing guessing level which declines toward the reciprocal of the number of responses only as the lag tends toward infinity.[11]

We now consider the implicit assumption involved in fitting curves for group data--namely that the subjects are homogeneous. A direct approach would be to fit the model to each subject's data separately. This was done

---

[11]Our use of the term "guessing level" in this context is itself misleading because it seems clear that the subject is using stored information concerning recent responses while "guessing."

under the restriction that three adjacent lags be lumped into a single point (there were not enough observations to guarantee stable lag curves from individual subjects without lumping adjacent points). Thus the model was fit independently to the data from each subject in the same manner that the group data was fit (naturally, for each set of parameter values considered, the predicted lag curves were lumped in the same manner as the observed data). The predictions of the model yielding minimum χ^2's for each subject are presented in Table 1 along with the observed data. Also given are the minimum χ^2 values and the parameter estimates for each subject. It is somewhat difficult at this point to decide the question of homogeneity of the subjects. In order to do so, the lag curves for each subject were predicted using a single set of parameters; namely those values estimated from the group data. When this was done the sum of the χ^2 values over subjects was 359.9 with 131 degrees of freedom. The sum of the χ^2 when each subject was fit with a separate best set of parameters was 285.4 with 99 degrees of freedom. The ratio of the two χ^2's, each divided by its respective degrees of freedom, is 1.05. This suggests that the assumption of homogeneity of subjects is not unreasonable.

Experiment II

Experiment II was identical to Experiment I in all respects except the following. In Experiment I the set of s stimuli was the same throughout an experimental session, with only the associated responses being changed on each trial, while in Experiment II all 100 stimuli were available for use in each session. In fact, every stimulus was effectively an N-item since the stimulus for each study trial was selected randomly from the set of all 100 stimuli under the restriction that no stimulus could be used if it had been tested or studied in the previous fifty trials. There were still three experimental conditions with s equal to 4, 6, or 8 denoting the number of items that the subject was required to try to remember at any point in time. Thus a session began with either 4, 6, or 8 study trials on different randomly selected stimuli each of which was paired with a randomly selected response (from the 26 letters). On each trial a stimulus in the current to-be-remembered set was presented for test. After the subject made his response he was instructed to forget the item he had just been tested on, since he would not be tested on it again. Following the test a new stimulus was selected (one that had not appeared for at least fifty trials) and randomly paired with a response for the subject to study. This procedure is quite different from Experiment I where the study stimulus was always the one just tested.

Denote an item presented for study on a trial as an O-item (old item) if the item just tested was at the moment of test in the buffer. Denote an item presented for study as an N-item (new item) if the item just tested was not in the buffer. This terminology conforms precisely to that used to describe Experiment I. If an O-item is presented there will be at least one spot in the buffer occupied by a useless item (the one just tested). If an N-item is presented, the buffer will be filled with information of the same value as that before the test. If we assume that an N-item has probability α of entering the buffer, and that an O-item will always enter the buffer and knock out the item just made useless, then the theory used to analyze Experiment I will apply here with no change whatsoever. In this case we again expect that the lag curves for $s = 4$, 6, and 8 would be separated. In fact, given the same parameter values, exactly the same predicted curves would be expected in Experiment II as in Experiment I.

We may have some doubt, however, that the assumptions regarding N-items and O-items will still hold for Experiment II. In Experiment I the stimulus just tested was re-paired with a new response, virtually forcing the subject to replace the old response with a new one if the item was in the buffer. To put this another way, if an item is in the buffer when tested, only a minor change need be made in the buffer to enter the succeeding study item: a single response is replaced by another. In Experiment II, however, a greater change needs to be made in order to enter an O-item; both a stimulus and a response member have to be replaced. Thus an alternative hypothesis which could be entertained holds that every entering item (whether an N-item or an O-item) has the same probability α of entering the buffer, and will knock out any item currently in the buffer with equal likelihood. In this case there will be no predicted differences among the lag curves for the $s = 4$, 6, and 8 conditions.

Results

The observed lag curves for Experiment II are displayed in Fig. 5. The number of observations at each point range from 1069 for lag 0 in condition $s = 4$ to 145 for lag 17 in condition $s = 8$. It should be emphasized that except for the procedural changes described above and the fact that a new sample of subjects was used in Experiment II, the experimental conditions and operations were identical in the two experiments. The important point of interest in this data is that lag curves for the three conditions appear

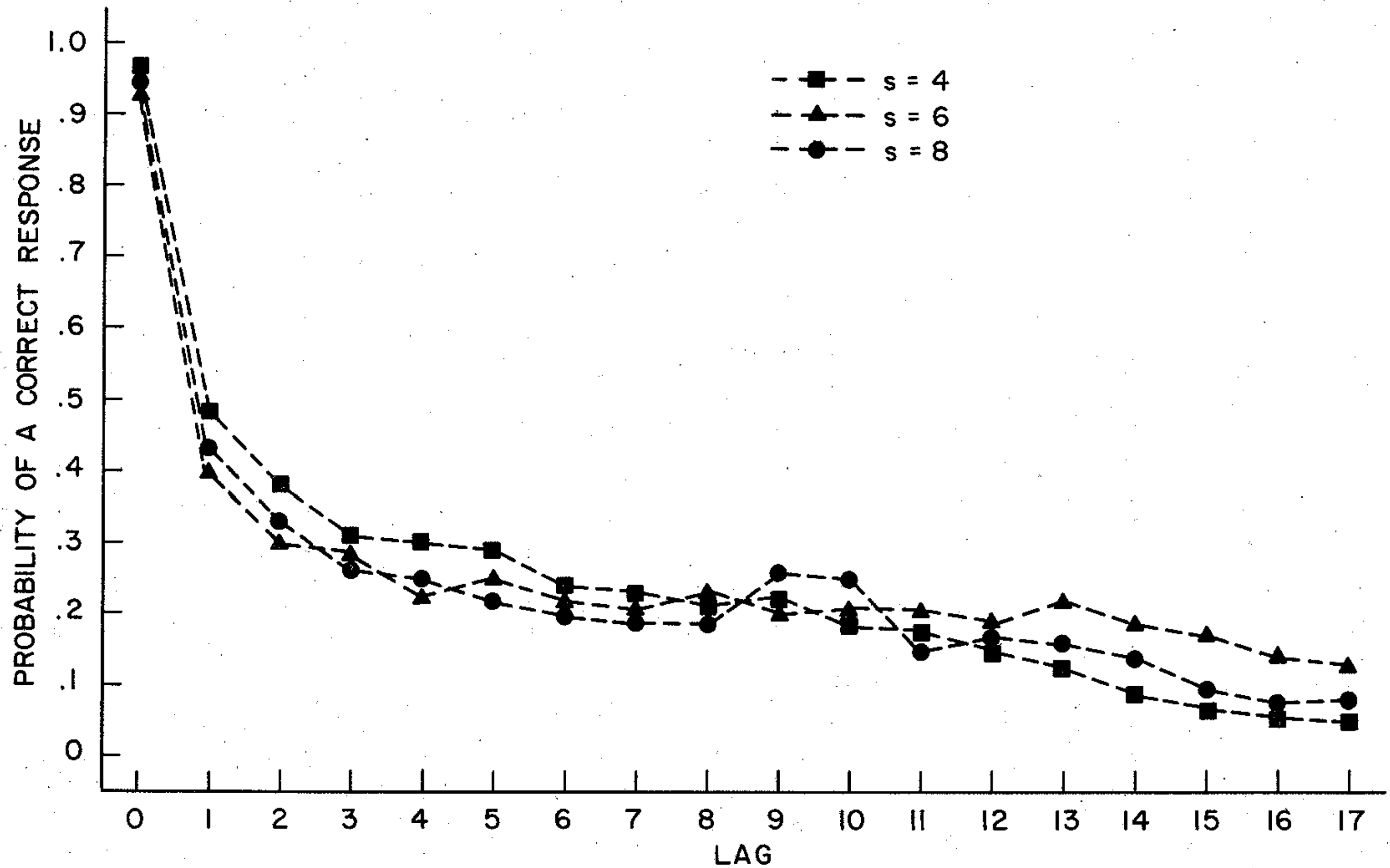


Figure 5. Observed and theoretical probabilities of a correct response as a function of lag. (Experiment II).

to overlap each other.[12] For this reason we lump the three curves to form the single lag curve displayed in Fig. 6.

Theoretical Analysis

Because the lag curves for the three conditions are not separated we assume that every item has an independent probability, α, of entering the buffer. If an item does enter, it randomly knocks out any one of the items already there. Under these assumptions we define

β°_j = probability that an item will be knocked out of the buffer by exactly the j^{th} succeeding item.

For this event to happen the following must hold: (1) the item must enter the buffer initially; (2) the item must not be knocked out for $j - 1$ trials; (3) the item must be knocked out by the j^{th} following item. Therefore

$$\beta^{\circ}_j = \begin{cases} 1 - \alpha & , \text{ for } j = 0 \\ (1 - \beta^{\circ}_0)(1 - \frac{\alpha}{r})^{j-1} \frac{\alpha}{r} & , \text{ for } j > 0 \end{cases} \tag{11}$$

where α/r is the probability that an intervening item will knock out the item of interest.

The curve in Fig. 6 was then fit using the minimum χ^2 technique. The parameter estimates were $r = 2$, $\alpha = .52$, $\theta = .17$, and $\tau = .90$; the minimum χ^2 value was 14.62 based on 13 degrees of freedom. It can be seen that the fit is excellent. Except for r, the parameters differ somewhat from those found in Experiment I. This result is not too

[12] To determine whether the three curves in Fig. 5 differ reliably, the proportions correct for each subject and condition calculated and then ranked. An analysis of variance for correlated means did not yield significant effects ($F = 2.67$, df = 2/16, $p > .05$).

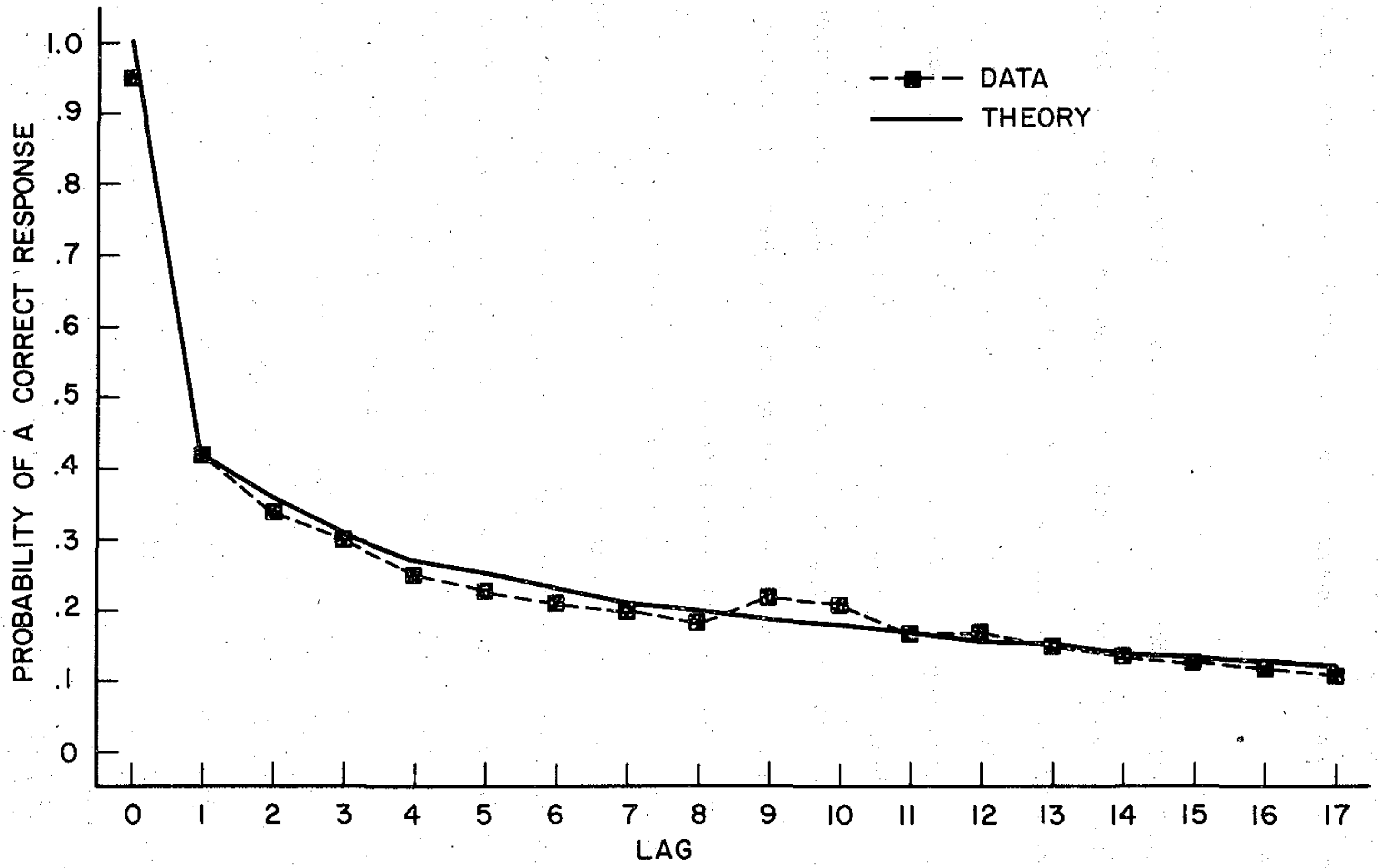


Figure 6. Observed and theoretical probabilities of a correct response as a function of lag. Data from the $s = 4, 6,$ and 8 conditions have been pooled to obtain the observed curve (Experiment II).

surprising considering the fact that the two experiments employ quite different procedures even though on logical grounds they can be regarded as equivalent.

Discussion

The difference in the effects of stimulus set size found in Experiments I and II suggests that the subject engages in an active decision process as each item is presented. This decision involves whether or not to enter the item into the memory buffer. The subject may also engage in a related decision regarding whether or not to transfer information on a given item to LTS. The experiments reported in this paper do not bear on this second point, but this type of decision undoubtedly would be important in studies of learning where each entering item may have been studied before as in the typical paired-associate paradigm.

An extended discussion of the relation of this model to other theories of memory may be found in Atkinson and Shiffrin (1965). The following points, however, are worth brief mention here. The model contains both all-or-none and incremental components: retrieval from the buffer is all-or-none and the buildup and decline of information in LTS in incremental. It is possible, however, to view LTS in a more discrete fashion than was done in this paper. For example, the transfer process might involve making partial copies of items in the buffer and then placing them in LTS. The number of copies made, of course, could depend on the length of time the item resided in the buffer. With one such copy the subject may be able to make a correct recognition response, whereas multiple copies would be needed for a correct recall response. Retroactive interference effects are also represented in the model. A sharp retroactive interference effect occurs in the buffer caused by the knockout process; a weaker effect occurs in LTS which is represented by the decay process. While proactive interference effects are not explicitly handled in the present paper, the general statement of the model includes a representation of them (Phillips, Shiffrin, and Atkinson, 1966). In the present study it is assumed that interference caused by preceding items in the sequence averages out at each lag. Finally, we note that other writers, in particular Broadbent (1963), Bower (1966), and Estes (1966), have presented theoretical models which mesh nicely with the conceptualization presented here.

Appendix

Throughout this paper it has been assumed that information is transferred to LTS at a constant rate, θ, during the entire period that an item resides in the buffer. Thus, if an item remains in the buffer for j trials, $j\theta$ is the amount of information transferred to LTS. Although this process seems reasonable to us, alternative schemes can be proposed. In particular, it can be assumed that an amount of information equal to θ is transferred to LTS at the time an item enters the memory buffer, and that this ends the transfer process for that item independent of any further time that it stays in the buffer. Thus any item that enters the memory buffer would have the same amount of information transferred to LTS. Two versions of this new model now come to mind: the information in LTS may start decaying at once, or the information may not start decaying until the item is knocked out of the memory buffer. These two versions are represented by the following retrieval functions:

$$\rho_{ij}^{(A)} = 1 - (1-g)\exp[-\theta\tau^{i}] \qquad (A)$$

$$\rho_{ij}^{(B)} = 1 - (1-g)\exp[-\theta\tau^{i-j}] \ . \qquad (B)$$

In order to make predictions from these models $\rho_{ij}^{(A)}$ and $\rho_{ij}^{(B)}$ were substituted for ρ_{ij} in Eq. 4. These two models were then fit to the unconditional lag curves from Experiment I using the same method as before; i.e., a minimum χ^2 estimate of the four parameters was obtained. For Model A the minimum χ^2 was 51.47 and the parameter estimates were $r = 2$, $\alpha = .30$, $\theta = .90$, and $\tau = 1.0$. For Model B the χ^2 procedure also yielded a best fit when $\tau = 1.0$. Since the Models A and B are identical when $\tau = 1.0$ the χ^2 and the parameter estimates are the same for both models.

Because the minimum χ^2's for Models A and B were somewhat larger than that for the version in the body of the paper, and because the earlier version seemed more reasonable, we have relegated these two models to an appendix. It should be noted, however, that these models do not require the assumption of a decay process. More precisely, the assumption of a decay process does not improve the fit of Models A and B (i.e., when τ equals one the models predict no decay in LTS). These alternative models are of interest also because they represent various branches of the general family of multi-process memory models formulated by Atkinson and Shiffrin (1965). There remain many other branches, however, that are as yet unexplored. In this regard, it is interesting to speculate that a model postulating a larger amount of information transfer when an item first enters the buffer, with smaller amounts thereafter, might fit the data as well as the version in this paper without requiring an LTS decay process.

References

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

Atkinson, R. C., and Shiffrin, R. M. Mathematical models for memory and learning. Tech. Rep. 79, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965. (To be published in The anatomy of memory, vol. 3, proceedings of the third conference on learning, remembering and forgetting, edited by D. P. Kimble, Palo Alto, California: Science and Behavior Books and Company, 1966.)

Bower, G. H. Notes on a descriptive theory of memory. In D. A. Kimble (Ed.) Learning, remembering, and forgetting: proceedings of the second conference. New York: New York Acad. Sci., 1966.

Brelsford, J. W., Jr., Keller, L., Shiffrin, R. M., and Atkinson, R. C. Short-term recall of paired-associates as a function of the number of interpolated pairs. Psychonomic Science, 1966, 4, 73-74.

Broadbent, D. E. Flow of information within the organism. Journal of Verbal Learning and Verbal Behavior, 1963, 4, 34-39.

Conrad, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-84.

Estes, W. K. Learning and memory. In E. F. Beckenbach (Ed.) Conceptual bases and applications of the communications sciences, 1966, in press.

Holland, P. A variation on the minimum chi-square test. Tech. Rep. 85, Institute for Mathematical Studies in the Social Sciences, Stanford Univer., 1965.

Katz, L. A technique for the study of steady-state short-term memory. Psychonomic Science, 1966, 4, 361-362.

Peterson, L. R., and Peterson, Margaret. Short-term retention of individual verbal items. Journal of Experimental Psychology, 1959, 58, 193-198.

Phillips, J. L., Shiffrin, R. M., and Atkinson, R. C. The effects of list length on short-term memory. Journal of Verbal Learning and Verbal Behavior, 1966, in press.

Wickelgren, W. A. Acoustic similarity and retroactive interference in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 53-61.

Wickelgren, W. A., and Norman, D. A. Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 1966, 3, in press.

Yntema, D. B., and Mueser, G. D. Keeping track of variables that have few or many states. Journal of Experimental Psychology, 1962, 63, 391-395.

SOME TWO-PROCESS MODELS FOR MEMORY

by

R. C. Atkinson and R. M. Shiffrin

TECHNICAL REPORT NO. 107

September 30, 1966

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Abstract

A general theoretical framework is developed in which to view memory and learning. The basic model is presented in terms of a memory system having two central components: a transient-memory buffer and a long-term store. Each stimulus item is postulated to enter a constant-sized push-down memory buffer, stay a variable amount of time and leave on a probabilistic basis when displaced by succeeding inputs. During the period that each item resides in the buffer, copies of the item are placed in the long-term store. The remaining feature of the model is concerned with the recovery of items from the memory system at the time of test. If at this time an item is still present in the buffer, it is perfectly retrieved. If an item is not present in the buffer, a search of the long-term store is made. This search is imperfect and the greater the number of items in the long-term store, the smaller the probability that any particular one will be retrieved. The model is applied to a set of experiments on paired-associate memory with good success.

Some Two-Process Models for Memory[1]

R. C. Atkinson and R. M. Shiffrin

Stanford University

A model for memory will be outlined in this paper. The experimental framework for which the model was constructed is that in which a series of items is presented to the subject who is then required to recall one or more of them. A familiar example is the digit span test in which the subject is required to repeat a series of digits read to him. A typical finding in digit span studies is that performance is error free until a critically large number of digits is reached. Thus a short-term memory system, called the "buffer," is proposed which may hold a fixed number of digits and allows perfect retrieval of those digits currently held. Errors are made only when the number of digits presented exceeds the capacity of the buffer, at which time the previous digits are forced out of the buffer. We propose, in addition, a long-term memory system (abbreviated LTS for long-term store) which allows items not present in the buffer to be recalled with some probability between 0 and 1. This two-process model will be presented in the first part of the paper and then applied to data from an experiment in paired-associate memory in the second part of the paper.

Insert Figure 1 about here

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

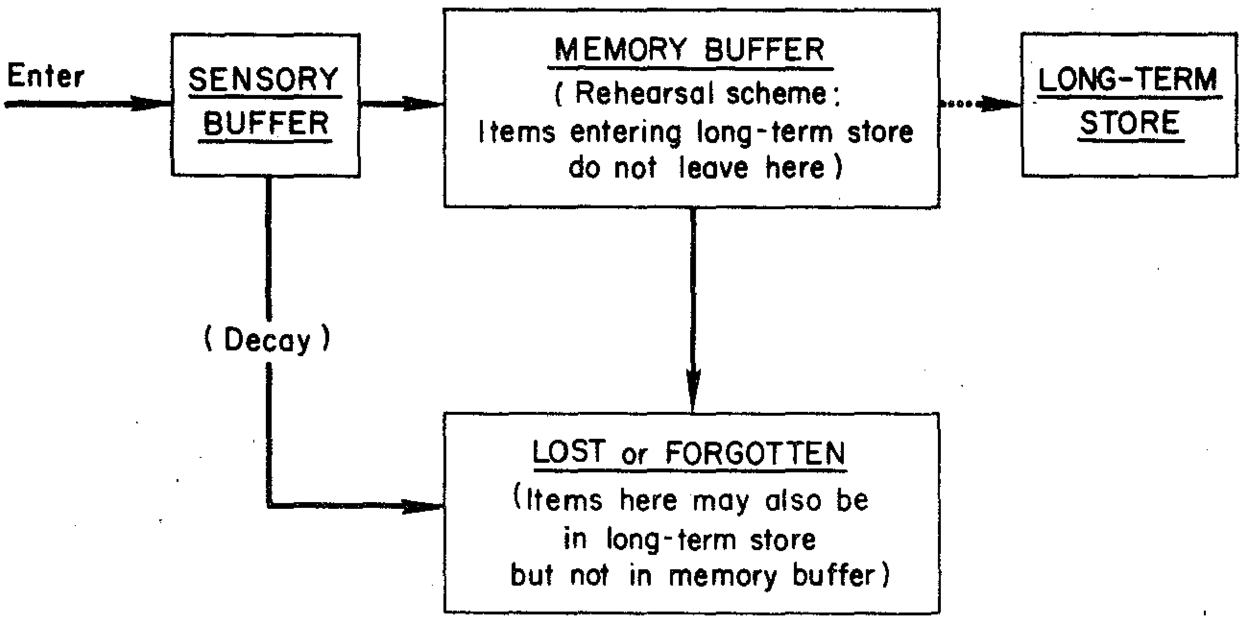


Fig. 1. Flow chart for the general system.

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., Estes and Taylor, 1964; Mackworth, 1963; Sperling, 1960), but 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 usccessfully 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 an analysis of what is meant by an "item." If the word "cat" 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 cats) is sufficient to permit the subject to report back the word "cat" 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, 1956). We prefer the term buffer because of the wide range of applications for which the term short-term store has been used.

Insert Figure 2 about here

Figure 2 illustrates the workings of the memory buffer. The properties of the buffer will be examined successively.

1. Constant size. The buffer can contain exactly r items and no more. This statement holds within any experimental situation. The buffer size will change when the type of items change. For example, if the items are single digits, the buffer size might be five, but if the items are five-digit numbers the buffer size would correspondingly be one. We should like eventually to be able to permanently fix the buffer size on a more molecular basis than "items": for example, on some such basis as the amount of information transmitted, or the length of the auditory code for the items. This is still an open question and at present the buffer size must be estimated separately for each experiment.

A second important point concerns what we mean by an item. In the experiments that the model is designed to handle there is a clearly separated series of inputs and a clearly defined response. In these cases, the "item" that is placed in the buffer may be considered to be an amount of information which is sufficient to allow emission of the correct response.

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

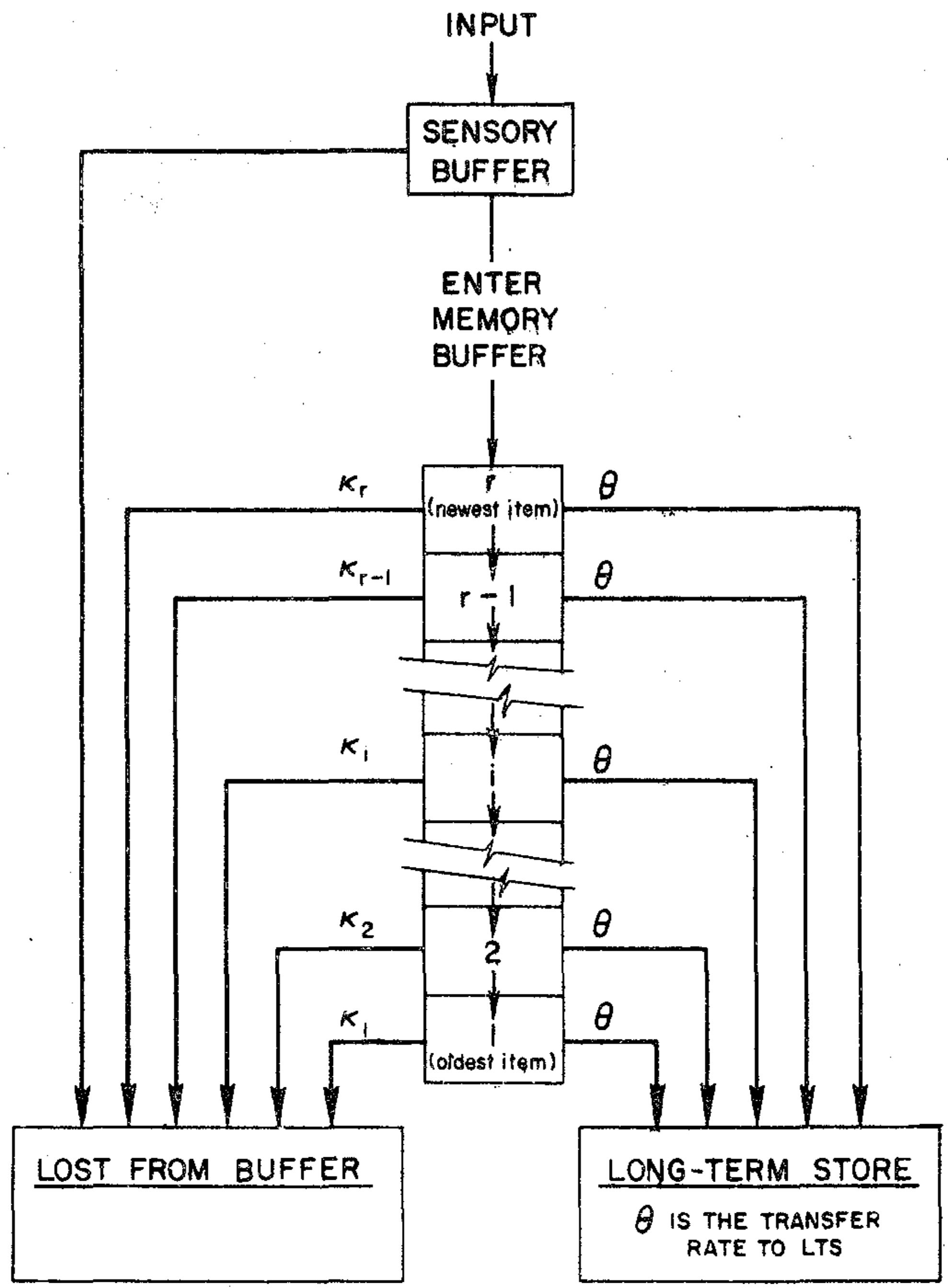


Fig. 2. A flow chart characterizing inputs to the memory system.

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. It should be noted that the lost state refers only to the fact that an item has left the buffer and says nothing regarding the item's presence in LTS.

3. <u>Buffer stays filled</u>. 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. This state of affairs is assumed to hold as long as the subject is paying attention. In this matter we tend to follow Broadbent (1963) and view the buffer as the input-output mechanism for information transmission between the subject and the environment. At the end of a trial, for example, attention ceases, the subject "thinks" of other things, and the buffer gradually empties of that trial's items.

4. <u>Each new item bumps out an old item</u>. 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 $\kappa_1 + \kappa_2 + \cdots + \kappa_r = 1$, since exactly one item is lost. Various schemes can be proposed for the generation of the κ's. The simplest scheme, requiring no additional parameters, is to equalize the κ's: i.e., let $\kappa_j = 1/r$ for all j.

A useful one-parameter scheme can be derived as follows: the oldest item (in slot 1) is dropped with probability δ. If that item is not 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 first slot. This process continues until an item is dropped. Hence

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

It is easy to see that as δ approaches 0, κ_j approaches $1/r$ for all j, which was the earlier case mentioned. On the other hand, when $\delta = 1$, the oldest item is always the one lost. Intermediate values of δ allow a bump-out process between these two extremes. We would expect that the tendency to bump out the oldest item first would depend on such factors as the serial nature of the task, the subject's instructions, and the subject's knowledge concerning the length of the list he is to remember.

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 fairly slow inputs, such as the experiment to be considered later in this paper.

6. Perfect recovery of items 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 point leads to the question,

"What is stored in the buffer?" and "What is an item?" In terms of the preceding requirement (and in accord with the mathematical structure of the model) we may be satisfied with the definition, "an item is that amount of information that allows correct performance at the time of test." Because the model does not require a more precise statement than the above, it is not necessary in the present analysis to spell out the matter in detail. Nevertheless, in view of the work of Conrad (1964), Wickelgren (1965), and others on auditory confusions in short-term memory, we would be satisfied with the view that items in the buffer are acoustic mnemonics and are kept there via rehearsal, at least for experiments of a verbal character.

7. Buffer is unchanged by the transfer process to LTS. We will say more about LTS and transfer to it in the next section, but here it may be said that whatever transfer takes place, and whenever the transfer takes place, the buffer remains unchanged. That is, if a copy of an item is placed in LTS, the item remains represented in the buffer, and the buffer remains unchanged.

This set of seven assumptions characterizes the memory buffer. Now we consider the long-term memory system. In recent years a number of mathematical models for memory and learning have made use of a state labeled "long-term store." In most of these cases, however, the term is used to denote a completely learned state. LTS in this case is used in a very different manner; information concerning each item is postulated to enter LTS during the period the item remains in the buffer. This information may or may not be sufficient to allow recall of the item, and even if sufficient information to allow recall is stored, the subject may fail to recall because he still must search LTS for the appropriate information.

There are many possible representations of the transfer process to LTS. Let θ_{ij} be the transfer parameter representing the amount transferred to LTS of an item in slot i of the buffer between one item presentation and the next if there are currently j items in the buffer. In the present version θ_{ij} is the probability of copying an item into LTS during each presentation period.

For this discussion we will assume that θ does not depend on the position in the buffer, but does depend on the number of other items currently in the buffer. The justification for this is based on the amount of attention that an item will receive during each presentation period; thus an item will receive r times as much attention if it is the only item in the buffer than if all r buffer positions were filled. Hence θ_{ij} is set equal to θ/j. It is further assumed that there may be more than one copy of any item in LTS. Since one copy may be made during each presentation period, the maximum number of copies that can exist in LTS for a particular item equals the number of presentation periods that the item stayed in the buffer.

The retrieval rules are relatively simple. At the time of test any item in the buffer is recalled perfectly. If the item is not present in the buffer then a search of LTS is made. If the item is found in LTS it is recalled; if not, then the subject guesses. The search process the subject engages in is postulated to be a search made uniformly with replacement from the pool of items in LTS which are not in the buffer. (An alternative scheme is to pick from all the items in LTS, which gives very similar results to those given by the stated scheme.) In particular, the subject is said to make R random picks in LTS; if none of these picks finds the desired item, it is reported; otherwise the subject guesses.

The mathematical development of this model is presented in Atkinson and Shiffrin (1965). For present purposes, it is sufficient to note that there are four parameters available to fit the data: r, the buffer size; θ, the transfer probability; δ, the tendency to bump out the oldest item in the buffer first; and R, the number of searches into LTS.

We now turn to an experiment in human paired-associate memory (Phillips, Shiffrin, and Atkinson, 1967). 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 the 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.

Following the subject's response, 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.

Insert Figure 3 about here

Figure 3 presents the proportion of correct responses as a function of the test position in the display. Display sizes 3 and 4 are not graphed because performance was essentially perfect for these cases. Observed points 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. 3 also indicates that the overall proportion correct is a decreasing function of display size.

The model was fit to the data using a minimum chi-square technique.[2] The details are presented in Atkinson and Shiffrin (1965). It will merely be pointed out here that the value of r was set equal to 5 before the minimization because performance was essentially error free for list lengths of 5 and less. The other three parameters were fit using a grid search

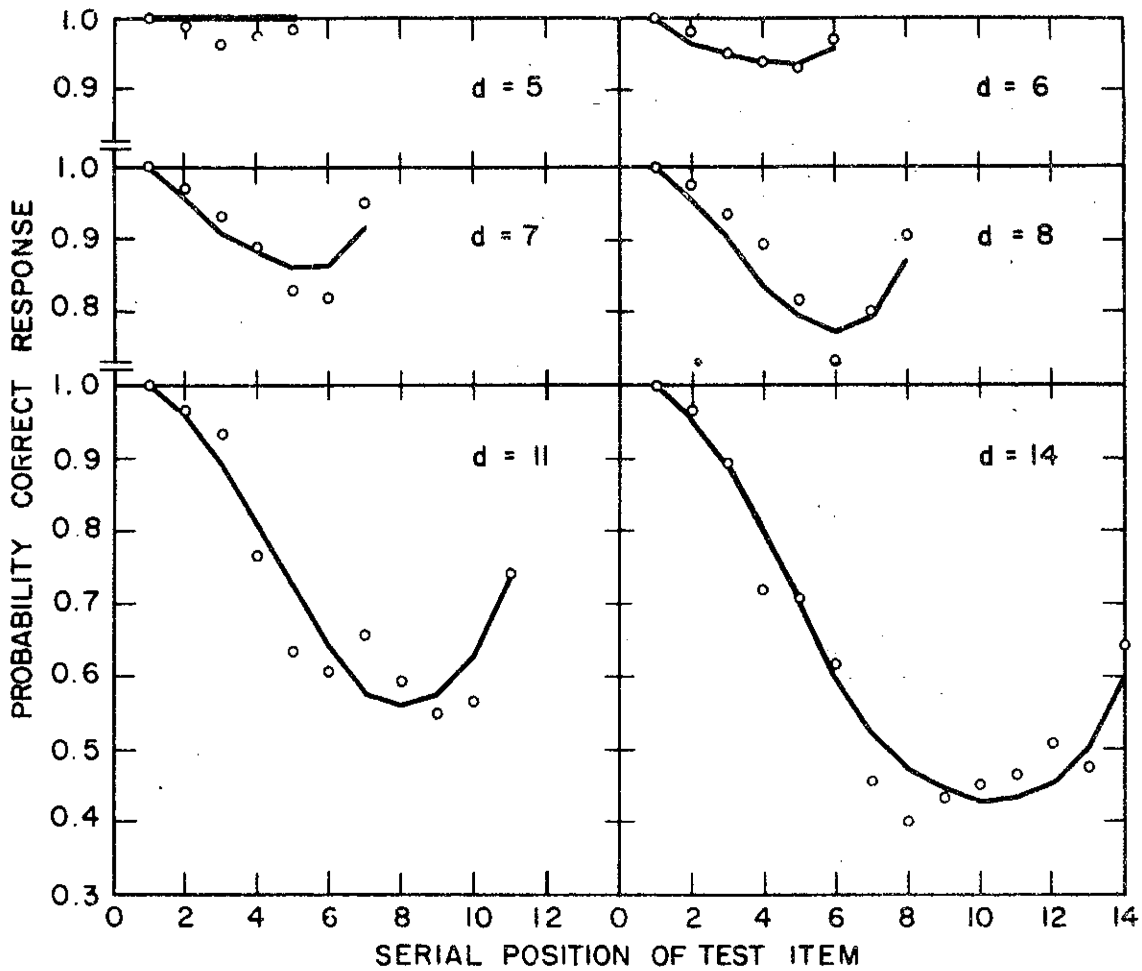


Fig. 3. Goodness-of-fit results for the paired-associate memory experiment.

procedure on a computer. The parameter estimates were as follows:

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

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

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

The predicted curves are given in Fig. 3. It should be emphasized that the same 4 parameters are used to fit the serial position curves for all five list lengths. It can be seen that the fit is quite good with a minimum chi-square of 46.2 based on 43 degrees of freedom.

We have outlined only one example of how this model can be applied to data. Other applications of the model have been made including experiments involving a continuous-presentation memory task, free-verbal recall, paired-associate learning, and serial-anticipatory learning; also, the model has been used to perdict not only response probabilities, but confidence ratings and latency data. Time does not permit us to present these developments here; for a review of such applications see Atkinson and Shiffrin (1965), Atkinson, Brelsford, and Shiffrin (1967), Brelsford and Atkinson (1967), and Phillips, Shiffrin, and Atkinson (1967). In conclusion, it should be pointed out that of all the assumptions 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., Bower, G. H., and Crothers, E. J. *An introduction to mathematical learning theory*. New York: Wiley, 1965.

Atkinson, R. C., Brelsford, J. W., Jr., and Shiffrin, R. M. Multi-process models for memory with applications to a continuous presentation task. *Journal of Mathematical Psychology*, 1967, 4, in press.

Atkinson, R. C., and Shiffrin, R. M. Mathematical models for memory and learning. Technical Report 79, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965. (To be published in D. P. Kimble (Ed.), *Proceedings of the Third Conference on Learning, Remembering, and Forgetting*. New York: The New York Academy of Sciences, 1966.)

Bower, G. H. Notes on a descriptive theory of memory. In D. P. Kimble (Ed.), *Proceedings of the Second Conference on Learning, Remembering, and Forgetting*. New York: The New York Academy of Sciences, 1966, in press.

Brelsford, J. W., Jr., and Atkinson, R. C. Recall of paired-associates as a function of overt and covert rehearsal procedures. Technical Report 111, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California, 1966.

Broadbent, D. E. Flow of information within the organism. *Journal of Verbal Learning and Verbal Behavior*, 1963, 4, 34-39.

Brown, J. Short-term memory. *British Medical Bulletin*, 1964, 20(1), 8-11.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, 55, 75-84.

Estes, W. K., and Taylor, H. A. A detection method and probabilistic models for assessing information processing from brief visual displays. _Proceedings of the National Academy of Sciences_, 1964, 52, 2, 446-454.

Mackworth, J. F. The relation between the visual image and post-perceptual immediate memory. _Journal of Verbal Learning and Verbal Behavior_, 1963, 2, 113-119.

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.

Phillips, J. L., Shiffrin, R. M., and Atkinson, R. C. The effects of display size on short-term memory. _Journal of Verbal Learning and Verbal Behavior_, 1967, in press.

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

Wickelgren, W. A. Acoustic similarity and retroactive interference in short-term memory. _Journal of Verbal Learning and Verbal Behavior_, 1965, 4, 53-61.

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

Footnotes

[1]Support for this research was provided by the National Aeronautics and Space Administration, Grant NGR-05-020-036, and United States Public Health Service MH06154. This paper was read at the Symposium on Mathematical Models of Psychological Processes, XVIII International Congress of Psychology, Moscow, USSR, August, 1966.

[2]For a discussion of minimum chi-square procedures for parameter estimation, see Atkinson, Bower, and Crothers (1965).

HUMAN MEMORY: A PROPOSED SYSTEM AND ITS CONTROL PROCESSES

R. C. Atkinson and R. M. Shiffrin

Stanford University

Prepublication draft of a chapter to appear in K. W. Spence and J. T. Spence (Eds.), The Psychology of Learning and Motivation: Advances in Research and Theory. Vol. II. New York: Academic Press.

* This research was supported by the National Aeronautics and Space Administration, Grant No. NGR-05-020-036. The authors are indebted to W. K. Estes and G. H. Bower who provided many valuable suggestions and comments at various stages of the work. Special credit is due J. W. Brelsford who was instrumental in carrying out the research discussed in Section 4 and whose overall contributions are too numerous to report in detail. We should also like to thank those co-workers who carried out a number of the experiments discussed in the latter half of the paper; rather than list them here, each will be acknowledged at the appropriate place.

TABLE OF CONTENTS

Section 1. INTRODUCTION 1

Section 2. STRUCTURAL FEATURES OF THE MEMORY SYSTEM 5
2.1 Sensory Register 9
2.2 Short-Term Store 12
2.3 Long-Term Store 24

Section 3. CONTROL PROCESSES IN MEMORY 29
3.1 Control Processes in Sensory Register 30
3.2 Control Processes in Short-Term Store 32
3.3 Control Processes in Long-Term Store 46

Section 4. EXPERIMENTS CONCERNED WITH SHORT-TERM PROCESSES . . . 57
4.1 A Continuous Paired-Associate Memory Task (Experiment 1) 57
4.2 The "All-Different" Stimulus Procedure (Experiment 2) 78
4.3 A Continuous Paired-Associate Memory Task with Multiple Reinforcements (Experiment 3) 88
4.4 Overt vs. Covert Study Procedures (Experiment 4). 110
4.5 Additional Variables Related to the Rehearsal Buffer (Experiments 5, 6, and 7) 118

Section 5. EXPERIMENTS CONCERNED WITH LONG-TERM SEARCH AND RETRIEVAL 129
5.1 A Serial Display Procedure Involving Single Tests (Experiment 8) 130
5.2 Free-Verbal Recall Experiments 145
5.3 Further Considerations Involving LTS 161

Section 6. CONCLUDING REMARKS 173

LIST OF FIGURES

Figure		Page
1	Structure of the memory system	6
2	The rehearsal buffer and its relation to the memory system .	41
3	A sample sequence of trials for Experiment 1	62
4	Observed and theoretical probabilities of a correct response as a function of lag (Experiment 1)	70
5	Observed and theoretical probabilities of a correct response as a function of lag when every intervening item uses the same stimulus (Experiment 1)	74
6	Observed and theoretical probabilities of a correct response as a function of lag when every intervening item uses a different stimulus (Experiment 1)	75
7	Observed and theoretical probabilities of a correct response as a function of lag (Experiment 2)	82
8	Observed and theoretical probabilities of a correct response as a function of lag. Data from the $s = 4, 6,$ and 8 conditions have been pooled (Experiment 2)	83
9	Probability of a correct response as a function of the number of consecutive preceding items tested at zero lag (Experiment 1 and Experiment 2)	86
10	A sample sequence of trials for Experiment 3	90
11	Observed and theoretical probabilities of a correct response as a function of lag for items tested following their 1st, 2nd or 3rd reinforcement (Experiment 3)	93
12	Observed and theoretical probabilities of a correct response as a function of lag $\underline{a}$ (the spacing between the 1st and 2nd reinforcement) (Experiment 3)	95
13	Observed and theoretical probabilities of a correct response as a function of lag for the "all-same" and "all-different" conditions (Experiment 3)	99
14	Observed and theoretical probabilities of a correct response as a function of the number of consecutive preceding items using the same stimulus (Experiment 3)	101

LIST OF FIGURES (cont.)

Figure		Page
15	Observed and theoretical probabilities of a correct response as a function of lag $\underline{a}$ (the lag of the item preceding the item tested, but using the same stimulus) (Experiment 3) . .	103
16	Observed and theoretical probabilities of a correct response as a function of lag $\underline{a}$ conditionalized on errors or successes on the test at lag $\underline{a}$ (Experiment 3)	105
17	Probability that the correct response for the preceding item using the same stimulus will be given in error to the present item (Experiment 3)	108
18	Observed and theoretical probabilities of a correct response as a function of lag (Experiment 4)	112
19	Observed and theoretical probabilities of a correct response as a function of lag for the "all-same" and "all-different" conditions (Experiment 4)	116
20	Observed and theoretical probabilities of a correct response as a function of the number of consecutive preceding items all using the same stimulus (Experiment 4)	117
21	Observed and theoretical probabilities of a correct response as a function of lag, for the overall condition and for the "all-known-intervening" and "all-unknown-intervening" conditions (Experiment 5)	120
22	Observed and theoretical mean latencies as a function of lag for correct and incorrect responses (Experiment 6)	124
23	Observed and theoretical probabilities of a correct response as a function of lag (Experiment 7)	126
24	Observed and theoretical mean lag judgments as a function of the actual lag (Experiment 7)	127
25	Observed and theoretical probabilities of a correct response as a function of serial position (Experiment 8)	132
26	Observed and predicted probabilities of a confidence rating R_1 as a function of serial position (Experiment 8)	139
27	Probability of correct recall as a function of serial position for free verbal recall (after Murdock, 1962)	147

LIST OF FIGURES (cont.)

Figure | Page

28 Probability of correct recall as a function of serial position for free verbal recall with test following 0 seconds and 30 seconds of intervening arithmetic (after Postman and Phillips, 1965) 151

29 Probability of a correct response prior to the last error as a function of lag (after Bjork, 1966) 171

SECTION 1. INTRODUCTION

This paper is divided into two major portions; the first outlines a general theoretical framework in which to view human memory, and the second describes the results of a number of experiments designed to test specific models that can be derived from the overall theory.

The general theoretical framework, set forth in Sections 2 and 3, categorizes the memory system along two major dimensions. One categorization distinguishes permanent, structural features of the system from control processes that can be readily modified or reprogrammed at the will of the subject. Because we feel that this distinction helps clarify a number of results, we will take time to elaborate it at the outset. The permanent features of memory, which will be referred to as the memory structure, include both the physical system and the built-in processes that are unvarying and fixed from one situation to another. Control processes, on the other hand, are selected, constructed, and used at the option of the subject and may vary dramatically from one task to another even though superficially the tasks may appear very similar. The use of a particular control process in a given situation will depend upon such factors as the nature of the instructions, the meaningfulness of the material, and the individual subject's history.

A computer analogy might help illustrate the distinction between memory structure and control processes. If the memory system is viewed as a computer under the direction of a programmer at a remote console, then both the computer hardware and those programs built into the system that cannot be modified by the programmer are analogous to our structural features; those programs and instruction sequences, which the programmer can write at his console, and which determine the operation of the computer, are analogous to our control processes. In the sense that the computer's method of processing a given batch of data depends on the operating program, so the way a stimulus input is processed depends on the particular control processes the subject brings into play. The structural components include the basic memory stores; examples of control processes are coding procedures, rehearsal operations, and search strategies.

Our second categorization divides memory into three structural components: the sensory register, the short-term store, and the long-term store. Incoming sensory information first enters the sensory register where it resides for a very brief period of time, then decays and is lost. The short-term store is the subject's working memory; it receives selected inputs from the sensory register and also from long-term store. Information in the short-term store decays completely and is lost within a period of about 30 seconds, but a control process called rehearsal can maintain a limited amount of information in this store as long as the subject desires. The long-term store is a fairly permanent repository for information, information which is transferred from the short-term store. Note that "transfer" is not meant to imply that information is removed from one store and placed in the next; we use transfer to mean the copying of selected information from one store into the next without removing this information from the original store.

In presenting our theoretical framework we will consider first the structural features of the system (Section 2) and then some of the more generally used control processes (Section 3). In both of these sections the discussion is organized first around the sensory register, then the short-term store, and finally the long-term store. Thus, the outline of Sections 2 and 3 can be represented as follows:

	Sensory Register	Short-Term Store	Long-Term Store
Structure	Sec. 2.1	Sec. 2.2	Sec. 2.3
Control Processes	Sec. 3.1	Sec. 3.2	Sec. 3.3

These first sections of the paper do not present a finished theory; instead they set forth a general framework within which specific models can be formulated. We attempt to demonstrate that a large number of results may be handled parsimoniously within this framework, even without coming to final decisions at many of the choice points that occur. At many choice points several hypotheses will be presented, and the evidence that is available to help make the choice will be reviewed. The primary goal of Sections 2 and 3 is to justify our theoretical framework and to demonstrate that it is a useful way of viewing a wide variety of memory phenomena.

The remaining sections of the paper present a number of precise models that satisfy the conditions imposed by our general theoretical framework. These sections also present data from a series of experiments designed to evaluate the models. Section 4 is concerned with an analysis of short-term memory; the model used to analyze the data emphasizes a control process based in the short-term store which we designate a rehearsal buffer. Section 5 presents several experiments which shed some light upon processes in the long-term store, especially subject-controlled search processes. Some of the experiments in Sections 4 and 5 have been reported by us and our co-workers in previous publications, but the earlier treatments were primarily mathematical whereas the present emphasis is upon discussion and overall synthesis.

If the reader is willing to accept our overall framework on a provisional basis and wishes to proceed at once to the specific models and experiments, then he may begin with Section 4 and as a prerequisite need only read that portion of Section 3.2 concerned with the rehearsal buffer.

SECTION 2. STRUCTURAL FEATURES OF THE MEMORY SYSTEM

This section of the paper will describe the permanent, structural features of the memory system. The basic structural division is into the three components diagramed in Figure 1; the sensory register, the short-term store, and the long-term store.

When a stimulus is presented there is an immediate registration of that stimulus within the appropriate sensory dimensions. The form of this registration is fairly well understood in the case of the visual system (Sperling, 1960); in fact, the particular features of visual registration (including a several hundred millisecond decay of an initially accurate visual image) allow us positively to identify this system as a distinct component of memory. It is obvious that incoming information in other sense modalities also receives an initial registration, but it is not clear whether these other registrations have an appreciable decay period or any other features which would enable us to refer to them as components of memory.

The second basic component of our system is the short-term store. This store may be regarded as the subject's "working memory." Information entering the short-term store is assumed to decay and disappear completely, but the time required for the information to be lost is considerably longer than for the sensory register. The character of the information in the short-term store does not depend necessarily upon the form of the sensory input. For example, a word presented visually may be encoded from the visual sensory register into an auditory short-term

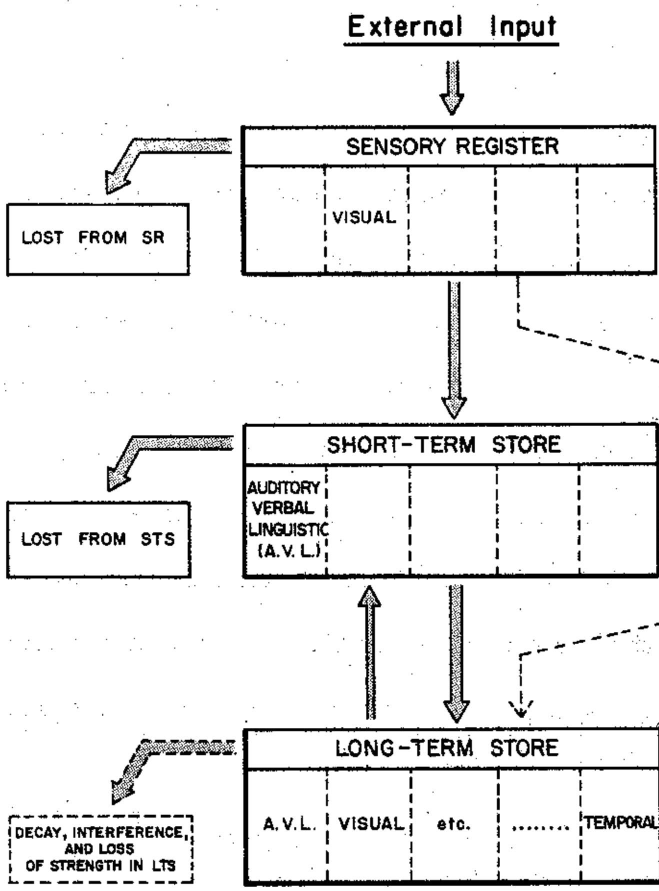


Figure 1 Structure of the memory system

store. Since the auditory short-term system will play a major role in subsequent discussions, we shall use the abbreviation a-v-l to stand for auditory-verbal-linguistic store. The triple term is used because, as we shall see, it is not easy to separate these three functions.

The exact rate of decay of information in the short-term store is difficult to estimate because it is greatly influenced by subject-controlled processes. In the a-v-l mode, for example, the subject can invoke rehearsal mechanisms that maintain the information in STS and thereby complicate the problem of measuring the structural characteristics of the decay process. However, the available evidence suggests that information represented in the a-v-l mode decays and is lost within a period of about 15 to 30 seconds. Storage of information in other modalities is less well understood and, for reasons to be discussed later, it is difficult to assign values to their decay rates.

The last major component of our system is the long-term store. This store differs from the preceding ones in that information stored here does not decay and become lost in the same manner. All information eventually is completely lost from the sensory register and the short-term store, whereas information in the long-term store is relatively permanent (although it may be modified or rendered temporarily irretrievable as the result of other incoming information). Most experiments in the literature dealing with long-term store have been concerned with storage in the a-v-l mode, but it is clear that there is long-term memory in each of the other sensory modalities, as demonstrated by an ability to recognize stimuli presented to these senses. There may even be information in the long-term store which is not classifiable into any of the sensory modalities, the prime example being temporal memory.

The flow of information among the three systems is to a large degree under the control of the subject. Note that by information flow and transfer between stores we refer to the same process: the copying of selected information from one store into the next. This copying takes place without the transferred information being removed from its original store. The information remains in the store from which it is transferred and decays according to the decay characteristics of that store. In considering information flow in the system we start with its initial input into the sensory register. The next step is a subject-controlled scan of the information in the register; as a result of this scan and an associated search of long-term store, selected information is introduced into short-term store. We assume that transfer to the long-term store takes place throughout the period that information resides in the short-term store, although the amount and form of the transferred information is markedly influenced by control processes. The possibility that there may be direct transfer to the long-term store from the sensory register is represented by the dashed line in Figure 1; we do not know whether such transfer occurs. Finally, there is transfer from the long-term store to the short-term store, mostly under the control of the subject; such transfer occurs, for example, in problem solving, hypothesis-testing, and "thinking" in general.

This brief encapsulation of the system raises more questions than it answers. Not yet mentioned are such features as the cause of the decay in each memory store and the form of the transfer functions between the stores. In an attempt to specify these aspects of the system, we now turn to a more detailed outline including a review of some relevant literature.

2.1. Sensory Register.

The prime example of a sensory register is the short-term visual image investigated by Sperling (1960, 1963), Averbach and Coriell (1961), Estes and Taylor (1964, 1966) and others. As reported by Sperling (1967), if an array of letters is presented tachistoscopically and the subject is instructed to call out as many letters as possible, usually about six letters are reported. Further, a 30-second delay between presentation and report does not cause a decrement in performance. This fact (plus the facts that confusions tend to be based on auditory rather than visual similarities, and that subjects report rehearsing and subvocalizing the letters) indicates that the process being examined is in the a-v-l short-term store; i.e., subjects scan the visual image and transfer a number of letters to the a-v-l short-term store for rehearsal and output. In order to study the registered visual image itself, partial-report procedures (Sperling, 1960; Averbach and Sperling, 1961; Averbach and Coriell, 1961; Sperling, 1963) and forced-choice detection procedures (Estes, 1965; Estes and Taylor, 1964, 1966; Estes and Wessel, 1966) have been employed. The partial-report method typically involves presenting a display (usually a 3 x 4 matrix of letters and numbers) tachistoscopically for a very brief period. After the presentation the subject is given a signal that tells him which row to report. If the signal is given almost immediately after stimulus offset, the requested information is reported with good precision, otherwise considerable loss occurs. Thus we infer that a highly accurate visual image lasts for a short period of time and then decays. It has also been established that succeeding visual stimulation can wipe out or replace prior stimulation.

By using a number of different methods, the decay period of the image has been estimated to take several hundred milliseconds, or a little more depending on experimental conditions; that is, information can not be recovered from this store after a period of several hundred milliseconds.

Using the detection method, in which the subject must report which of two critical letters was presented in a display, Estes and Taylor (1964, 1966) and Estes and Wessel (1966) have examined some models for the scanning process. Although no completely satisfactory models have yet been proposed, it seems reasonably certain that the letters are scanned serially (which letters are scanned seems to be a momentary decision of the subject), and a figure of about 10 milliseconds to scan one letter seems generally satisfactory.

Thus it appears fairly well established that a visual stimulus leaves a more or less photographic trace which decays during a period of several hundred milliseconds, and is subject to masking and replacement by succeeding stimulation. Not known at present is the form of the decay, that is, whether letters in a display decay together or individually, probabilistically or temporally, all-or-none or continuously. The reader may ask whether these results are specific to extremely brief visual presentations; although presentations of long duration complicate analysis (because of eye movements and physical scanning of the stimulus), there is no reason to believe that the basic fact of a highly veridical image quickly decaying after stimulus offset does not hold also for longer visual presentations. It is interesting that the stimulation seems to be transferred from the visual image to the a-v-l short-term store, rather than to a visual short-term store. The fact that a verbal report was requested may provide the explanation, or it may be that the visual short-term store lacks rehearsal capacity.

There is not much one can say about registers in sensory modalities other than the visual. A fair amount of work has been carried out on the auditory system without isolating a registration mechanism comparable to the visual one. On the other hand, the widely differing structures of the different sensory systems makes it questionable whether we should expect similar systems for registration.

Before leaving the sensory register it is worth adding a few comments about the transfer to higher order systems. In the case of the transfer from the visual image to the a-v-l short-term store it seems likely that a selective scan is made at the discretion of the subject.* As each element in the register is scanned, a matching program of some sort is carried out against information in long-term store and the verbal "name" of the element is recovered from long-term memory and fed into the short-term store. Other information might also be recovered in the long-term search; for example, if the scanned element was a pineapple, the word, its associates, the taste, smell and feel of a pineapple might all be recovered and transferred to various short-term stores. This communication between the sensory register and long-term store does not, however, permit us to infer that information is transferred directly to long-term store from the register. Another interesting theoretical question is

---

* Sperling (1960) has presented evidence relating the type of scan used to the subject's performance level.

whether the search into long-term store is necessary to transfer information from the sensory register to the short-term store within a modality. We see no a-priori theoretical reason to exclude non-mediated transfer (e.g., why should a scan or match be necessary to transfer a spoken word to the a-v-l short-term store). For lack of evidence, we leave these matters unspecified.

2.2. Short-Term Store.

The first point to be examined in this section is the validity of the division of memory into short- and long-term stores. Workers of a traditional bent have argued against dichotomizing memory (e.g., Postman, 1964; Melton, 1963). We, however, feel there is much evidence indicating the parsimony and usefulness of such a division. The argument is often given that one memory is somehow "simpler" than two; but quite the opposite is usually the case. A good example may be found in a comparison of the model for free recall presented in this paper and the model proposed by Postman and Phillips (1965). Any single-process system making a fair attempt to explain the mass of data currently available must, of necessity, be sufficiently complex that the term "single process" becomes a misnomer. We do not wish, however, to engage in the controversy here. We ask the reader to accept our model provisionally until its power to deal with data becomes clear. Still, some justification of our decision would seem indicated at this point. For this reason, we turn to what is perhaps the single most convincing demonstration of a dichotomy in the memory system: the effects of hippocampal lesions reported by Milner (1959, 1966, 1967). In her words:

"Bilateral surgical lesions in the hippocampal region, on the mesial aspect of the temporal lobes, produce a remarkably severe and persistent memory disorder in human patients, the pattern of breakdown providing valuable clues to the cerebral organization of memory. Patients with these lesions show no loss of preoperatively acquired skills, and intelligence as measured by formal tests is unimpaired, but, with the possible exception of acquiring motor skill, they seem largely incapable of adding new information to the long-term store. This is true whether acquisition is measured by free recall, recognition, or learning with savings. Nevertheless, the immediate registration of new input (as measured, for example, by digit span and dichotic listening tests) appears to take place normally and material which can be encompassed by verbal rehearsal is held for many minutes without further loss than that entailed in the initial verbalization. Interruption of rehearsal, regardless of the nature of the distracting task, produces immediate forgetting of what went before, and some quite simple material which cannot be categorized in verbal terms decays in 30 seconds or so, even without an interpolated distraction. Material already in long-term store is unaffected by the lesion, except for a certain amount of retrograde amnesia for preoperative events." (Milner, 1966). Apparently, a short-term store remains to the patients, but the lesions have produced a breakdown either in the ability to store new information in long-term store or to retrieve new information from it. These patients appear to be incapable of retaining new material on a long-term basis.*

As with most clinical research, however, there are several problems that should be considered. First, the patients were in a general sense abnormal to begin with; second, once the memory defect had been discovered, the operations were discontinued, leaving only a few subjects for observation; third, the results of the lesions seem to be somewhat variable, depending for one thing upon the size of the lesion, the larger lesions giving rise to the full syndrome. Thus there are only a few patients who exhibit the deficit described above in full detail. As startling as these patients are, there might be a temptation to discount them as anomalies but for the following additional findings. Patients who had known damage to the hippocampal area in one hemisphere were tested for memory deficit after an intracarotid injection of sodium amytal temporarily inactivated the other hemisphere. Controls were patients without known damage, and patients who received injections inactivating their

---

* A related defect in short-term memory, called Korsakoff's Syndrome, has been known for many years. Patients suffering from this abnormal condition are unable to retain new events for longer than a few seconds or minutes (e.g., they cannot recall the meal they have just eaten or recognize the face of the doctor who treated them a few minutes earlier) but their memory for events and people prior to their illness remains largely unimpaired and they can perform adequately on tests of immediate memory span. Recent evidence suggests that Korsakoff's Syndrome is related to damage of brain tissue, frequently as the result of chronic alcoholism, in the hippocampal region and the mammillary body (Barbizet, 1963).

damaged side. A number of memory tests were used as a criterion for memory deficit; the easiest consisted of presenting four pictures, distracting the patient, and then presenting nine pictures containing the original four. If the patient cannot identify the critical four pictures then evidence of memory deficit is assumed. The results showed that in almost all cases memory deficit occurs only after bilateral damage; if side A is damaged and side B inactivated memory deficit appears, but if the inactivated side is the damaged side, no deficit occurs. These results suggest that the patients described above by Milner were not anomalous cases and their memory deficits therefore give strong support to the hypothesis of distinct short- and long-term memory stores.

Mechanisms Involved in Short-Term Store. We now turn to a discussion of some of the mechanisms involved in the short-term store. The purpose of this section is not to review the extensive literature on short-term memory, but rather to describe a few experiments which have been important in providing a basis for our model. The first study in this category is that of Peterson and Peterson (1959). In their experiment subjects attempted to recall a single trigram of three consonants after intervals of 3, 6, 9, 12, 15 and 18 seconds. The trigram, presented auditorily, was followed immediately by a number, and the subject was instructed to count backwards by three's from that number until he received a cue to recall the trigram. The probability of a correct answer was nearly perfect at 3 seconds, then dropped off rapidly and seemed to reach an asymptote of about .08 at 15 to 18 seconds. Under the assumption that the arithmetic task played the role of preventing rehearsal and had no direct interfering effect, it may be concluded that a consonant trigram decays from short-term store within a period of about 15 seconds. In terms of the model, the following events are assumed to occur in this situation: the consonant trigram enters the visual register and is at once transferred to the a-v-l short-term store where an attempt is made to code or otherwise "memorize" the item. Such attempts terminate when attention is given to the task of counting backwards. In this initial period a trace of some sort is built up in long-term store and it is this long-term trace which accounts for the .08 probability correct at long intervals. Although discussion of the long-term system will come later, one point should be noted in this context; namely, that the long-term trace should be more powerful the more repetitions of the trigram before arithmetic, or the longer the time before arithmetic. These effects were found by Hellyer (1962); that is, the model predicts the probability correct curve will reach an asymptote that reflects long-term strength, and in the aforementioned experiment, the more repetitions before arithmetic, the higher the asymptote.

It should be noted that these findings tie in nicely with the results from a similar experiment that Milner (1967) carried out on her patients. Stimuli that could not be easily coded verbally were used; for example, clicks, light flashes, and nonsense figures. Five values were assigned to each stimulus; a test consisted of presenting a particular value of one stimulus, followed by a distracting task, followed by another value of the stimulus. The subject was required to state whether the two stimuli were the same or different. The patient with the most complete memory deficit was performing at a chance level after

60 seconds, whether or not a distracting task was given. In terms of the model, the reduction to chance level is due to the lack of a long-term store. That the reduction occurred even without a distracting task indicates that the patient could not readily verbalize the stimuli, and that rehearsal in modes other than the verbal one was either not possible or of no value. From this view, the better asymptotic performance demonstrated by normal subjects on the same tasks (with or without distraction) would be attributed to a long-term trace. At the moment, however, the conclusion that rehearsal is lacking in non-verbal modes can only be considered a highly tentative hypothesis.

We next ask whether or not there are short-term stores other than in the a-v-l mode, and if so, whether they have a comparable structure. A natural approach to this problem would use stimuli in different sense modalities and compare the decay curves found with or without a distracting task. If there was reason to believe that the subjects were not verbally encoding the stimuli, and if a relatively fast decay curve was found, then there would be evidence for a short-term memory in that modality. Furthermore, any difference between the control group and the group with a distracting task should indicate the existence of a rehearsal mechanism. Posner (1966) has undertaken several experiments of this sort. In one experiment the subject saw the position of a circle on a 180 millimeter line and later had to reproduce it; in another the subject moved a lever in a covered box a certain distance with only kinesthetic feedback and later tried to reproduce it. In both cases, testing was performed at 0, 5, 10, and 20 seconds; the interval was filled with either rest, or one of three intervening tasks of varying difficulty.

These tasks in order of increasing difficulty consisted of reading numbers, adding numbers, and classifying numbers into categories. For the kinesthetic task there was a decline in performance over 30 seconds, but with no obvious differences among the different intervening conditions. This could be taken as evidence for a short-term kinesthetic memory without a rehearsal capability. For the visual task, on the other hand, there was a decline in performance over the 30 seconds only for the two most difficult intervening tasks; performance was essentially constant over time for the other conditions. One possibility, difficult to rule out, is that the subjects' performance was based on a verbal encoding of the visual stimulus. Posner tends to doubt this possibility for reasons that include the accuracy of the performance. Another possibility is that there is a short-term visual memory with a rehearsal component; this hypothesis seems somewhat at variance with the results from Milner's patient who performed at chance level in the experiment cited above. Inasmuch as the data reported by Posner (1966) seem to be rather variable, it would probably be best to hold off a decision on the question of rehearsal capability until further evidence is in.

<u>Characteristics of the a-v-l Short-Term Store</u>. We restrict ourselves in the remainder of this section to a discussion of the characteristics of the a-v-l short-term store. Work by Conrad (1964) is particularly interesting in this regard. He showed that confusions among visually presented letters in a short-term memory task are highly correlated with the confusions that subjects make when the same letters are read aloud in a noise background; that is, the letters most confused

are those sounding alike. This might suggest an auditory short-term store, essentially the auditory portion of what has been called to this point an a-v-l store. In fact, it is very difficult to separate the verbal and linguistic aspects from the auditory ones. Hintzman (1965, 1967) has argued that the confusions are based upon similar kinesthetic feedback patterns during subvocal rehearsal. When subjects were given white noise on certain trials several could be heard rehearing the items aloud, suggesting subvocal rehearsal as the usual process. In addition, Hintzman found that confusions were based upon both the voicing qualities of the letters and the place of articulation. The place-of-articulation errors indicate confusion in kinesthetic feedback, rather than in hearing. Nevertheless, the errors found cannot be definitely assigned to a verbal rather than an auditory cause until the range of auditory confusions is examined more thoroughly. This discussion should make it clear that it is difficult to distinguish between the verbal, auditory, and linguistic aspects of short-term memory; for the purposes of this paper, then, we group the three together into one short-term memory, which we have called the a-v-l short-term store. This store will henceforth be labeled STS. (Restricting the term STS to the a-v-l mode does not imply that there are not other short-term memories with similar properties.)

The notation system should be made clear at this point. As just noted, STS refers to the auditory-verbal-linguistic short-term store. LTS will refer to the comparable memory in long-term store. It is important not to confuse our theoretical constructs STS and LTS (or the more general terms short-term store and long-term store) with the terms short-term memory (STM) and long-term memory (LTM) used in much of the psychological literature. These latter terms have come to take on an operational definition in the literature; STM refers to the memory examined in experiments with short durations or single trials, and LTM to the memory examined in long-duration experiments, typically list learning, or multiple-list learning experiments. According to our general theory, both STS and LTS are active in both STM and LTM experiments. It is important to keep these terms clear lest confusion results. For example, the Keppel and Underwood (1962) finding that performance in the Peterson situation is better on the first trials of a session has been appropriately interpreted as evidence for proactive interference in short-term memory (STM). The model we propose, however, attributes the effect to changes in the long-term store over the session, hence placing the cause in LTS and not STS.

At this point a finished model would set forth the structural characteristics of STS. Unfortunately, despite a large and growing body of experiments concerned with short-term memory, our knowledge about its structure is very limited. Control processes and structural features are so complexly interrelated that it is difficult to isolate those aspects of the data that are due solely to the structure of the memory system. Consequently, this paper presumes only a minimal structure for STS; we assume a trace in STS with auditory or verbal components which decays fairly rapidly in the absence of rehearsal, perhaps within 30 seconds. A few of the more promising possibilities concerning the precise nature of the trace will be considered next. Because most workers in this area make no particular distinction between traces in the two systems, the comments to follow are relevant to the memory trace in the long-term as well as the short-term store.

Bower (1967) has made a significant exploration of the nature of the trace. In his paper, he has demonstrated the usefulness of models based on the assumption that the memory trace consists of a number of pieces of information (possibly redundant, correlated, or in error, as the case may be), and that the information ensemble may be construed as a multi-component vector. While Bower makes a strong case for such a viewpoint, the details are too lengthy to review here. A somewhat different approach has been proposed by Wickelgren and Norman (1966) who view the trace as a unidimensional strength measure varying over time. They demonstrate that such a model fits the results of certain types of recognition-memory experiments if the appropriate decay and retrieval assumptions are made. A third approach is based upon a phenomenon reported by Murdock (1966), which has been given a theoretical analysis by Bernbach (1967). Using methods derived from the theory of signal detectability, Bernbach found that there was an all-or-none aspect to the confidence ratings that subjects gave regarding the correctness of their response. The confidence ratings indicated that an answer was either "correct" or "in error" as far as the subject could tell; if intermediate trace strengths existed, the subject was not able to distinguish between them. The locus of this all-or-none feature, however, may lie in the retrieval process rather than in the trace; that is, even if trace strengths vary, the result of a retrieval attempt might always be one of two distinct outcomes: a success or a failure. Thus, one cannot rule out models that assume varying trace strengths. Our preference is to consider the trace as a multicomponent array of information (which we shall often represent in experimental models by a unidimensional strength measure), and reserve judgment on the locus of the all-or-none aspect revealed by an analysis of confidence ratings.

There are two experimental procedures which might be expected to shed some light on the decay characteristics of STS and both depend upon controlling rehearsal; one is similar to the Peterson paradigm in which rehearsal is controlled by an intervening activity and the other involves a very rapid presentation of items followed by an immediate test. An example of the former procedure is Posner's (1966) experiment in which the difficulty of the intervening activity was varied. He found that as the difficulty of an intervening task increased, accuracy of recall decreased.

Although this result might be regarded as evidence that decay from STS is affected by the kind of intervening activity, an alternative hypothesis would ascribe the result to a reduction in rehearsal with more difficult intervening tasks. It would be desirable to measure STS decay when rehearsal is completely eliminated, but it has proved difficult to establish how much rehearsal takes place during various intervening tasks.

Similar problems arise when attempts are made to control rehearsal by increasing presentation rates. Even at the fastest conceivable presentation rates subjects can rehearse during presentation if they attend to only a portion of the incoming items. In general, experiments manipulating presentation rate have not proved of value in determining decay characteristics for STS, primarily because of the control processes the subject brings into play. Thus Waugh and Norman (1965) found no difference between 1-second and 4-second rates in their probe digit experiment; Conrad and Hille (1958) found improvement with faster rates; and Buschke (1967) found increases in the amount of primacy in his missing-span serial position curves as input rate increased from 1 item per second to 4 items per second. Complex results of this sort make it difficult to determine the structural decay characteristics of STS. Eventually, models that include the control processes involved in these situations should help clarify the STS structure.

<u>Transfer from STS to LTS</u>. The amount and form of information transferred from STS to LTS is primarily a function of control processes. We will assume, however, that transfer itself is an unvarying feature of the system; throughout the period that information resides in the short-term store, transfer takes place to long-term store. Support for such an assumption is given by studies on incidental learning which indicate that learning takes place even when the subject is not trying to store material in the long-term store. Better examples may be the experiments reported by Hebb (1961) and Melton (1963). In these experiments subjects had to repeat sequences of digits. If a particular sequence was presented every several trials, it was gradually learned. It may be assumed that subjects in this situation attempt to perform solely by rehearsal of the sequence within STS; nevertheless, transfer to LTS clearly takes place. This Hebb-Melton procedure is currently being used to explore transfer characteristics in some detail. Cohen and Johansson (in press), for example, have found that an overt response to the repeated sequence was necessary for improvement in performance

to occur in this situation; thus information transfer is accentuated by overt responses and appears to be quite weak if no response is demanded.

The form of the STS-LTS transfer may be probabilistic, continuous, or some combination; neither the literature nor our own data provide a firm basis for making a decision. Often the form of the information to be remembered and the type of test used may dictate a particular transfer process, as for example in Bower's (1961) research on an all-or-none paired-associate learning model, but the issue is nevertheless far from settled. In fact, the changes in the transfer process induced by the subject effectively alter the transfer function from experiment to experiment, making a search for a universal, unchanging process unproductive.

2.3. Long-Term Store.

Because it is easiest to test for recall in the a-v-l mode, this part of long-term store has been the most extensively studied. It is clear, however, that long-term memory exists in each of the sensory modalities; this is shown by subjects' recognition capability for smells, taste, and so on. Other long-term information may be stored which is not necessarily related to any of the sensory modalities. Yntema and Trask (1963), for example, have proposed that temporal memory is stored in the form of "time-tags." One again, however, lack of data forces us to restrict our attention primarily to the a-v-l mode, which we have designated LTS.

First a number of possible formulations of the LTS trace will be considered. The simplest hypothesis is to assume that the trace is all-or-none; if a trace is placed in memory then a correct retrieval and response will occur. Second-guessing experiments provide evidence concerning an hypothesis of this sort.

Binford and Gettys (1965) presented the subject with a number of alternatives, one of which was the correct answer. If his first response is incorrect, he picks again from the remaining alternatives. The results indicate that second guesses are well above the chance level to be expected if the subject were guessing randomly from the remaining alternatives. This result rules out the simple trace model described above because an all-or-none trace would predict second guesses to be at the chance level. Actually, the above model was a model of both the form of the trace and the type of retrieval. We can expand the retrieval hypothesis and still leave open the possibility of an all-or-none trace. For example, in searching for a correct all-or-none trace in LTS, the subject might find a similar but different trace and mistakenly terminate the search and generate an answer; upon being told that the answer is wrong the subject renews the search and may find the correct trace the next time. Given this hypothesis, it would be instructive to know whether the results differ if the subject must rank the response alternatives without being given feedback after each choice. In this case all the alternatives would be ranked on the basis of the same search of LTS; if the response ranked second was still above chance then it would become difficult to defend an all-or-none trace.

A second source of information about the nature of the trace comes from the tip-of-the-tongue phenomenon examined by Hart (1965), Brown and McNeill (1966), and Freedman and Landauer (1966). This phenomenon refers to a person's ability to predict accurately that he will be able to recognize a correct answer even though he cannot recall it at the moment. He feels as if the correct answer were on the "tip of the tongue." Experiments have shown that if subjects who cannot recall an answer are asked to estimate whether they will be able to choose the correct answer from a set of alternatives, they often show good accuracy in predicting their success in recognition. One explanation might be that the subject recalls some information, but not enough to generate an answer and feels that this partial information is likely to be sufficient to choose among a set of alternatives. Indeed, Brown and McNeill found that the initial sound of the word to be retrieved was often correctly recalled in cases where a correct identification was later made. On the other hand, the subject often is absolutely certain upon seeing the correct response that it is indeed correct. This might indicate that some new, relevant information has become available after recognition. In any case, a simple trace model can probably not handle these results. A class of models for the trace which can explain the tip-of-the-tongue phenomenon are the multiple-copy models suggested by Atkinson and Shiffrin (1965). In these schemes there are many traces or copies of information laid in long-term store, each of which may be either partial or complete. In a particular search of LTS perhaps only a small number or just one of these copies is retrieved, none complete enough to generate the correct answer; upon recognition, however, access is gained to the other copies, presumably through some associative process. Some of these other copies contain enough information to make the subject certain of his choice. These multiple-copy memory models are described more fully in Atkinson and Shiffrin (1965). Bernbach (1967) has successfully applied a model of this type to a variety of data.

The decay and/or interference characteristics of LTS have been studied more intensively over the last 50 years than any other aspect of memory. Partly for this reason a considerable body of theory has been advanced known as interference theory.* We tend to regard this theory as descriptive rather than explanatory; this statement is not meant to detract from the value of the theory as a whole, but to indicate that a search for mechanisms at a deeper level might prove to be of value. Thus, for example, if the interfering effect of a previously learned list upon recall of a second list increases over time until the second list is retested, it is not enough to accept "proactive interference increasing over time" as an explanation of the effect; rather one should look for the underlying search, storage, and retrieval mechanisms responsible.

We are going to use a very restricted definition of interference in the rest of this paper, interference will be considered a structural feature of memory not under the control of the subject. It will refer to such possibilities as disruption and loss of information. On the other hand, there are search mechanisms which generate effects like those of structural interference, but which are control processes. Interference theory, of course, includes both types of possibilities, but we prefer to break down interference effects into those which are structurally based, and those under the control of the subject. Therefore

* For an overview of interference theory see Postman (1961).

the term interference is used henceforth to designate a structural feature of the long-term system.

It is important to realize that often it is possible to explain a given phenomena with either interference or search notions. Although both factors will usually be present, the experimental situation sometimes indicates which is more important. For example, as we shall see in Section 5, the decrease in the percentage of words recalled in a free verbal-recall experiment with increases in list length could be due either to interference between items or to a search of decreasing effectiveness as the number of items increase. The typical free recall situation, however, forces the subject to engage in a search of memory at test and indicates to us that the search process is the major factor. Finally, note that the interference effect itself may take many forms and arise in a number of ways. Information within a trace may be destroyed, replaced, or lessened in value by subsequent information. Alternatively, information may never be destroyed but may become irretrievable, temporarily or permanently.

In this section an attempt has been made to establish a reasonable basis for at least three systems -- the sensory register, the short-term store, and the long-term store; to indicate the transfer characteristics between the various stores; and to consider possible decay and interference functions within each store.

SECTION 3: CONTROL PROCESSES IN MEMORY

The term "control process" refers to those processes that are not permanent features of memory, but are instead transient phenomena under the control of the subject; their appearance depends on such factors as instructional set, the experimental task, and the past history of the subject. A simple example of a control process can be demonstrated in a paired-associate learning task involving a list of stimuli each paired with either an A or B response (Bower, 1961). The subject may try to learn each stimulus-response pair as a separate, integral unit or he may adopt the more efficient strategy of answering B to any item not remembered and attempting to remember only the stimuli paired with the A response. This latter scheme will yield a radically different pattern of performance than the former; it exemplifies one rather limited control process. The various rehearsal strategies, on the other hand, are examples of control processes with almost universal applicability.

Since subject-controlled memory processes include any schemes, coding techniques, or mnemonics used by the subject in his effort to remember, their variety is virtually unlimited and classification becomes difficult. Such classification as is possible arises because these processes, while under the voluntary control of the subject, are nevertheless dependent upon the permanent memory structures described in the previous section. This section therefore will follow the format of Section 2, organizing the control processes into those primarily associated with the sensory register, STS, and LTS. Apart from this, the presentation will be somewhat fragmentary, drawing upon examples from many disparate experiments in an attempt to emphasize the variety, pervasiveness, and importance of the subject-controlled processes.

3.1. Control Processes in the Sensory Register

Because a large amount of information enters the sensory register and then decays very quickly, the primary function of control processes at this level is the selection of particular portions of this information for transfer to the short-term store. The first decision the subject must make concerns which sensory register to attend to. Thus, in experiments with simultaneous inputs from several sensory channels, the subject can readily report information from a given sense modality if so instructed in advance, but his accuracy is greatly reduced if instructions are delayed until after presentation. A related attention process is the transfer to STS of a selected portion of a large information display within a sensory modality. An example to keep in mind here is the scanning process in the visual registration system. Letters in a tachistiscopically presented display may be scanned at a rate of about 10 milliseconds a letter, the form of the scan being under the control of the subject. Sperling (1960) found the following result. When the signal identifying which row to report from a matrix of letters was delayed for an interval of time following stimulus offset, the subjects developed two observing strategies. One strategy consisted of obeying the experimenter's instructions to pay equal attention to all rows; this strategy resulted in evenly distributed errors and quite poor performance at long delays. The other strategy consisted of anticipating which row would be tested and attending to only that row; in this case the error variance is increased but performance is better at longer delay intervals than for the other strategy. The subjects were aware of, and reported using, these strategies. For example, one experienced subject reported switching from the first to the second strategy in an effort to maximize performance when the delay between presentation and report rose above .15 seconds. The graph of his probability of a correct response plotted against delay interval, while generally decreasing with delay, showed a dip of about .15 seconds indicating that he did not switch strategies soon enough for optimal performance.

The decisions as to which sensory register to attend to, and where and what to scan within the system, are not the only choices that must be made at this level. There are a number of strategies available to the subject for matching information in the register against the long-term store and thereby identifying the input. In an experiment by Estes and Taylor (1966) for example, the subject had to decide whether an F or B was embedded in a matrix display of letters. One strategy would have the subject scan the letters in order, generating the "name" of each letter and checking to see whether it is a B or an F. If the scan ends before all letters are processed, and no B or F has been found, the subject would presumably guess according to some bias. Another strategy might have the subject do a features match on each letter against B and then F, moving on as soon as a difference is found; in this strategy it would not be necessary to scan all features of each letter (i.e., it would not be necessary to generate the name of each letter). A third strategy might have the subject compare with only one of the crucial letters, guessing the other if a match is not found by the time the scan terminates.

3.2. Control Processes in Short-Term Store

Storage, Search and Retrieval Strategies. Search processes in STS, while not as elaborate as those in LTS because of the smaller amount of information in STS through which the search must take place, are nevertheless important. Since information in STS in excess of the rehearsal capability is decaying at a rapid rate, a search for a particular datum must be performed quickly and efficiently. One indirect method of examining the search process consists of comparing the results of recognition and recall experiments in which STS plays the major role. Presumably there is a search component in the recall situation that is absent in the recognition situation. It is difficult to come to strong conclusions on this basis, but recognition studies such as Norman and Wickelgren (1966) have usually given rise to less complicated models than comparable recall experiments, indicating that the search component in STS might be playing a large role.

One result indicating that the STS search occurs along ordered dimensions is based upon binaural stimulus presentation (Broadbent, 1954, 1956, 1958). A pair of items is presented, one to each ear simultaneously. Three such pairs are given, one every half second. Subjects perform best if asked to report the items first from one ear and then the other, rather than, say, in pairs. While Broadbent interprets these results in terms of a postulated time needed to switch attention from one ear to the other (a control process in itself), other interpretations are possible. In particular, part of the information stored with each item might include which ear was used for input. This information might then provide a simple dimension along which to search STS and report during recall. Another related possibility would have the subject group the items along this dimension during presentation. In any case we would expect similar results if another dimension other than "sides" (which ear) were provided. Yntema and Trask (1963) used three word-number pairs presented sequentially, one every half second; one member of a pair was presented to one ear and the other member to the other ear. There were three conditions: the first in which three words were presented consecutively on one side (and therefore the three numbers on the other), the second in which two words and one number were presented consecutively on one side, the third in which a number separated the two words on one side. Three test conditions were used: the subject was asked to report words, then numbers (types); or to report one ear followed by the other (sides); or the simultaneous pairs in order (pairs). The results are easy to describe. In terms of probability correct, presentation condition one was best, condition two next, and condition three worst. For the test conditions "types" yielded the highest probability of correct response, followed by "sides" and then "pairs." "Sides" being better than "pairs" was one of the results found by Broadbent, but "types" being even better than "sides" suggests that the organization along available dimensions, with the concomitant increase of efficiency in the search process, is the dominant factor in the situation.

One difficulty in studying the search process in STS is the fact that the subject will perform perfectly if the number of items presented is within his rehearsal span. Sternberg (1966) has overcome this difficulty by examining the latency of responses within the rehearsal span. His typical experiment consists of presenting from one to six digits to the subject at the rate of 1.2 seconds each. Following a 2-second delay, a single digit is presented and the subjects must respond "yes" or "no" depending on whether or not the test digit was a member of the set just presented. Following this response the subject is required to recall the complete set in order. Since the subjects were 98.7 percent correct on the recognition test and 98.6 percent correct on the recall test it may be assumed that the task was within their rehearsal span. Interesting results were found in the latencies of the recognition responses: there was a linear increase in latency as the set size increased from 1 to 6 digits. The fact that there was no difference in latencies for "yes" versus "no" responses indicates that the search process in this situation is exhaustive and does not terminate the moment a match is found. Sternberg concludes that the subject engages in an exhaustive serial comparison process which evaluates elements at the rate of 25 to 30 per second. The high processing rate makes it seem likely that the rehearsal the subjects report is not an integral part of the scanning process, but instead maintains the image in STS so that it may be scanned at the time of the test. This conclusion depends upon accepting as a reasonable rehearsal rate for digits the values reported by Landauer (1962) which were never higher than 6 per second.

Buschke's (1963) missing-span method provides additional insight into search and retrieval processes in STS. The missing-span procedure consists of presenting in a random order all but one of a previously specified set of digits; the subject is then asked to report the missing digit. This technique eliminates the output interference associated with the usual digit-span studies in which the entire presented set must be reported. Buschke found that subjects had superior performance on a missing-span task as compared with an identical digit-span task in which all of the presented items were to be reported in any order. A natural hypothesis would explain the difference in performance as being caused by output interference; that is, the multiple recalls in the digit-span procedure produce interference not seen in the single test procedure of the missing-span. An alternative explanation would hold that different storage and search strategies were being employed in the two situations. Madsen and Drucker (1966) examined this question by comparing test instructions given just prior to or immediately following each presentation sequence; the instructions specify whether the subject is to report the set of presented digits or simply to report the missing digit. Output interference would imply that the difference between missing-span and digit-span would hold up in both cases. The results showed that the missing-span procedure with prior instructions was superior to both missing-span and digit-span with instructions following presentation; the latter two conditions produced equal results and were superior to digit-span with prior instructions. It seems clear, then, that two storage and search strategies are being used: a missing-span type, and a digit-span type. Prior instructions (specifying the form of the subject's report) lead the subject to use one or the other of these strategies, but instructions following presentation are associated with a mixture of the two strategies. It appeared in this case that the strategies differed in terms of the type of storage during presentation; the digit-span group with prior instructions tended to report their digits in their presentation order, while the digit-span group with instructions after presentation more often reported the digits in their numerical order. This indicates that the missing-span strategy involved checking off the numbers as they were presented against a fixed, numerically-ordered list, while the digit-span strategy involved rehearsing the items in their presented order. It is interesting to note that if the subjects had been aware of the superiority of the missing-span strategy, they could have used it in the digit-span task also, since the two types of tests called for the same information.

It should be noted that retrieval from STS depends upon a number of factors, some under the control of the subject and some depending upon the decay characteristics of STS. If the decay is partial in some sense, so that the trace contains only part of the information necessary for direct output, then the problem arises of how the partial information should be used to generate a response. In this case, it would be expected that the subject would then engage in a search of LTS in an effort to match or recognize the partial information. On the other hand, even though traces may decay in a partial manner, the rehearsal capability can hold a select set of items in a state of immediate recall availability and thereby impart to these items what is essentially an all-or-none status. It is to this rehearsal process that we now turn.

Rehearsal Processes. Rehearsal is one of the most important factors in experiments on human memory. This is particularly true in the laboratory because the concentrated, often meaningless, memory tasks used increase the relative efficacy of rehearsal as compared with the longer term coding and associative processes. Rehearsal may be less pervasive in everyday memory, but nevertheless has many uses, as Broadbent (1958) and others have pointed out. Such examples as remembering a telephone number or table-tennis score serve to illustrate the primary purpose of rehearsal, the lengthening of the time period information stays in the short-term store. A second purpose of rehearsal is illustrated by the fact that even if one wishes to remember a telephone number permanently, one will often rehearse the number several times. This rehearsal serves the purpose of increasing the strength built up in a long-term store, both by increasing the length of stay in STS (during which time a trace is built up in LTS) and by giving coding and other storage processes time to operate. Indeed, almost any kind of operation on an array of information (such as coding) can be viewed as a form of rehearsal, but this paper reserves the term only for the duration-lengthening repetition process.

In terms of STS structure, we can imagine that each rehearsal regenerates the STS trace and thereby prolongs the decay. This does not imply that the entire information ensemble available in STS immediately after presentation is regenerated and maintained at each rehearsal. Only that information selected by the subject, often a small proportion of the initial ensemble, is maintained. If the word "cow" is presented, for example, the sound of the word cow will enter STS; in addition,associates of cow, like milk, may be retrieved from LTS and also entered in STS; furthermore, an image of a cow may be entered into a short-term visual store. In succeeding rehearsals, however, the subject may rehearse only the word "cow" and the initial associates will decay and be lost. The process may be similar to the loss of meaningfulness that occurs when a word is repeated over and over (Lambert and Jakobovitz, 1960).

An interesting question concerns the maximum number of items that can be maintained via rehearsal. This number will depend upon the rate of STS decay and the form of the trace regenerated in STS by rehearsal. With almost any reasonable assumptions about either of these processes, however, an ordered rehearsal will allow the greatest number of items to be maintained. To give a simple example, suppose that individual items take 1.1 seconds to decay and may be restarted if rehearsal begins before decay is complete. Suppose further that each rehearsal takes .25 seconds. It is then clear that 5 items may be maintained indefinitely if they are rehearsed in a fixed order over and over. On the other hand, a rehearsal scheme in which items are chosen for rehearsal on a random basis will quickly result in one or more items decaying and becoming lost. It would be expected, therefore, that in situations where subjects are relying primarily upon their rehearsal capability in STS, rehearsal will take place in an ordered fashion. One such situation, from which we can derive an estimate of rehearsal capability, is the digit-span task. A series of numbers is read to the subject who is then required to recall them, usually in the forward or backward order. Because the subject has a long-term store which sometimes can be used to supplement the short-term rehearsal memory, the length of a series which can be correctly recalled may exceed the rehearsal capacity. A lower limit on this capacity can be found by identifying the series length at which a subject never errs; this series length is usually in the range of 5 to 8 numbers.*

*Wickelgren (1965) has examined rehearsal in the digit-span task in greater detail and found that rehearsal capacity is a function of the groupings engaged in by the subject; in particular, rehearsal in distinct groups of three was superior to rehearsal in fours and fives.

The above estimates of rehearsal capability are obtained in a discrete-trial situation where the requirement is to remember every item of a small input. A very similar rehearsal strategy can be employed, however, in situations such as free recall where a much greater number of items is input than rehearsal can possibly encompass. One strategy in this case would be to replace one of the items currently being rehearsed by each new item input. In this case every item would receive at least some rehearsal. Because of input and reorganization factors, which undoubtedly consume some time, the rehearsal capacity would probably be reduced. It should be clear that under this scheme a constant number of items will be undergoing rehearsal at any one moment. As an analogy, one might think of a bin always containing exactly n items; each new item enters the bin and knocks out an item already there. This process has been called in earlier reports a "rehearsal buffer," or simply a "buffer," and we will use this terminology here. (Atkinson and Shiffrin, 1965).

In our view the maintainence and use of the buffer is a process entirely under the control of the subject. Presumably a buffer is set up and used in an attempt to maximize performance in certain situations. In setting up a maximal sized buffer, however, the subject is devoting all his effort to rehearsal and not engaging in other processes such as coding and hypothesis-testing. In situations, therefore, where coding, long-term search, hypothesis-testing and other mechanisms appreciably improve performance, it is likely that a trade-off will occur in which the buffer size will be reduced and rehearsal may even become somewhat random while coding and other strategies increase.

At this point we want to discuss various buffer operations in greater detail. Figure 2 illustrates a fixed size buffer and its relation to the rest of the memory system. The content of the buffer is constructed from items that have entered STS, items which have been input from the sensory register or from LTS. The arrow going toward LTS indicates that some long-term trace is being built up during an item's stay in the buffer. The other arrow from the buffer indicates that the input of a new item into the buffer causes an item currently in the buffer to be bumped out; this item then decays from STS and is lost (except for any trace which has accumulated in LTS during its stay). An item dropped from the buffer is likely to decay more quickly in STS than a newly presented item which has just entered STS. There are several reasons for this. For one thing, the item is probably already in some state of partial decay when dropped; in addition, the information making up an item in the buffer is likely to be only a partial copy of the ensemble present immediately following stimulus input.

There are two additional processes not shown in Figure 2 that the subject can use on appropriate occasions. First, the subject may decide not to enter every item into the buffer; the reasons are manifold. For example, the items may be presented at a very fast rate so that input and reorganization time encroach too far upon rehearsal time. Another possibility is that some combinations of items are particularly easy to rehearse, making the subject loath to break up the combination. In fact, the work involved in introducing a new item into the buffer and deleting and old one may alone give the subject incentive to keep the buffer unchanged. Judging from these remarks, the choice of which items to

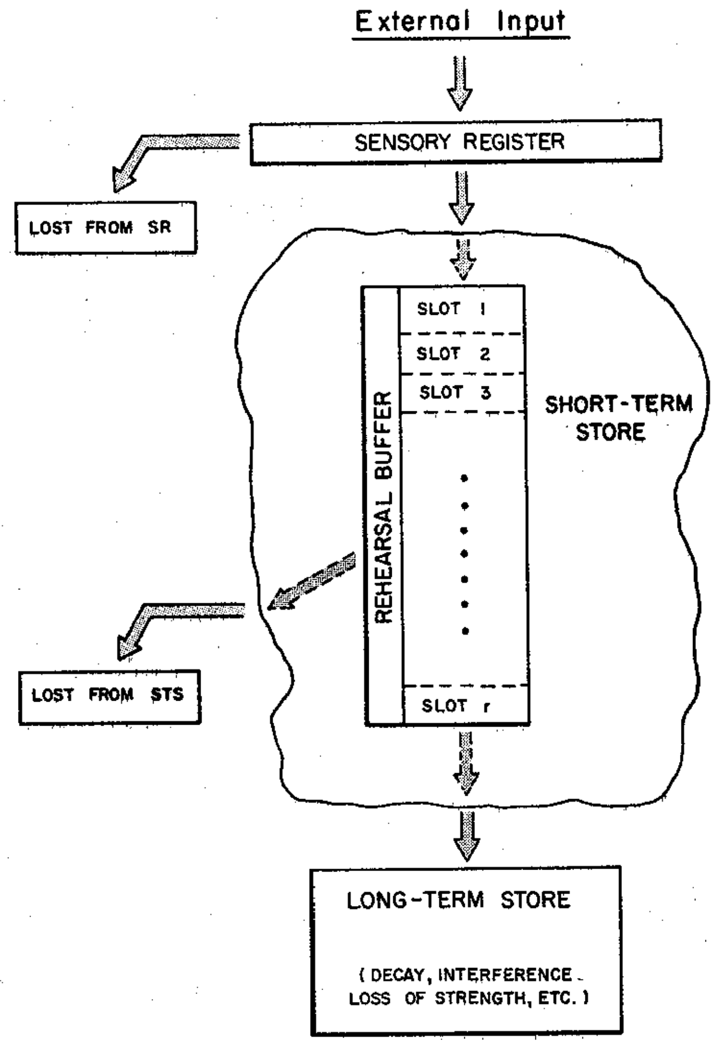


Figure 2 The rehearsal buffer and its relation to the memory system

enter into the buffer is based on momentary characteristics of the current string of input items and may appear at times to be essentially random.

The second process not diagrammed in Figure 2 is the choice of which item to eliminate from the buffer when a new item is entered. There are several possibilities. The choice could be random; it could be based upon the state of decay of the current items; it could depend upon the ease of rehearsing the various items; most important, it could be based upon the length of time the various items have resided in the buffer. It is not unreasonable that the subject should have a fairly good idea which items he has been rehearsing the longest, as he might if rehearsal takes place in a fixed order. It is for this reason that the slots or positions of the buffer have been numbered consecutively in Figure 2; that is, to indicate that the subject might have some notion of the relative recency of the various items in the buffer.

The experimental justification for these various buffer mechanisms will be presented in Section 4. It should be emphasized that the subject will use a fixed size buffer of the sort described here only in select situations, primarily those in which he feels that trading off rehearsal time for coding and other longer term control processes would not be fruitful. To the extent that long-term storage operations prove to be successful as compared with rehearsal, the structure of the rehearsal mechanism will tend to become impoverished. One other point concerning the buffer should be noted. While this paper consistently considers a fixed size short-term buffer as a rehearsal strategy of the subject, it is possible to apply a fixed-size model of a similar kind to the structure of the short-term system as a whole, that is, to consider a short-term buffer as a permanent feature of memory. Waugh and Norman (1965), for example, have done this in their paper on primary memory. The data on the structure of STS is currently so nebulous that such an hypothesis can be neither firmly supported nor rejected.

<u>Coding Processes and Transfer Between Short- and Long-Term Store</u>. It should be evident that there is a close relationship between the short- and long-term store. In general, information entering STS comes directly from LTS and only indirectly from the sensory register. For example, a visually presented word cannot be entered into STS as an auditory-verbal unit until a long-term search and match has identified the verbal representation of the visual image. For words, letters, and highly familiar stimuli, this long-term search and match process may be executed very quickly, but one can imagine unfamiliar stimuli, such as, say, a nonsense scribble, where considerable search might be necessary before a suitable verbal representation is found to enter into STS. In such cases, the subject might enter the visual image directly into his short-term visual memory and not attempt a verbal coding operation.

Transfer from STS to LTS may be considered a permanent feature of memory; any information in STS is transferred to LTS to some degree throughout its stay in the short-term store. The important aspect of this transfer, however, is the wide variance in the amount and form of the transferred information that may be induced by control processes. When the subject is concentrating upon rehearsal, the information transferred would be in a relatively weak state and easily subject to interference. On the other hand, the subject may divert his effort from rehearsal to various coding operations which will increase the strength of the stored information. In answer to the question of what is a coding process, we can most generally state that a coding process is a select alteration and/or addition to the information in the short-term store as the result of a search of the long-term store. This change may take a number of forms, often using strong pre-existing associations already in long-term store. A number of these coding possibilities will be considered later.

Experiments may be roughly classified in terms of the control operations the subject will be led to use. Concept formation problems or tasks where there is a clear solution will lead the subject to strategy selection and hypothesis-testing procedures (Restle, 1964). Experiments which do not involve problem solving, where there are a large number of easily coded items, and where there is a long period between presentation and test, will prompt the subject to expend his efforts on long-term coding operations. Finally, experiments in which memory is required, but long-term memory is not efficacious, will lead the subject to adopt rehearsal strategies that maintain the information the limited period needed for the task. Several examples of the latter experiment will be examined in this paper; they are characterized by the fact that the responses assigned to particular stimuli are continually changing, so that coding of a specific stimulus-response pair will prove harmful to succeeding pairs using the same stimulus. There are experiments, of course, for which it will not be possible to decide on a priori grounds which control processes are being used. In these cases the usual identification procedures must be used, including model fits and careful questioning of the subjects.

There are other short-term processes that do not fit easily into the above classification. They include grouping, organizing, and chunking strategies. One form that organizing may take is the selection of a subset of presented items for special attention, coding and/or rehearsal. This selection process is clearly illustrated in a series of studies on magnitude of reward by Harley (1965 a,b). Items in a paired-associate list were given two monetary incentives, one high and one low. In one experiment the subjects learned two paired-associate lists, one consisting of all high incentive items, the other consisting of all low incentive items; there were no differences in the learning rates for these lists. In a second experiment, subjects learned a list which included both high and low incentive items; in this case learning was faster for the high than the low incentive items. However, the overall rate of learning for the mixed list was about the same as for the two previous lists. It seems clear that when the high and low incentive items are mixed, the subject selectively attends to, codes and rehearses those items with the higher payoffs. A second kind of organizing that occurs is the grouping of items into small sets, often with the object of memorizing the set as a whole, rather than as individual items. Typically in this case the grouped items will have some common factor. A good example may be found in the series of studies by Battig (1966) and his colleagues. He found a tendency to group items according to difficulty and according to degree of prior learning; this tendency was found even in paired-associate tasks where an extensive effort had been made to eliminate any basis for such grouping. A third type of information organization is found in the "chunking" process suggested by

Miller (1956). In his view there is some optimal size that a set of information should have in order to best facilitate remembering. The incoming information is therefore organized into chunks of the desired magnitude.

3.3. Control Processes in Long-Term Store.

Control processes to be considered in this section fall roughly into two categories: those concerned with transfer between short-term and long-term store and those concerned with search for and retrieval of information from LTS.

Storage in Long-Term Store. It was stated earlier that some information is transferred to LTS throughout an item's stay in STS, but that its amount and form is determined by control processes. This proposition will now be examined in greater detail. First of all it would be helpful to consider a few simple examples where long-term storage is differentially affected by the coding strategy adopted. One example is found in a study on mediators performed by Montague, Adams and Kiess (1966). Pairs of nonsense syllables were presented to the subject who had to write down any natural language mediator (word, phrase, or sentence associated with a pair) which occurred to him. At test 24 hours later the subject attempted to give the response member of each pair and the natural language mediator (NLM) that had been used in acquisition. Proportion correct for items on which the NLM was retained was 70 percent, while the proportion correct was negligible for items where the NLM was forgotten or significantly changed. Taken in conjunction with earlier studies showing that a group using NLMs was superior to a group learning by rote (Runquist and Farley, 1964), this result indicates a strong dependence of recall upon natural language mediators. A somewhat different encoding technique has been examined by Clark and Bower (personal communication). Subjects were required to learn several lists of paired-associate items, where each item was a pair of familiar words. Two groups of subjects were given identical instructions, except for an extra section read to the experimental group explaining that the best method of learning the pairs was to form an elaborate visual image containing the objects designated by the two words. This experimental group was then given a few examples of the technique. There was a marked difference in performance between the groups on both immediate and delayed tests, the experimental group outperforming the control group by better than 40 percent in terms of probability correct. In fact, postexperimental questioning of the subjects revealed that the occasional high performers in the control group were often using the experimental technique even in the absence of instructions to do so. This technique of associating through the use of visual images is a very old one; it has been described, for example, by Cicero in his De oratore when he discusses memory as one of the five parts of rhetoric, and is clearly very effective.

We now consider the question of how these encoding techniques improve performance. The answer depends to a degree upon the fine structure of long-term store, and therefore cannot be stated precisely. Nevertheless, a number of possibilities should be mentioned. First, the encoding may make use of strong pre-existing associations, eliminating the necessity of making new ones. Thus in mediating a word pair in a paired-associate task, word A might elicit word A' which in turn elicits the response. This merely moves the question back a level: how does the subject know which associates are the correct ones? It may be that the appropriate associations are identified by temporal position; that is, the subject may search through the associations looking for one which has been elicited recently. Alternatively, information could be stored with the appropriate association identifying it as having been used in the current paired-associates task. Second, the encoding might greatly decrease the effective area of memory which must be searched at the time of test. A response word not encoded must be in the set of all English words, or perhaps in the set of all words presented "recently," while a code may allow a smaller search through the associates of one or two items. One could use further search-limiting techniques such as restricting the mediator to the same first letter as the stimulus. A third possibility, related to the second, is that encoding might give some order to an otherwise random search. Fourth, encoding might greatly increase the amount of information stored. Finally, and perhaps most important, the encoding might protect a fledgling association from interference by succeeding items. Thus if one encodes a particular pair through an image of, say a specific room in one's home, it is unlikely that future inputs will have any relation to that image; hence they will not interfere with it. In most cases coding probably works well for all of the above reasons.

There is another possible set of effects of the coding process which should be mentioned here. As background, we need to consider the results of several recent experiments which examine the effect of spacing between study and test in paired-associate learning (Bjork, 1966; Young, 1966). The result of primary interest to us is the decrease in probability correct as the number of other paired-associate items presented between study and test increases. This decrease seems to reach asymptote only after a fairly large number (e.g., 20) of intervening items. There are several possible explanations for this "short-term" effect. Although the effect probably occurs over too great an interval to consider direct decay from STS as an explanation, any of several rehearsal strategies could give rise to an appropriate looking curve. Since a paired-associate task usually requires coding, a fixed-size rehearsal buffer may not be a reasonable hypothesis, unless the buffer size is fairly small; on the other hand, a variable rehearsal set with semi-randomly spaced rehearsals may be both reasonable and accurate. If, on the other hand, one decides that almost no continuing rehearsal occurs in this task, what other hypotheses are available? One could appeal to retroactive interference but this does little more than name the phenomenon. Greeno (1967) has proposed a coding model which can explain the effect. In his view, the subject may select one of several possible codes at the time of study. In particular, he might select a "permanent" code, which will not be disturbed by any other items or codes in the experiment; if this occurs, the item is said to be learned. On the other hand, a "transitory" code might be selected, one which is disturbed or eliminated as succeeding items are presented. This transitory code will last for a probabilistically determined number of trials before becoming useless or lost. The important point to note here is the fact that a decreasing "short-term" effect can occur as a result of solely long-term operations. In experiments emphasizing long-term coding, therefore, the decision concerning which decay process, or combination of decay processes, is operative will not be easy to make in an a priori manner; rather the decision would have to be based upon such a postiori grounds as goodness-of-fit results for a particular model and introspective reports from the subject.

Long-Term Search Processes. One of the most fascinating features of memory is the long-term search process. We have all, at one time or another, been asked for information which we once knew, but which is now momentarily unavailable, and we are aware of the ensuing period (often lasting for hours) during which memory was searched, occasionally resulting in the correct answer. Nevertheless, there has been a marked lack of experimental work dealing with this rather common phenomenon. For this reason, our discussion of search processes will be primarily theoretical, but the absence of a large experimental literature should not lead us to underestimate the importance of the search mechanism.

The primary component of the search process is locating the sought-for trace (or one of the traces) in long-term store. This process is seen in operation via several examples. The occasionally very long latencies prior to a correct response for well-known information indicates a non-perfect search. A subject reporting that he will think "of it the moment he thinks about something else" indicates a prior fixation on an unsuccessful search procedure. Similarly the tip-of-the-tongue phenomenon mentioned earlier indicates a failure to find an otherwise very strong trace. We have also observed the following while quizzing a graduate student on the names of state capitals. The student gave up trying to remember the capital of the state of Washington after pondering for a long time. Later this student quickly identified the capital of Oregon as Salem and then said at once that the capital of Washington was Olympia. When asked how he suddenly remembered, he replied that he had learned the two capitals together. Presumably this information would have been available during the first search if the student had known where to look: namely in conjunction with the capital of Oregon. Such descriptive examples are numerous and serve to indicate that a search can sometimes fail to uncover a very strong trace. One of the decisions the subject must make is when to terminate an unsuccessful search. An important determiner of the length of search is the amount of order imposed during the search; if one is asked to name all the states and does so strictly geographically, one is likely to do better than someone who spews out names in a haphazard fashion. The person naming states in a haphazard fashion will presently encounter in his search for new names those which he has already given; if this occurs repeatedly, the search will be terminated as being unfruitful. The problem of terminating the search is especially acute in the case of recalling a set of items without a good natural ordering. Such a case is found in free-verbal-recall experiments in which a list of words is presented to the subject who must then recall as many as possible. The subject presumably searches along some sort of temporal dimension, a dimension which lets the subject know when he finds a word whether or not it was on the list presented most recently. The temporal ordering is by no means perfect, however, and the search must therefore be carried out with a degree of randomness. This procedure may lead to missing an item which has a fairly strong trace. It has been found in free-verbal-recall experiments, for example, that repeated recall tests on a given list sometimes result in the inclusion on the second test of items left out on the first test. In our own experiments we have even observed intrusions from an earlier list that had not been recalled during the test of that list.

It would be illustrative at this point to consider an experiment carried out by Norma Graham at Stanford University. Subjects were asked to name the capitals of the states. If a correct answer was not given within 5 seconds following presentation of the state name, the subjects were then given a hint and allowed 30 seconds more to search their memory. The hint consisted of either 1, 2, 4, 12, or 24 consecutive letters of the alphabet, one of which was the first letter in the name of the state capital. The probability correct dropped steadily as the hint size increased from 1 to 24 letters. The average response latencies for correct answers, however, showed a different effect; the one-letter hint was associated with the fastest response time, the two-letter hint was slower, the four-letter hint was slower yet, but the 12- and 24-letter hints were faster than the four-letter hint. One simple hypothesis that can explain why latencies were slower after the four-letter hint than after the 12- and 24-letter hints depends upon differing search processes. Suppose the subject in the absence of a hint engages in "normal" search, or N-search. When given the first letter, however, we will assume the subject switches to a first letter search, or L-search, consisting of a deeper exploration of memory based upon the first letter. This L-search might consist of forming possible sounds beginning with the appropriate letter, and matching them against possible city names. When the size of the hint increases, the subject must apply the L-search to each of the letters in turn, obviously a time consuming procedure. In fact, for twelve or twenty-four letter hints the probability is high that the subject would use up the entire thirty-second search period without carrying out an L-search on the correct first letter. Clearly a stage is reached, in terms of hint size, where the subject will switch from an L-search to N-search in order to maximize performance. In the present experiment it seems clear that the switch in strategy occurred between the 4- and 12-letter hints.

In the above experiment there were two search-stopping events, one subject controlled and the other determined by the thirty-second time limit. It is instructive to consider some of the possible subject-controlled stopping rules. One possibility is simply an internal time limit, beyond which the subject decides further search is useless. Related to this would be an event-counter stopping rule that would halt the subject when a fixed number of pre-specified events had occurred. The events could be total number of distinct "searches," total number of incorrect traces found, and so on. A third possibility is dependent on a consecutive-events counter. For example, search could be stopped whenever x consecutive searches recovered traces that had been found in previous searches.

It was noted earlier that searches may vary in their apparent orderliness. Since long-term memory is extremely large, any truly random search would invariably be doomed to failure. The search must always be made along some dimension, or on the basis of some available cues. Nevertheless searches do vary in their degree of order; a letter by letter search is highly structured, whereas a free associative search that proceeds from point to point in a seemingly arbitrary manner will be considerably less restrained, even to the point where the same ground may be covered many times. One other possible feature of the search process is not as desirable as the ones previously mentioned. The search itself might prove destructive to the sought after trace. That is, just as new information transferred to the long-term store might interfere with previous material stored there, the generation of traces during the search might prove to have a similar interfering effect.

A somewhat different perspective on search procedures is obtained by considering the types of experimental tests that typically are used. Sometimes the very nature of the task presumes a specific search procedure. An example is found in the free-verbal-recall task in which the subject must identify a subset of a larger well-learned group of words. A search of smaller scope is made in a paired-associate task; when the set of possible responses is large, the search for the answer is similar to that made in free recall, with a search component and a recognition component to identify the recovered trace as the appropriate one. When the set of responses in a paired-associate task is quite small, the task becomes one of recognition alone: the subject can generate each possible response in order and perform a recognition test on each. The recognition test presumably probes the trace for information identifying it as being from the correct list and being associated with the correct stimulus.

It was said that the primary component of the search process is locating the desired memory trace in LTS. The secondary component is the recovery of the trace once found. It has been more or less assumed for simplicity in the above discussions that the trace is all-or-none. This may not be the case, and the result of a search might be the recovery of a partial trace. Retrieval would then depend either upon correctly guessing the missing information or performing a further search to match the partial trace with known responses. It is possible, therefore, to divide the recovery processes into a search component and retrieval component, both of which must be successfully concluded in order to output the correct response. The two components undoubtedly are correlated in the sense that stronger, more complete traces will both be easier to find and easier to retrieve having been found.

One final problem of some importance should be mentioned at this time. The effects of trace interference may be quite difficult to separate from those of search failure. Trace interference here refers either to loss of information in the trace due to succeeding inputs or to confusions caused by competition among multiple traces at the moment of test. Search failure refers to an inability to find the trace at all. Thus a decrease in the probability of a correct response as the number of items intervening between study and test increases could be due to trace interference generated by those items. It could also be due to an increased likelihood of failing to find the trace because of the increasing number of items that have to be searched in memory. One way these processes might be separated experimentally would be in a comparison of recognition and recall measures, assuming that a failure to find the trace is less likely in the case of recognition than in the case of recall. At the present, research along these lines has not given us a definitive answer to this question.

SECTION 4. EXPERIMENTS CONCERNED WITH SHORT-TERM PROCESSES

Sections 2 and 3 of this paper have outlined a theoretical framework for human memory. As we have seen, the framework is extremely general, and there are many alternative choices that can be made in formulating models for particular experimental situations. The many choice points make it impossible for us to examine each process experimentally. Instead we shall devote our attention to a number of processes universally agreed to occur in experiments on memory, namely rehearsal and search processes. In Section 5 the LTS search processes will be examined in detail; in the present section the major emphasis will be on STS mechanisms, particularly the control process designated as the rehearsal buffer. The sensory registration system is not an important factor in these models; the experiments are designed so that all items enter the sensory register and then are transferred to STS. The long-term store will be presented in the models of this section but only in the simplest possible manner. We now turn to a series of experiments designed to establish in some detail the workings of the buffer mechanism.

4.1. A Continuous Paired-Associate Memory Task (Experiment 1).

This study is the prototype for a series of experiments reported in this section designed specifically to study buffer processes. The buffer is a fixed-size rehearsal scheme in STS; conditions which prompt the subject to make use of a buffer include difficulty in using long-term store, a large number of short study-test intervals, and a presentation rate slow enough that cognitive manipulations in STS are not excessively rushed. The task that was developed to establish these conditions is described below.*

The subject was required to keep track of constantly changing responses associated with a fixed set of stimuli.** The stimuli were two-digit numbers chosen from the set 00 - 99; the responses were letters of the alphabet. At the start of a particular subject-session a set of s stimuli was chosen randomly from the numbers 00 to 99; these stimuli were not changed over the course of that day's session. To begin the session each stimulus was paired with a letter chosen randomly from the alphabet. Following this initial period, a continuous sequence of trials made up the rest of the session, each trial consisting of a test phase followed by a study phase. During the test phase, one of the s stimuli was randomly selected and presented alone for test. The subject was required to respond with the most recent response paired with that stimulus. No feedback was given to the subject. Following his response the study portion of the trial began. During the study portion the stimulus just presented for test was paired with a new response selected randomly from the alphabet; the only restriction was that the previous response (the correct response during the immediately preceding test phase) was not used during the study phase of the same trial.

---

* The reader may consult Atkinson, Brelsford, and Shiffrin (1967) for details of the experimental procedure and theoretical analyses that are not covered in the present discussion. Also presented there is an account of the mathematics of the model.

** The task is similar to those used by Yntema and Mueser (1960, 1962), Brelsford, Keller, Shiffrin, and Atkinson (1966), and Katz (1966).

The subject was instructed to forget the previous pairing and try to remember the new pairing currently being presented for study. Following the study period, a stimulus was again selected randomly from the set of *s* stimuli and the test portion of the next trial began.

The result of this procedure is as follows: a particular stimulus-response pair is presented for study, followed by a randomly determined number of trials involving other stimuli, and then tested. Having been tested, the pair is broken up and the stimulus is paired with a different response; in other words, no stimulus-response pair is presented for study twice in succession. It is easy to imagine the effects of this procedure on the subject's long-term memory processes. If any particular pair is strongly stored in long-term memory, it will interfere with subsequent pairings involving that same stimulus. In addition, the nature of the stimuli and responses used makes coding a difficult task. For these reasons, the subject soon learns that the usual long-term storage operations, such as coding, are not particularly useful; in fact, the subject is forced to rely heavily on his short-term store and his rehearsal capacity. The experimental procedure also was designed so that it would be possible to carry out extensive parametric analyses on data from individual subjects. This was accomplished by running each subject for twelve or more days and collecting the data on a system under the control of a time-sharing computer, a procedure which made the precise sequence of events during each session available for analysis.

Method. The subjects were nine students from Stanford University who received $2 per experimental session. This experiment, and most of the others reported in this paper, was conducted in the Computer-Based Learning Laboratory at Stanford University. The control functions were performed by computer programs run on a modified PDP-1 computer manufactured by the Digital Equipment Corporation, and under control of a time-sharing system. The subject was seated at a cathode-ray-tube display terminal; there were six terminals each located in a separate 7 x 8 ft. sound-shielded room. Stimuli were displayed on the face of the cathode ray tube (CRT); responses were made on an electric tyepwriter keyboard located immediately below the lower edge of the CRT.

For each session the subject was assigned to one of the three experimental conditions. The three conditions were defined in terms of s, the size of the set of stimuli to be remembered, which took on the values 4, 6 or 8. An attempt was made to assign subjects to each condition once in consecutive three-session blocks. Every session began with a series of study trials: one study trial for each stimulus to be used in the session. On a study trial the word "study" appeared on the upper face of the CRT. Beneath the word "study" one of the stimuli (a two-digit number) appeared along with a randomly-selected letter from the alphabet. Subjects were instructed to try to remember the stimulus-response pairs. Each of these initial study trials lasted for 3 seconds with a 3-second intertrial interval. As soon as there had been an initial study trial for each stimulus to be used in the session, the session proper began.

Each subsequent trial involved a fixed series of events. (1) The word test appeared on the upper face of the CRT. Beneath the word test a randomly selected member of the stimulus set appeared. Subjects were instructed that when the word test and a stimulus appeared on the CRT, they were to respond with the last response that had been associated with that stimulus, guessing if necessary. This test portion of a trial lasted for 3 seconds. (2) The CRT was blacked out for 2 seconds. (3) The word study appeared on the upper face of the CRT for 3 seconds. Below the word study a stimulus-response pair appeared. The stimulus was the same one used in the preceding test portion of the trial. The response was randomly selected from the letters of the alphabet, with the stipulation that it be different from the immediately preceding response assigned to that stimulus. (4) There was a 3-second intertrial interval before the next trial. Thus a complete trial (test plus study) took 11 seconds. A subject was run for 220 such trials during each experimental session.

Theoretical Analysis. In order that the reader may visualize the sequence of events which occurs in this situation, a sample sequence of 18 trials is illustrated in Figure 3. Within the boxes are the displays seen on the CRT screen. In this session the stimulus set includes the four stimuli 20, 31, 42, and 53 (i.e., $s = 4$). On trial n, item 31-Q is presented for study. On trial n+1, 42 is tested and 42-B presented for study. Then on trial n+2, 31 is tested; the correct answer is Q as is seen by referring to trial n. After the subject answers he is given 31-S to study. He is instructed to forget the previous pair, 31-Q, and remember only the new pair, 31-S. The response letter S was selected randomly from the alphabet, with the restriction that the previous response, Q, could not be used. A previously used response may through chance, however, be chosen again later in the session; for example, on trial n+7, 31-Q is again presented for study. It is also possible that two or more stimuli might be paired with the same response concurrently; as an example, on trial

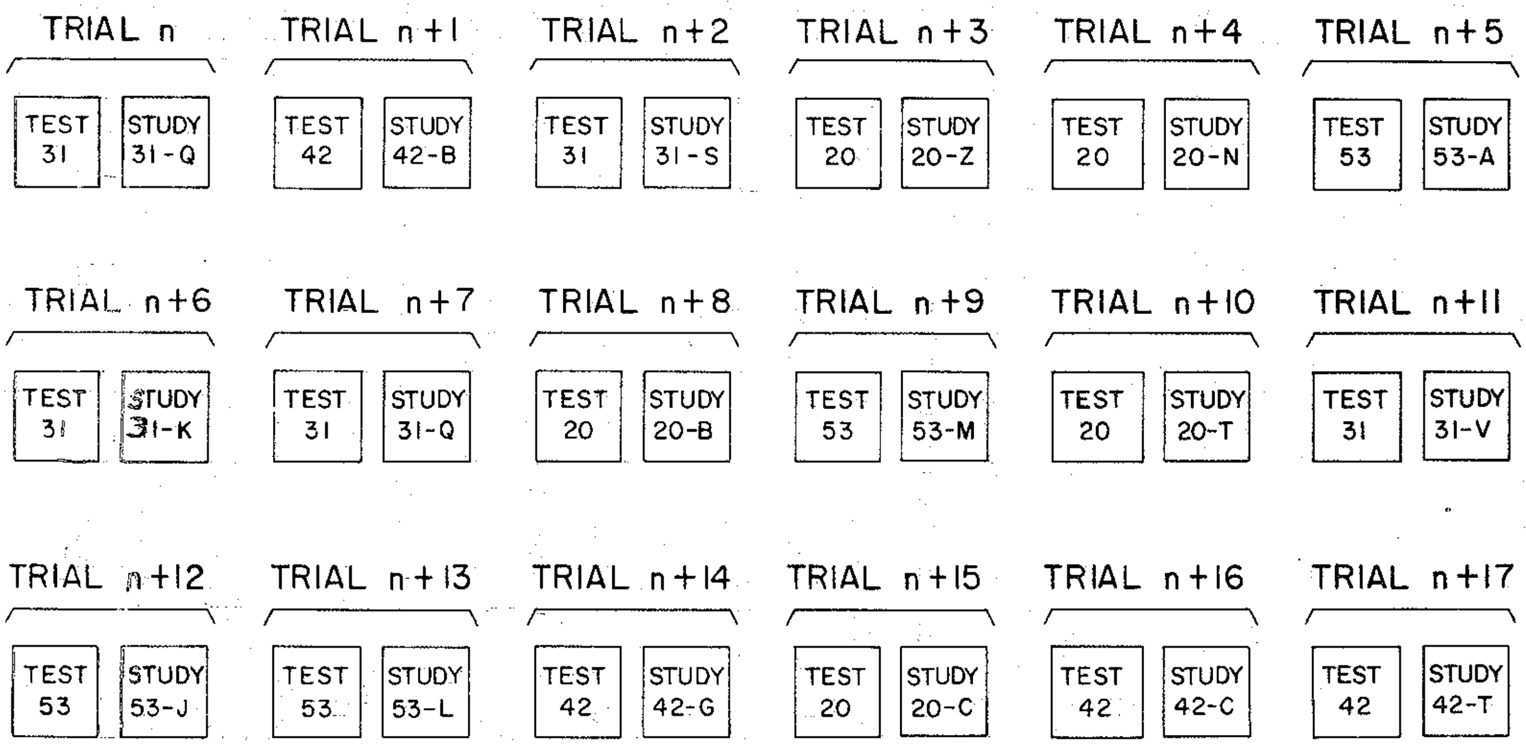


Figure 3 A sample sequence of trials for Experiment 1

n+15, 20 is paired with C and on trial n+16, 42 also is paired with C. The stimulus presented on each trial is chosen randomly; for this reason the number of trials intervening between study and test is a random variable distributed geometrically. In the analysis of the results, a very important variable is the number of trials intervening between study and test on a particular stimulus-response pair; this variable is called the lag. Thus 20 is tested on trial n+4 at a lag of 0 because it was studied on trial n+3. On the other hand, 42 is tested on trial n+14 at a lag of 12, because it was last studied on trial n+1.

Consider now the processes the subject will tend to adopt in this situation. The obvious difficulties involved in the use of LTS force the subject to rely heavily upon rehearsal mechanisms in STS for optimal performance.* A strategy making effective use of STS is an ordered rehearsal scheme of fixed size called the buffer in Section 3.2. The fixed size requirement may not be necessary for maximal utilization of STS, but is indicated by the following considerations. Keeping the size of the rehearsal set constant gives the subject a great deal of control over the situation; each rehearsal cycle will take about the same amount of time, and it is easier to reorganize the buffer when a new item is introduced. Furthermore, an attempt to stretch the rehearsal capacity to its limit may result in

---

* The usual examples given for the usefulness of a distinct short-term store do not stress the positive benefits of a memory decaying quickly and completely. Without such a memory, many minor tasks such as adding a long column of numbers might become far more difficult. The current experiment, in which associative bonds are frequently broken and reformed, is an example of a class of operations for which a short-term store is almost essential.

confusion which causes the entire rehearsal set to be disrupted; the confusion results from the variable time that must be allowed for operations such as responding at the keyboard and processing the new incoming items. The hypothesis of an ordered fixed-size buffer is given support by the subjects' reports and the authors' observations while acting as subjects. The reader is not asked, however, to take our word on this matter; the analysis of the results will provide the strongest support for the hypothesis.

It must be decided next just what is being rehearsed. The obvious candidate, and the one reported by subjects is the stimulus-response pair to be remembered. That is, the unit of rehearsal is the two-digit stimulus number plus the associated response letter. Under certain conditions, however, the subject may adopt a more optimal strategy in which only the responses are rehearsed. This strategy will clearly be more effective because many more items may be encompassed with the same rehearsal effort. The strategy depends upon ordering the stimuli (usually in numerical order in the present case) and rehearsing the responses in an order corresponding to the stimulus order; in this way the subject may keep track of which response goes with which stimulus. For a number of reasons, the scheme is most effective when the size of the stimulus set is small; for a large set the subject may have difficulty ordering the stimuli, and difficulty reorganizing the rehearsal as each new item is presented. When the number of stimulus-response pairs to be remembered is large, the subject may alter this scheme in order to make it feasible. The alteration might consist of rehearsing only the responses associated with a portion of the ordered stimuli. In a previous experiment (Brelsford, Atkinson, Keller, and Shiffrin, 1966) with a similar design, several subjects reported using such a strategy when the stimulus set size was four, and an examination of their results showed better performance than the other subjects. Subject reports lead us to believe that this strategy is used infrequently in the present experiment; consequently, our model assumes that the unit of rehearsal is the stimulus-response pair, henceforth called an "item."

Figure 2 outlines the structure of the model to be applied to the data. Despite the emphasis on rehearsal, a small amount of long-term storage occurs during the period that an item resides in the buffer. The information stored in LTS is comparatively weak and decays rapidly as succeeding items are presented. In accord with the argument that the long-term process is uncomplicated, we assume here that information stored in LTS increases linearly with the time an item resides in the buffer. Once an item leaves the buffer the LTS trace is assumed to decrease as each succeeding item is presented for study.

Every item is assumed to enter first the sensory register and then STS. At that point the subject must decide whether or not to place the new item in the rehearsal buffer. There are a number of reasons why every incoming item may not be placed in the buffer. For one thing, the effort involved in reorganizing the buffer on every trial may not always appear worthwhile, especially when the gains from doing so are not immediately evident; for another, the buffer at some particular time may consist of a combination of items especially easy to rehearse and the subject may not wish to destroy the combination. In order to be more specific about which items enter the buffer and which do not, two kinds of items must be distinguished. An O-item is an incoming stimulus-response pair whose stimulus is currently in the buffer. Thus if 52-L is currently in the buffer, 52 is tested, and 52-G is presented for study, then 52-G is said to be an O-item. Whenever an O-item is presented it is automatically entered into the buffer; this entry, of course, involves replacing the old response by the appropriate new response. Indeed, if an O-item did not enter the buffer, the subject would be forced to rehearse the now incorrect previous response, or to leave a useless blank spot in the buffer; for these reasons, the assumption that O-items are always entered into the buffer seems reasonable. The other kind of item that may be presented is an N-item. An N-item is a stimulus-response pair whose stimulus currently is not in the buffer. Whenever an N-item is entered into the buffer, one item currently in the buffer must be removed to make room for the new item (i.e., the buffer is assumed to be of fixed size, r, meaning that the number of items being rehearsed at any one time is constant). The assumption is made that an N-item enters into the buffer with probability α; whenever an N-item is entered one of the items currently in the buffer is randomly selected and removed to make room for it.

The model used to describe the present experiment is now almost complete. A factor still not specified is the response rule. At the moment of test any item which is in the buffer is responded to correctly. If the stimulus tested is not in the buffer, a search is carried out in LTS with the hope of finding the trace. The probability of retrieving the correct response from LTS depends upon the current trace strength, which in turn, depends on the amount of information transferred to LTS. Specifically we assume that information is transferred to LTS at a constant rate θ during the entire period an item resides in the buffer; θ is the transfer rate per trial. Thus, if an item remains in the rehearsal buffer for exactly j trials, then that item accumulated an amount of information equal to $j\theta$. We also assume that each trial following the trial on which an item is knocked out of the buffer causes the information stored in LTS for that item to decrease by a constant proportion τ. Thus, if an item were knocked out of the buffer at trial j, and i trials intervened between the original study and test on that item, then the amount of information in LTS at the time of the test would be $j\theta\tau^{i-j}$. We now want to specify the probability of a correct retrieval of an item from LTS. If the amount of information in LTS at the moment of test is zero, then the probability of a correct retrieval should be at the guessing level. As the amount of information increases, the probability of a correct retrieval should increase toward unity. We define ρ_{ij} as the probability of a correct response from LTS for an item that was tested at lag i, and resided in the buffer for exactly j trials. Considering the above specifications on the retrieval process,

$$\rho_{ij} = 1 - (1 - g)\exp[- j\theta(\tau^{i-j})]$$

where g is the guessing probability, which is 1/26 since there were 26 response alternatives.*

---

* Lest the use of an exponential function seem entirely arbitrary, it should be noted that this function bears a close relation to the familiar linear model of learning theory. If we ignore for the moment the decay feature, then $\rho_{ij} = 1 - (1 - g)\exp(-j\theta)$. It is easily seen that this is the linear model expression for the probability of a correct response after j reinforcements with parameter $e^{-\theta}$. Thus, the retrieval function ρ_{ij} can be viewed as a linear model (Cont'd on next page)

The basic dependent variable in the present experiment is the probability of a correct response at the time of a test, given lag $\underline{i}$. In order to derive this probability we need to know the length of time that an item resides in the memory buffer. Therefore, define

β_j = probability that an item resides in the buffer for exactly j trials, given that it is tested at a lag greater than j.

The probability of a correct response to an item tested at lag $\underline{i}$ can now be written in terms of the β_j's. Let "C_i" represent the occurence of a correct response to an item tested at lag $\underline{i}$. Then

$$\Pr(C_i) = \left[1 - \sum_{k=0}^{i} \beta_k\right] + \left[\sum_{k=0}^{i} \beta_k \rho_{ik}\right] .$$

The first bracketed term is the probability that the item is in the buffer at the time of the test. The second bracket contains a sum of probabilities, each term representing the probability of a correct retrieval from LTS of an item which remained in the buffer for exactly k trials and was then lost.**

---

* (Cont'd from previous page) with time in the buffer as the independent variable. To be sure, the decay process complicates matters, but the reason for choosing the exponential function becomes somewhat less arbitrary. A decay process is needed so that the probability of a correct retrieval from LTS will approach chance as the lag tends toward infinity.

** One factor which the model as outlined ignores is the probability of recovering from LTS an old, incorrect trace. In the interest of simplicity this process has not been introduced into the model, although it could be appended with no major changes.

There are four parameters in the model: r, the buffer size which must be an integer; α, the probability of entering an N-item into the buffer; θ, the transfer rate of information to LTS; and τ, the decay rate of information from LTS after an item has left the buffer.

One final process must be considered before the model is complete. This process is the recovery of information from STS which is not in the buffer. It will be assumed that the decay of an item which has entered and then left the buffer is very rapid, so rapid that an item which has left the buffer cannot be recovered from STS on the succeeding test.* The only time in which a recovery is made from STS, apart from the buffer, occurs if an item is tested immediately following its study (i.e., at a lag of 0). In this case there is virtually no time between study and test and it is assumed therefore that the recovery probability is one, regardless of whether the item was entered into the buffer or not. In other words, the probability correct is one when the lag is zero.

Data Analysis. Figure 4 presents the probability of a correct response as a function of lag for each of the three stimulus set sizes examined. It can be seen that the smaller the stimulus set size, the better the overall performance. It is important to note that the theory predicts such a difference on the following basis: the larger the size of the stimulus set, the more often an N-item will be presented; and the

---

* Clearly this assumption depends on the time intervals involved. In the present experiment the trials were quite slow; in experiments where a faster presentation rate is used, the model probably would need to be modified slightly to allow a non-zero probability of recovery of an item from STS on the test following its removal from the buffer.

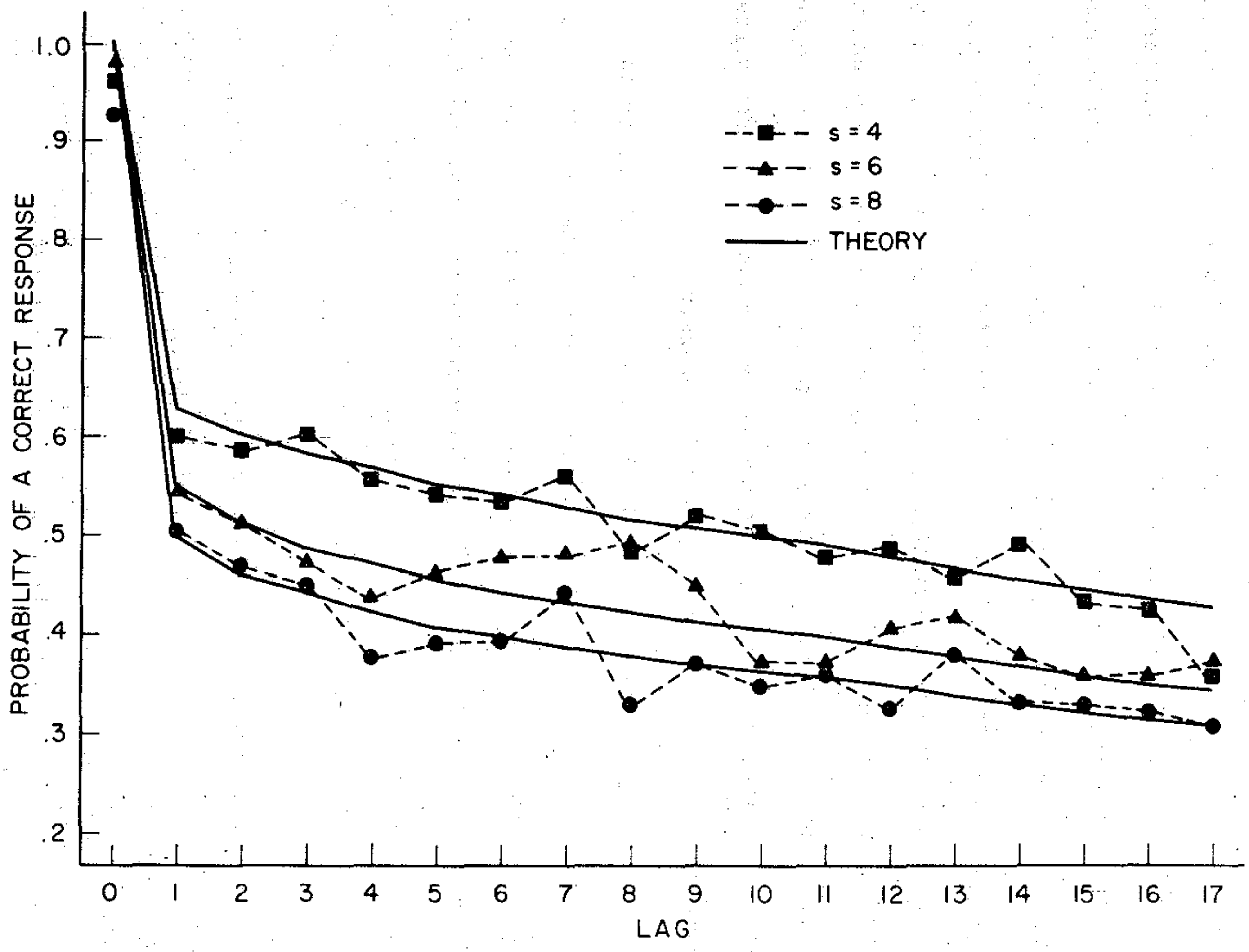


Figure 4 Observed and theoretical probabilities of a correct response as a function of lag (Experiment 1)

more often N-items will be presented, the more often items in the buffer will be knocked out. Recall that only N-items can knock items from the buffer; O-items merely replace themselves.

It can be seen that performance is almost perfect for lag 0 in all three conditions. This was expected because lag 0 means that the item was tested immediately following its study, and was therefore available in STS. The curves drop sharply at first and slowly thereafter, but have not yet reached the chance level at lag 17, the largest lag plotted. The chance level should be 1/26 since there were 26 response alternatives.

The four parameters of the model were estimated by fitting the model to the lag curves in Figure 4 using a minimum chi-square as a best fit criterion.* The solid lines in Figure 5 give the best fit of the model which occurred when the parameter values were: $r = 2$, $\alpha = .39$, $\theta = .40$, and $\tau = .93$. It can be seen that the observed data and the predictions from the model are in close agreement. It should be emphasized that the three curves are fit simultaneously using the same parameter values, and the differences between the curves depend only on the value of s (the stimulus set size) which, of course, is determined by the experimenter. The predicted probabilities of a correct response weighted and summed over all lag positions are .562, .469, and .426 for s equal to 4, 6, and 8, respectively; the observed values are .548, .472, and .421.

The estimated value of r might seem surprising at first glance; two items appear to be a rather small buffer capacity. But there are a number

* See Atkinson, Brelsford, and Shiffrin (1967) for details of the estimation procedure and a statistical evaluation of the goodness-of-fit.

of considerations which render this estimate reasonable. It seems clear that the capacity estimated in a task where the subject is constantly interrupted for tests must be lower than the capacity estimated, for example, in a typical digit-span task. This is so because part of the attention time that would be otherwise alloted to rehearsal must be used to search memory in order to respond to the continuous sequence of tests. Considering that two items in this situation consist of four numbers and two letters, an estimate of r equal to two is not particularly surprising. The estimated value of α indicates that only 39 percent of the N-items actually enter the buffer (remember that O-items always enter the buffer). This low value may indicate that a good deal of mental effort is involved in keeping an item in the buffer via rehearsal, leading to a reluctance to discard an item from the buffer which has not yet been tested. A similar reluctance to discard items would be found if certain combinations of items were particularly easy to rehearse. Finally, note that the theory predicts that, if there were no long-term storage, the subject's overall probability of a correct response would be independent of α. Thus it might be expected that α would be higher the greater the effectiveness of long-term storage. In accord with this reasoning, the low value of α found would result from the weak long-term storage associated with the present situation.

In addition to the lag curves in Figure 4, there are a number of other predictions that can be examined. One aspect of the theory maintains that O-items always enter the buffer and replace themselves, while N-items enter the buffer with probability α and knock an item out of the buffer whenever they do so. The effects of different stimulus-set sizes displayed in Figure 5 are due to this assumption. The assumption, however, may be examined in other ways; if it is true, then an item's probability of being correct will be affected by the specific items that intervene between its initial study and its later test. If every intervening trial uses the same stimulus, then the probability of knocking the item of interest from the buffer is minimized. This is so because once any intervening item enters the buffer, every succeeding intervening item is an O-item (since it uses the same stimulus), and hence also enters the buffer. Indeed, if α were one then every intervening item after the first would be an O-item, and hence only the first intervening item would have a chance of knocking the item of interest from the buffer; if $\alpha = 1$ and there were no long-term decay, then the lag curve for this condition would be flat from lag 1 onwards. In this case, however, α is not equal to one and there is long-term decay; hence the lag curve will decrease somewhat when the intervening items all have the same stimulus, but not to the extent found in Figure 4. This lag curve, called the "all-same" curve, is shown in Figure 5; it plots the probability of a correct response as a function of lag, when all the intervening trials between study and test involve the same stimulus. The parameters previously estimated were used to generate predictions for these curves and they are displayed as solid lines. It seems clear that the predictions are highly accurate.

A converse result, called the "all-different" lag curve, is shown in Figure 6. In this condition, every intervening item has a different stimulus, and therefore the probability of knocking the item of interest from the buffer is maximized. The lag curves for this condition, therefore, should drop faster than the unconditional lag curves of Figure 4. Predictions were again generated using the previous parameter values and are

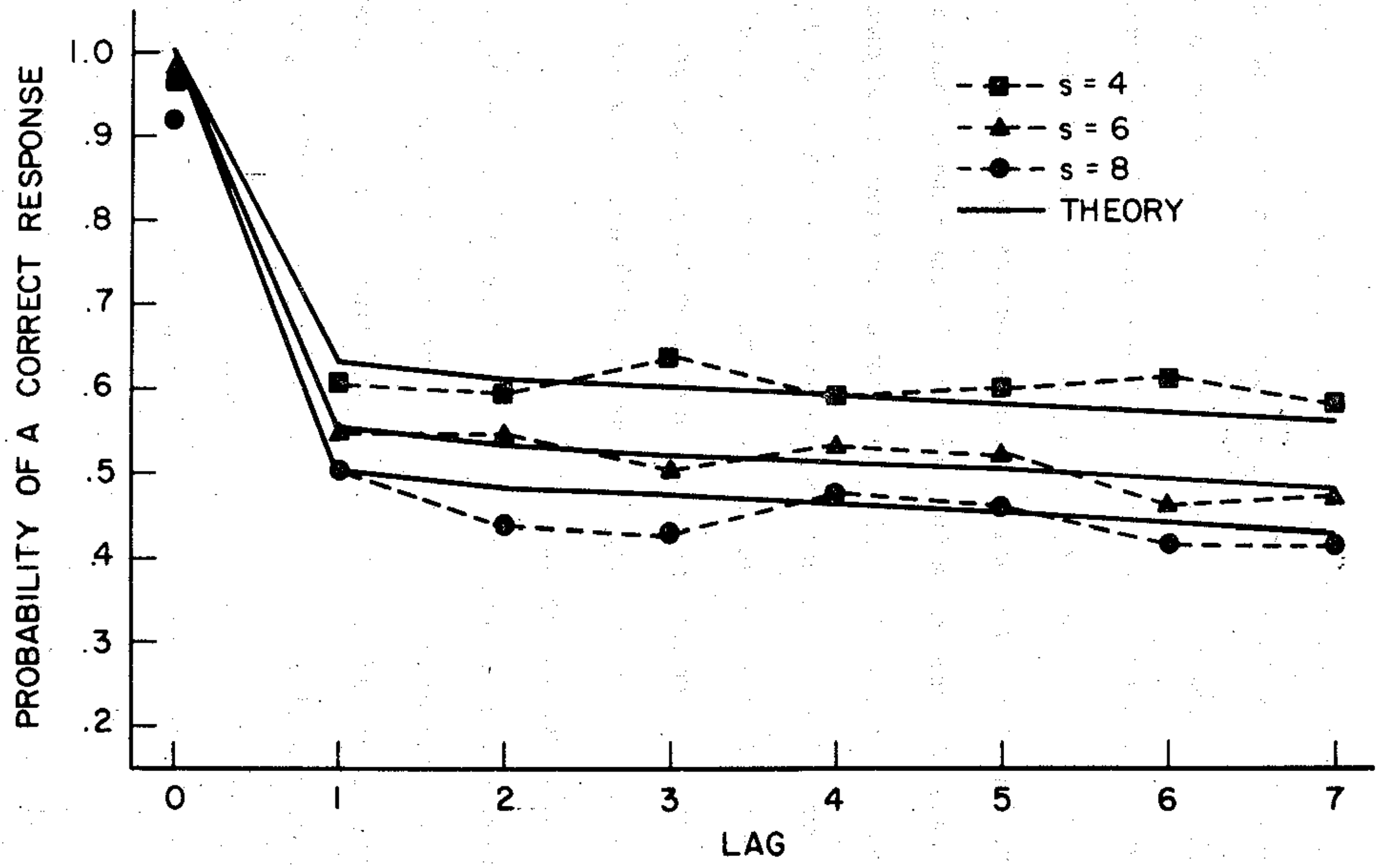


Figure 5 Observed and theoretical probabilities of a correct response as a function of lag when every intervening item uses the same stimulus (Experiment 1)

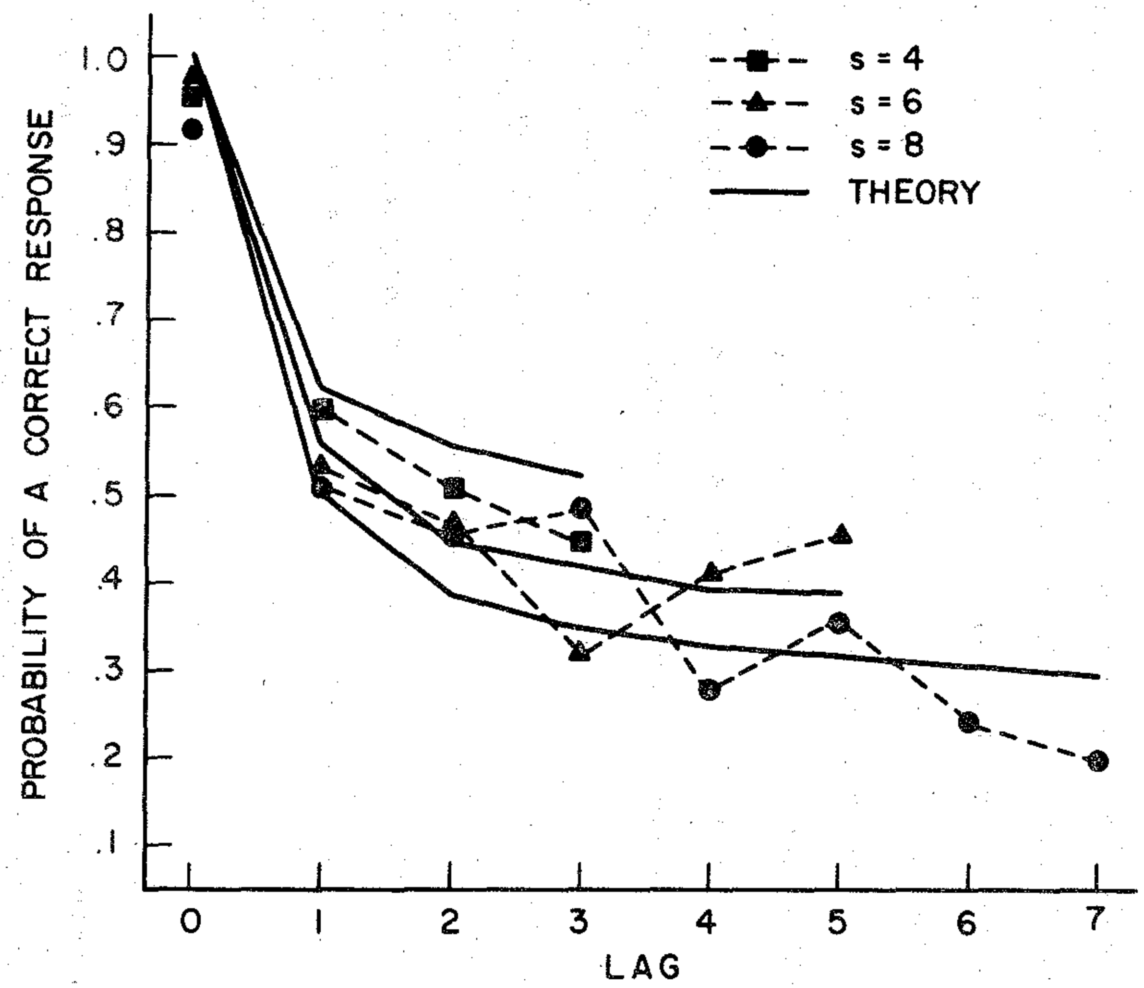


Figure 6 Observed and theoretical probabilities of a correct response as a function of lag when every intervening item uses a different stimulus (Experiment 1)

represented by the solid lines in Figure 6. Relatively few observations were available in this condition; considering the instability of the data the predictions seem reasonable.

The procedure used in this experiment is an excellent example of what has been traditionally called a negative transfer paradigm. The problems inherent in such a paradigm were mentioned earlier as contributing to the subjects' heavy reliance upon the short-term store. To the extent that there is any use of LTS, however, we would expect intrusion errors from previously correct responses. The model could be extended in several obvious ways to predict the occurrence of such intrusions. For example, the subject could, upon failing to recover the most recent trace from LTS, continue his search and find the remains of the previous, now incorrect, trace. In order to examine intrusion errors, the proportion of errors which were the correct response for the previous presentation of the stimulus in question were calculated for each lag and each condition. The proportions were quite stable over lags with mean values of .065, .068, and .073 for the 4, 6, and 8 stimulus conditions, respectively. If the previously correct response to an item is generated randomly for any given error, these values should not differ significantly from 1/25 = .04. In both the $s = 4$ and $s = 6$ conditions seven of the nine subjects had mean values above chance; in the $s = 8$ condition eight of the nine subjects were above chance. Intrusion errors may therefore be considered a reliable phenomenon in this situation; on the other hand, the relatively low frequency with which they occur indicates a rather weak and quickly decaying long-term trace.

A second error category of interest includes those responses that are members of the current set of responses to be remembered but are not the correct responses. This set, of course, includes the set of responses in the buffer at any one time; if the subject tends to give as a guess a response currently in the buffer (and therefore highly available), then the probability of giving as an error a response in the current to-be-remembered set will be higher than chance. Since responses may be assigned to more than one stimulus simultaneously, the number of responses in the to-be-remembered set is bound by, but may be less than, the size of the stimulus set, s. Thus, on the basis of chance the error probabilities would be bounded below .12, .20, and .28 for s = 4, 6, and 8, respectively. The actual values found were .23, .28, and .35, respectively. This finding suggests that when the subject cannot retrieve the response from his buffer or LTS and is forced to guess, he has a somewhat greater than chance likelihood of giving a response currently in the rehearsal set but assigned to another stimulus. It is not surprising that a subject will give as a guess one of the responses in his buffer since they are immediately available.

Other analyses have been performed on the data of this experiment, but the results will not be presented until a second experiment has been described. Before considering the second experiment, however, a few words should be said about individual differences. One of the reasons for running a single subject for many sessions was the expectation that the model could be applied to each subject's data separately. Such analyses have been made and are reported elsewhere (Atkinson, Brelsford, and Shiffrin, 1967). The results are too complex to go into here, but they establish that individual subjects by and large conform to the predictions of the model quite well. Since our aim in this paper is to present a non-technical discussion of the model, to simplify matters we will make most of our analyses on group data.

4.2. The "All-Different" Stimulus Procedure (Experiment 2).

In the preceding experiment, the number of stimuli used in a given experimental session and the size of the to-be-remembered set were identical. These two factors, however, can be made independent. Specifically, a set of all-different stimuli could be used while keeping the size of the to-be-remembered set constant. The name, all-different, for this experiment results from the use of all-different stimuli; i.e., once a given stimulus-response pair is presented for test, that stimulus is not used again. In other respects the experiment is identical to Experiment 1.

One reason for carrying out an experiment of this type is to gain some information about the replacement hypothesis for 0-items. In Experiment 1 we assumed that a new item with a stimulus the same as an item currently in the buffer automatically replaced that item in the buffer; that is, the response switched from old to new. In the all-different experiment subjects are instructed, as in Experiment 1, to forget each item once it has been tested. If an item currently in the buffer is tested (say, 52-G) and a new item is then presented for study (say 65-Q), we might ask whether the tested item will be automatically replaced by the new item (whether 65-Q will replace 52-G in the buffer). This replacement strategy is clearly optimal for it does no good to retain an item in the buffer that already has been tested. Nevertheless, if the reorganization of the buffer is difficult and time consuming, then the replacement of a tested item currently in the buffer might not be carried out. One simple assumption along these lines would postulate that every item has an independent probability α of entering the buffer.

The all-different experiment was identical to Experiment 1 in all respects except the following. In Experiment 1 the s stimuli were the same throughout an experimental session, with only the associated responses being changed on each trial, whereas in the all-different experiment 100 stimuli were available for use in each session. In fact, every stimulus was effectively new since the stimulus for each study trial was selected randomly from the set of all 100 stimuli under the restriction that no stimulus could be used if it had been tested or studied in the previous fifty trials. There were still three experimental conditions with s equal to 4, 6, or 8 denoting the number of items that the subject was required to try to remember at any point in time. Thus a session began with either 4, 6, or 8 study trials on different randomly selected stimuli,each of which was paired with a randomly selected response (from the 26 letters). On each trial a stimulus in the current to-be-remembered set was presented for test. After the subject made his response he was instructed to forget the item he had just been tested on, since he would not be tested on it again. Following the test a new stimulus was selected (one that had not appeared for at least fifty trials) and randomly paired with a response for the subject to study. Thus the number of items to be remembered at any one time stays constant throughout the session. However, the procedure is quite different from Experiment 1 where the study stimulus was always the one just tested.

Denote an item presented for study on a trial as an O-item (old item) if the item just tested was in the buffer. Denote an item presented for study as an N-item (new item) if the item just tested was not in the buffer. This terminology conforms precisely to that used to describe Experiment 1. If an O-item is presented there will be at least one spot in the buffer occupied by a useless item (the one just tested). If an N-item is presented, the buffer will be filled with information of the same value as that before the test. If we assume that an N-item has probability α of entering the buffer, and that an O-item will always enter the buffer and knock out the item just made useless, then the model for Experiment 1 will apply here with no change whatsoever. In this case we again expect that the lag curves for $s = 4$, 6, and 8 would be separated. In fact, given the same parameter values, exactly the same curves would be predicted for the all-different experiment as for Experiment 1.

As noted earlier, however, there is some doubt that the assumptions regarding N-items and O-items will still hold for the all-different experiment. In Experiment 1 the stimulus just tested was re-paired with a new response, virtually forcing the subject to replace the old response with a new one if the item was in the buffer. Put another way, if an item is in the buffer when tested, only a minor change need be made in the buffer to enter the succeeding study item: a single response is replaced by another. In the all-different experiment, however, a greater change needs to be made in order to enter an O-item; both a stimulus and a response member have to be replaced. Thus an alternative hypothesis might maintain that every entering item (whether an N-item or an O-item) has the same probability α of entering the buffer, and will knock out any item currently in the buffer with equal likelihood. In this case we predict no differences among the lag curves for the $s = 4, 6$, and 8 conditions.

Results. The observed lag curves for Experiment 2 are displayed in Figure 7. It should be emphasized that, except for the procedural changes described above and the fact that a new sample of subjects was used, the experimental conditions and operations were identical in experiments 1 and 2. The important point about this data is that the lag curves for the three conditions appear to overlap.* For this reason we lump the three curves to form the single lag curve displayed in Figure 8.

Because the three curves overlap, it is apparent that the theory used in Experiment 1 needs modification. The hypothesis suggested above will be used: every item enters the buffer with probability α. If an item enters the buffer it knocks out an item already there on a random basis. This model implies that useless items are being rehearsed on occasion, and subjects reported doing just that despite instructions to forget each item once tested.

The curve in Figure 8 was fit using a minimum χ^2 procedure; the parameter estimates were $r = 2$, $\alpha = .52$, $\theta = .17$, and $\tau = .90$. It can be seen that the fit is excellent. Except for r, the parameters differ somewhat from those found in Experiment 1, primarily in a slower transfer rate, θ. In Experiment 1 the estimate of θ was .40. This reduction in long-term storage is not too surprising since the subjects were on occasion rehearsing useless information. It could have been argued in advance of the data that the change away from a strong "negative-transfer" paradigm

---

* To determine whether the three curves in Figure 7 differ reliably, the proportions correct for each subject and condition were calculated and then ranked. An analysis of variance for correlated means did not yield significant effects ($\underline{F} = 2.67$, df $= 2/16$, $p > .05$).

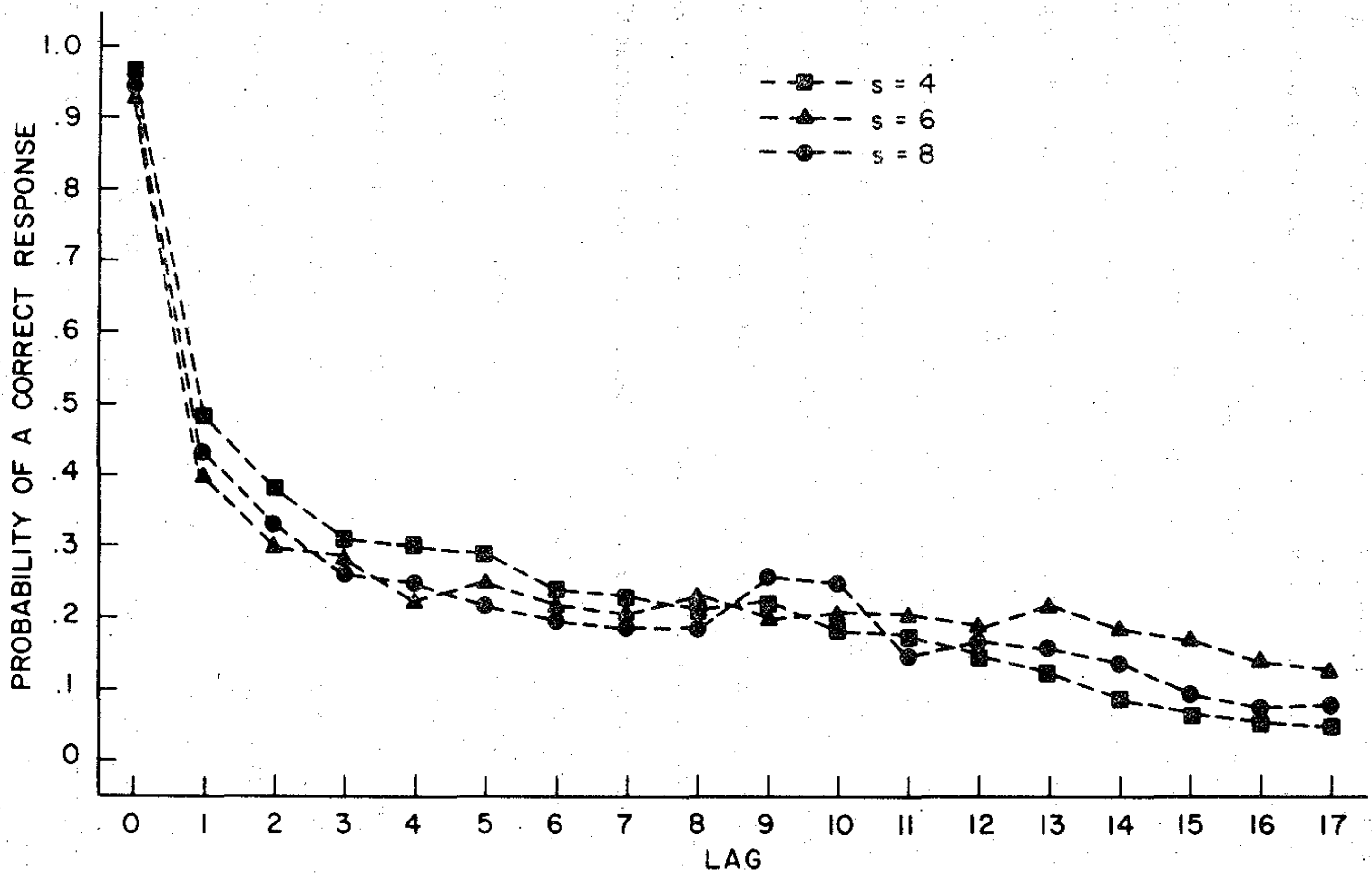


Figure 7 Observed and theoretical probabilities of a correct response as a function of lag (Experiment 2)

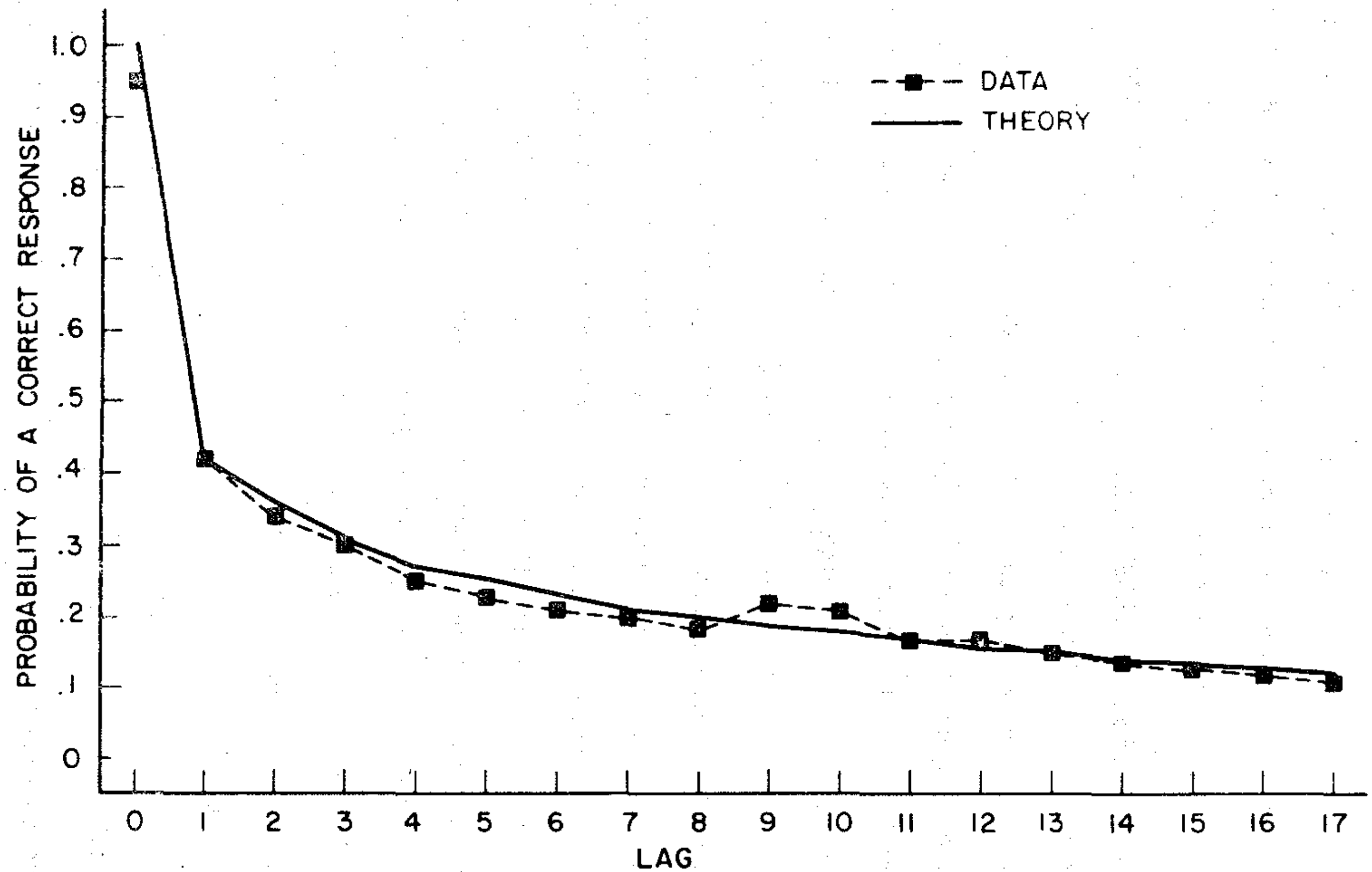


Figure 8 Observed and theoretical probabilities of a correct response as a function of lag. Data from the $s = 4$, 6, and 8 conditions have been pooled (Experiment 2)

in Experiment 2 would lead to increased use of LTS; that this did not occur is indicated not only by the low θ value, but also by the low probability of a correct response at long lags. One outcome of this result is the possibility that the all-different procedure would give superior long-term memory in situations where subjects could be induced to attempt coding or other long-term storage strategies. It seems apparent that LTS was comparatively useless in the present situation.

Some Statistics Comparing Experiments 1 and 2. In terms of the model, the only difference between Experiments 1 and 2 lies in the replacement assumption governing the buffer. In Experiment 1, an item in the buffer when tested is automatically replaced by the immediately succeeding study item; if the tested item is not in the buffer, the succeeding study item enters the buffer with probability α, randomly displacing an item already there. In Experiment 2, every study item, independent of the contents of the buffer, enters the buffer with probability α, randomly displacing an item already there. While these assumptions are given credence by the predictions of the various lag curves of Figures 4 and 8, there are other statistics that can be examined to evaluate their adequacy. These statistics depend upon the fact that items vary in their probability of entering the buffer. Since items which enter the buffer will have a higher probability correct than items which do not, it is relatively easy to check the veracity of the replacement assumptions in the two experiments.

In Experiment 1, the probability that an item will be in the buffer at test is higher the greater the number of consecutive preceding trials that involve the same stimulus. Thus if the study of 42-B is preceded, for example, by six consecutive trials using stimulus 42, there is a very high probability that 42-B will enter the buffer. This occurs because there is a high probability that the stimulus 42 already will be in the buffer when 42-B is presented, and if so, then 42-B will automatically enter the buffer. In any series of consecutive trials all with the same stimulus, once any item in the series enters the buffer, every succeeding item will enter the buffer. Hence the longer the series of items with the same stimulus, the higher the probability that that stimulus will be in the buffer. Figure 9 graphs the probability of a correct response to the last stimulus-response pair studied in a series of consecutive trials involving the same stimulus; the probability correct is lumped over all possible lags at which that stimulus-response pair is subsequently tested. This probability is graphed as a function of the length of the consecutive run of trials with the same stimulus and is the line labeled Experiment 1. These curves are combined over the three experimental conditions (i.e., s = 4, 6, 8). We see that the probability of a correct response to the last item studied in a series of trials all involving the same stimulus increases as the length of that series increases, as predicted by the theory.

In Experiment 2 stimuli are not repeated, so the above statistic cannot be examined. A comparable statistic exists, however, if we consider a sequence of items all of which are tested at zero lag (i.e., tested immediately after presentation). One could hypothesize that the effect displayed in Figure 9 for Experiment 1 was due to a consecutive sequence of zero-lag tests, or due to factors related to the sequence of correct answers (at zero-lag an item is always correct). These same arguments

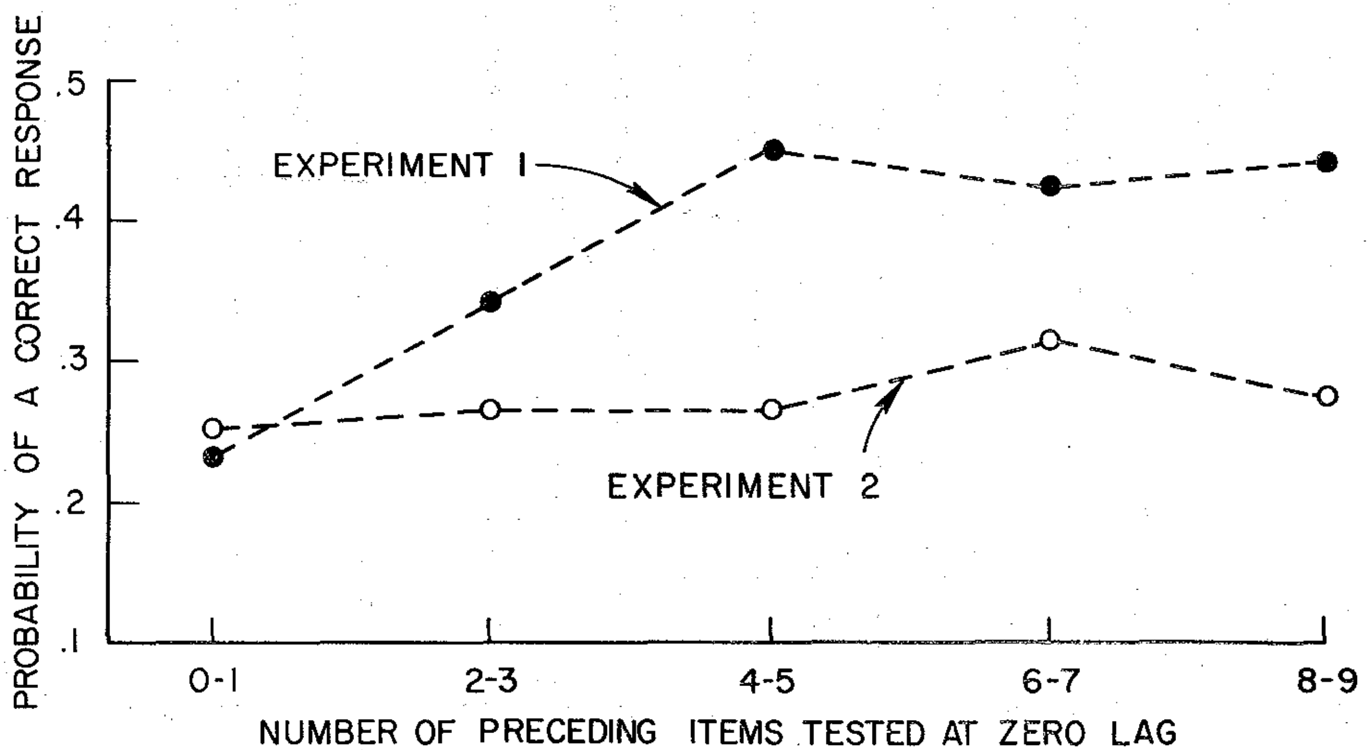


Figure 9 Probability of a correct response as a function of the number of consecutive preceding items tested at zero lag (Experiment 1 and Experiment 2)

would apply, however, to the sequence of zero-lag items in Experiment 2. In Figure 9, the line labeled Experiment 2 represents a probability measure comparable to the one displayed for Experiment 1. Specifically, it is the probability of a correct response on the eventual test of the last S-R pair studied in a consecutive sequence of trials all involving S-R pairs tested at lag zero, as a function of the length of the sequence. The model for Experiment 2 with its scheme for entering items in the buffer, predicts that this curve should be flat; the data seem to bear out this prediction.

The close correspondence between the predicted and observed results in Experiments 1 and 2 provides strong support for the theory. The assumptions justified most strongly appear to be the fixed-size rehearsal buffer containing number-letter pairs as units, and the replacement assumptions governing O- and N-items. It is difficult to imagine a consistent system without these assumptions that would give rise to similar effects. Some of the predictions supported by the data are not at all intuitive. For example, the phenomenon displayed in Figure 9 seems to be contrary to predictions based upon considerations of negative transfer. Negative transfer would seem to predict that a sequence of items having the same stimulus but different responses would lead to large amounts of interference and hence reduce the probability correct of the last item in the sequence; however, just the opposite effect was found. Furthermore, the lack of an effect in Experiment 2 seems to rule out explanations based on successive correct responses or successive zero-lag tests. Intuition notwithstanding, this effect was predicted by the model.

4.3. A Continuous Paired-Associate Memory Task with Multiple Reinforcements (Experiment 3).

In contrast to a typical short-term memory task, the subjects' strategy in paired-associate learning shifts from a reliance on rehearsal processes to a heavy emphasis on coding schemes and related processes that facilitate long-term storage. There are many factors, however, that contribute to such a shift, and the fact that items are reinforced more than once in a paired-associate learning task is only one of these. In the present experiment, all factors are kept the same as in Experiment 1, except for the number of reinforcements. It is not surprising, then, that subjects use essentially the same rehearsal strategy found in Experiment 1. It is therefore of considerable interest to examine the effects associated with repeated reinforcements of the same item.

In Experiment 3 only one stimulus set size, $s = 8$, was used. Each session began with eight study trials on which the eight stimuli were each randomly paired with a response. The stimuli and responses were two digit numbers and letters, respectively. After the initial study trials the session involved a series of consecutive trials each consisting of a test phase followed by a study phase. On each trial a stimulus was randomly selected for testing and the same stimulus was then presented for study on the latter portion of the trial. Whereas in Experiment 1, during the study phase of a trial, the stimulus was always re-paired with a new response, in the present experiment the stimulus was sometimes left paired with the old response. To be precise, when a particular S-R pair was presented for study the first time, a decision was made as to how many reinforcements (study periods) it would be given; it was given either 1, 2, 3, or 4 reinforcements with probabilities .30, .20, .40, and .10 respectively. When a particular S-R pair had received its assigned number of reinforcements, its stimulus was then re-paired with a new response on the next study trial, and this new item was assigned a number of reinforcements using the probability distribution specified above. In order to clarify the procedure, a sample sequence from trials n to n+19 is shown in Figure 10. On trial n+2 stimulus 22 is given a new response, L, and assigned three reinforcements, the first occurring on trial n+2. The second reinforcement occurs on trial n+3 after a lag of zero. After a lag of 6, the third reinforcement is presented on trial n+10. After a lag of 8, stimulus 22 is re-paired with a new response on trial n+19. Stimulus 33 is sampled for test on trial n+6 and during the study phase is assigned the new response, B, which is to receive two reinforcements, the second on trial n+9. Stimulus 44 is tested on trial n+4, assigned the new response X which is to receive only one reinforcement; thus when 44 is presented again on trial n+16 it is assigned another response which by chance also is to receive only one reinforcement, for on the next trial 44 is studied with response Q. The subject is instructed, as in Experiments 1 and 2, to respond on the test phase of each trial with the letter that was <u>last</u> studied with the stimulus being tested.

The same display devices, control equipment, and timing relations used in Experiment 1 were used in this study. There were 10 subjects, each run for at least 10 sessions; a session consisted of 220 trials. Details of the experimental procedure, and a more extensive account of the data analysis, including a fit of the model to response protocols

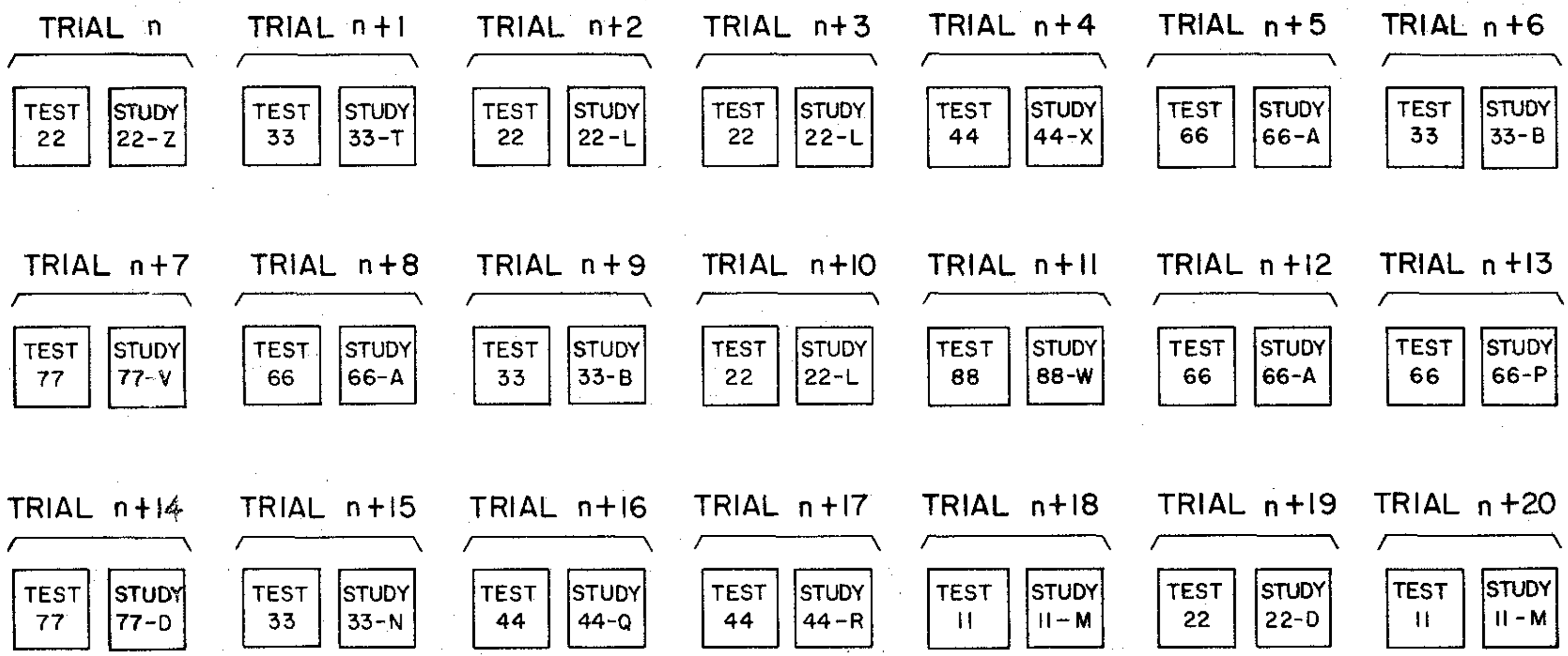


Figure 10 A sample sequence of trials for Experiment 3

of individual subjects, can be found in Brelsford and Atkinson (1967).

The model for Experiment 1 may be used without change in the present situation. There is some question, however, whether it is reasonable to do so. The assumptions concerning LTS storage and decay may be applied to items which are given multiple reinforcements: information is transferred to LTS at a rate θ whenever the item resides in the buffer, and decays from LTS by the proportion τ on each trial that the item is not present in the buffer. The assumption regarding O-items also may be applied: since the stimulus already is in the buffer, the new response replaces the old one thereby entering the item in the buffer (if, as is the case in this experiment, the old response is given yet another study, then nothing changes in the buffer). N-items, however, are not so easily dealt with. N-items, remember, are items whose stimuli are not currently represented in the buffer. In Experiment 1, the stimulus of every N-item also was being paired with a new response. In the current experiment this is not always the case; some N-items, although not in the buffer, will be receiving their 2nd, 3rd, or 4th reinforcement when presented for study. That is, some N-items in this experiment, will already have a substantial amount of information stored on them in LTS. It seems reasonable that subjects may not rehearse an item which has just been retrieved correctly from LTS. The assumption regarding N-items is therefore modified for purposes of the present experiment as follows. If a stimulus is tested and is not in the buffer, then a search of LTS is made. If the response is correctly retrieved from LTS, and if that stimulus-response pair is repeated for study, then that item will not be entered into the buffer (since the subject "knows" it already). If a new item is presented for study (i.e., the response to that stimulus is changed), or if the correct response is not retrieved from LTS (even though the subject may have made the correct response by guessing), then the study item enters the buffer with probability α. This slight adjustment of the replacement assumption allows for the fact that some items presented for study may already be known and will not enter the rehearsal buffer. This version of the model is the one used later to generate predictions for the data.

Results. Figure 11 presents the probability of a correct response as a function of lag for items tested after their first, second, and third reinforcements. The number of observations is weighted not only toward the short lags, but also toward the smaller numbers of reinforcements. This occurs because the one-reinforcement lag curve contains not only the data from the items given just one reinforcement, but also the data from the first reinforcement of items given two, three, and four reinforcements. Similarly, the lag curve following two reinforcements contains the data from the second reinforcement of items given two, three, and four reinforcements, and the three reinforcement curve contains data from the third reinforcement of items given three and four reinforcements. The lag curves in Figure 11 are comparable to those presented elsewhere in this paper. What is graphed is the probability of a correct response to an item that received its j^{th} reinforcement, and was then tested after a lag of n trials. The graph presents data for n ranging from 0 to 15 and for j equal to 1, 2, and 3. Inspecting the figure, we see that an item which received its first reinforcement and was then tested at a lag of 8 trials gave a correct response about 23 percent of the time; an item that received its second reinforcement and was then tested at lag 8 had about 44 percent correct responses; and an item that received its third reinforcement and was then tested at lag 8 had about 61 percent correct.

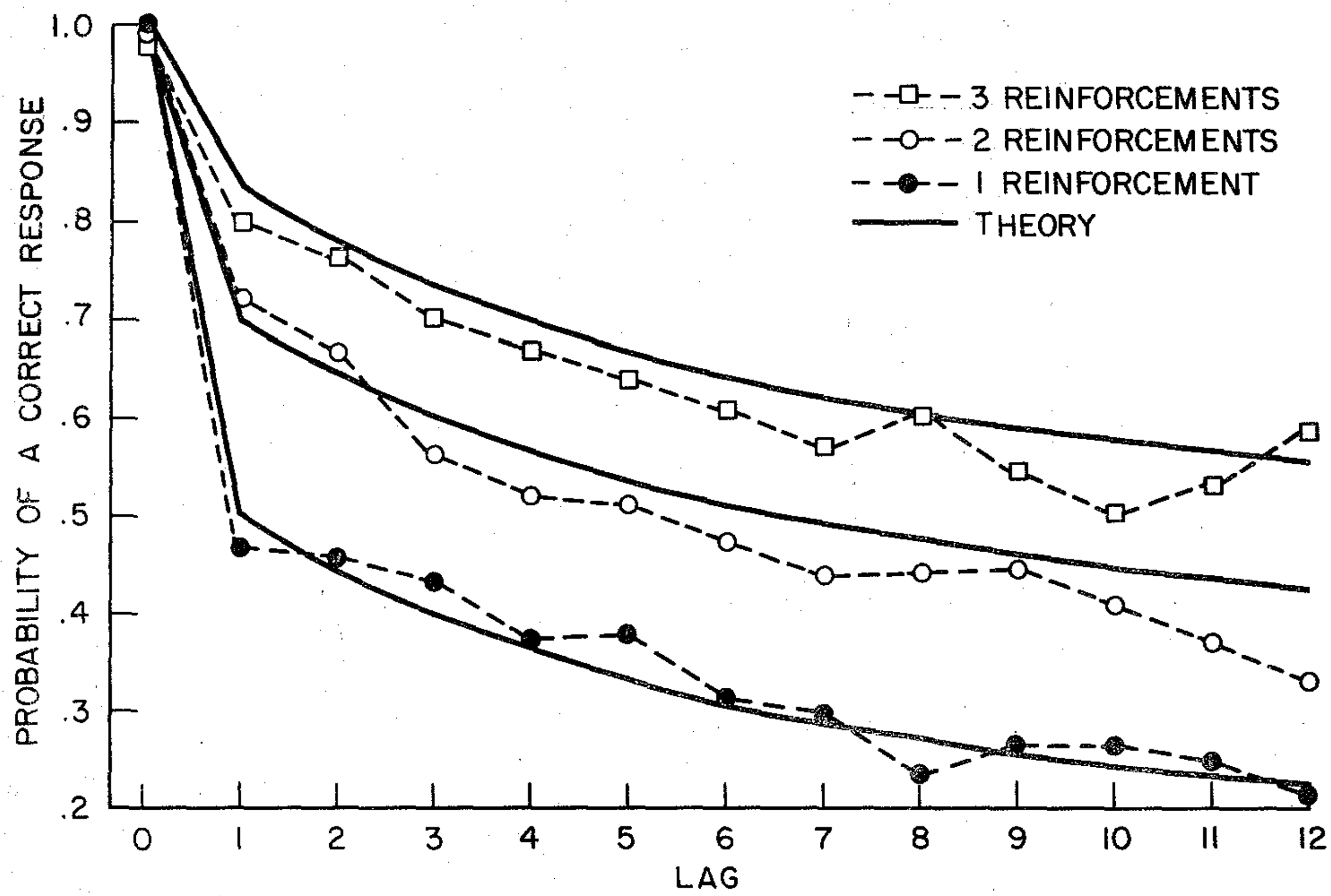


Figure 11 Observed and theoretical probabilities of a correct response as a function of lag for items tested following their 1st, 2nd or 3rd reinforcement (Experiment 3)

The curves in Figure 11 exhibit a consistent pattern. The probability correct decreases regularly with lag, starting at a higher value on lag 1 the greater the number of prior reinforcements. Although these curves are quite regular, there are a number of dependencies masked by them. For example, the probability of a correct response to an item that received its second reinforcement and was then tested at some later trial, will depend on the number of trials that intervened between the first and second reinforcments. To clarify this point consider the following diagram

$\boxed{\text{22-Z}}$	$\xrightarrow{\text{lag } \underline{a}}$	$\boxed{\text{22}}$	$\boxed{\text{22-Z}}$	$\xrightarrow{\text{lag } \underline{b}}$	$\boxed{\text{22}}$
(1st study)		(1st test)	(2nd study)		(2nd test)

Item 22-Z is given its first reinforcement, tested at lag *a* and given a second reinforcement, and then given a second test at lag *b*. For a fixed lag *b*, the probability of a correct response on the 2nd test will depend on lag *a*. In terms of the model it is easy to see why this is so. The probability correct for an item on the second test will depend upon the amount of information about it in LTS. If lag *a* is extremely short, then there will have been very little time for LTS strength to build up. Conversely, a very long lag *a* will result in any LTS strength decaying and disappearing. Hence the probability of a correct response on the second test will be maximal at some intermediate value of lag *a*; namely, at a lag which will give time for LTS strength to build up, but not so much time that excessive decay will occur. For this reason a plot of probability correct on the second test as a function of the lag between the first and second reinforcement should exhibit an inverted U-shape. Figure 12 is

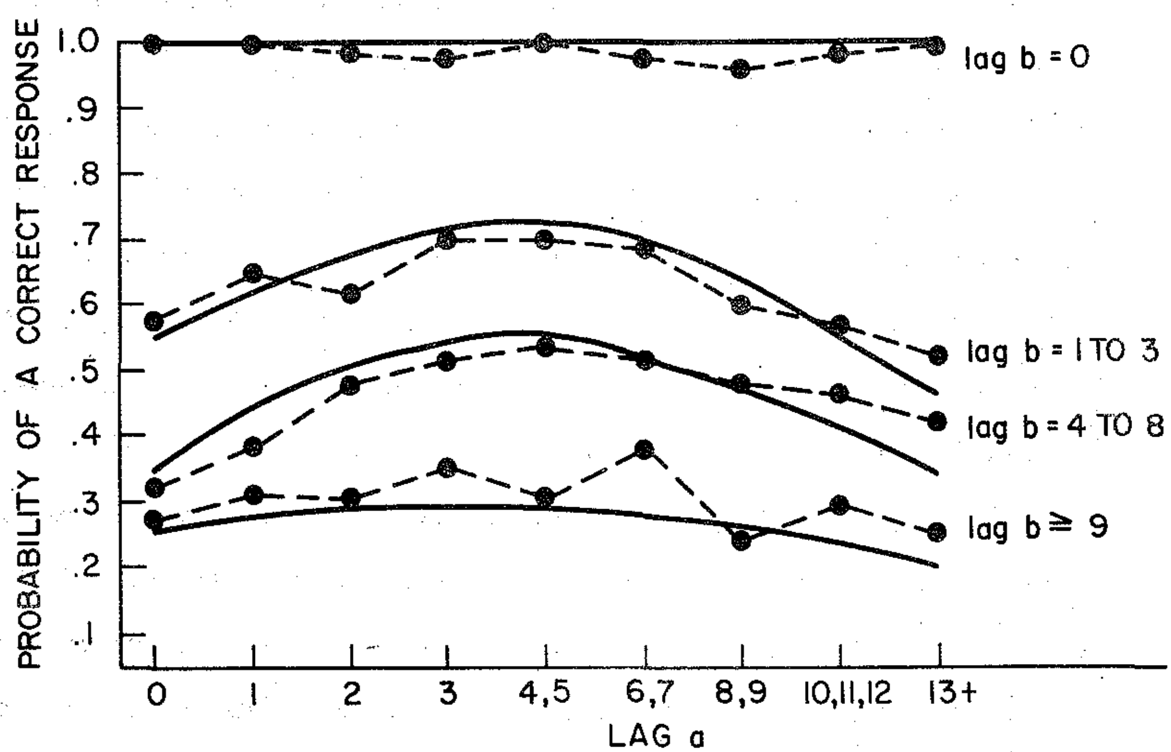


Figure 12 Observed and theoretical probabilities of a correct response as a function of lag a (the spacing between the 1st and 2nd reinforcement) (Experiment 3)

such a plot. The probability correct on the second test is graphed as a function of lag a. Four curves are shown for different values of lag b. The four curves have not been lumped over all values of lag b because we wish to indicate how the U-shaped effect changes with changes in lag b. Clearly, when lag b is zero, the probability correct is one and there is no U-shaped effect. Conversely, when lag b is very large, the probability correct will tend toward chance regardless of lag a, and again the U-shaped effect will disappear. The functions shown in Figure 12 give support to the assumption that information is being transferred to LTS during the entire period an item resides in the buffer. If information is transferred, for example, only when an item first enters the buffer, then it is difficult to explain the rise in the functions of Figure 12 for lag a going from zero to about five. The rise is due to the additional information transferred to LTS as lag a increases.

Theoretical Analysis. A brief review of the model is in order. O-items (whose stimulus is currently in the buffer) always enter the buffer. N-items (whose stimulus is not currently in the buffer) enter the buffer with probability α if they are also new items (i.e., receiving their first reinforcement). However, N-items do not enter the buffer if they are repeat items and were correctly retrieved from LTS on the immediately preceding test; if they are repeat items and a retrieval was not made, then they enter the buffer with probability α. An O-item entering the buffer occupies the position of the item already there with the same stimulus; an entering N-item randomly replaces one of the items currently in the buffer. During the period an item resides in the buffer information is transferred to LTS at a rate θ per trial. This information decays by a proportion τ on each trial after an item has left the buffer.* The subject is always correct at a lag of zero, or if the item is currently in the buffer. If the item is not in the buffer a search of LTS is made, and the correct response is retrieved with a probability that is an exponential function of the amount of information currently in LTS (i.e., the same function specified for Experiments 1 and 2). If the subject fails to retrieve from LTS, then he guesses. There are four parameters for this model: r, the buffer size; α, the buffer entry probability; θ, the transfer rate of information to LTS; and τ, the parameter characterizing the LTS decay rate once an item has left the buffer.

Estimates of r, α, θ, and τ were made using the data presented in Figures 11 and 12. We shall not go into the estimation procedures here for they are fairly complex; in essence they involve a modified minimum χ^2 procedure where the theoretical values are based on Monte Carlo runs. The parameter estimates that gave the best fit to the data displayed in Figures 11 and 12 were as follows: $r = 3$; $\alpha = .65$; $\theta = 1.24$; and $\tau = .82$. Once these estimates had been obtained they were then used to generate a large-scale Monte Carlo run of 12,500 trials. The Monte Carlo procedure involved generating pseudo-data following precisely the rules specified by the model and consulting a random number generator whenever an event occurred in the model that was probabilistically determined.

* In this experiment an item receiving x reinforcements may enter the buffer as many as x times. When the item is in the buffer the θ-process is activated, and when not in the buffer the τ-process takes over.

Thus the pseudo-data from a Monte Carlo run is an example of how real data would look if the model was correct, and the parameters had the values used in the Monte Carlo computation. In all subsequent discussions of Experiment 3, the predicted values are based on the output of the Monte Carlo run. The run was very long so that in all cases the theoretical curves are quite smooth, and we doubt if they reflect fluctuations due to sampling error. A detailed account of the estimation and prediction procedures for this experiment is given in Brelsford and Atkinson (1967).

The predictions from the theory are shown as the smooth curves in Figures 11 and 12. It should be evident that the predicted values are quite close to the observed ones. Note also that the seven curves in the two figures are fit simultaneously with the same four parameter values; the fact that the spacing of the curves is accurately predicted is particularly interesting.

We now examine a number of statistics that were not used in making parameter estimates. First consider the all-same and all-different curves shown in Figure 13; these are the same functions displayed in Figures 5 and 6 for Experiment 1. For the all-same curve, we compute the probability of a correct response as a function of the lag, when all the intervening items between study and test involve the same stimulus. There are three such curves depending on whether the study was the first, second or third reinforcement of the particular S-R pair. The model predicts that once the intervening stimulus enters the buffer, there will be no further chance of any other item being knocked out of the buffer. Hence these curves should drop at a much slower rate than the unconditional lag curves in Figure 11. The all-different curve plots the probability of a

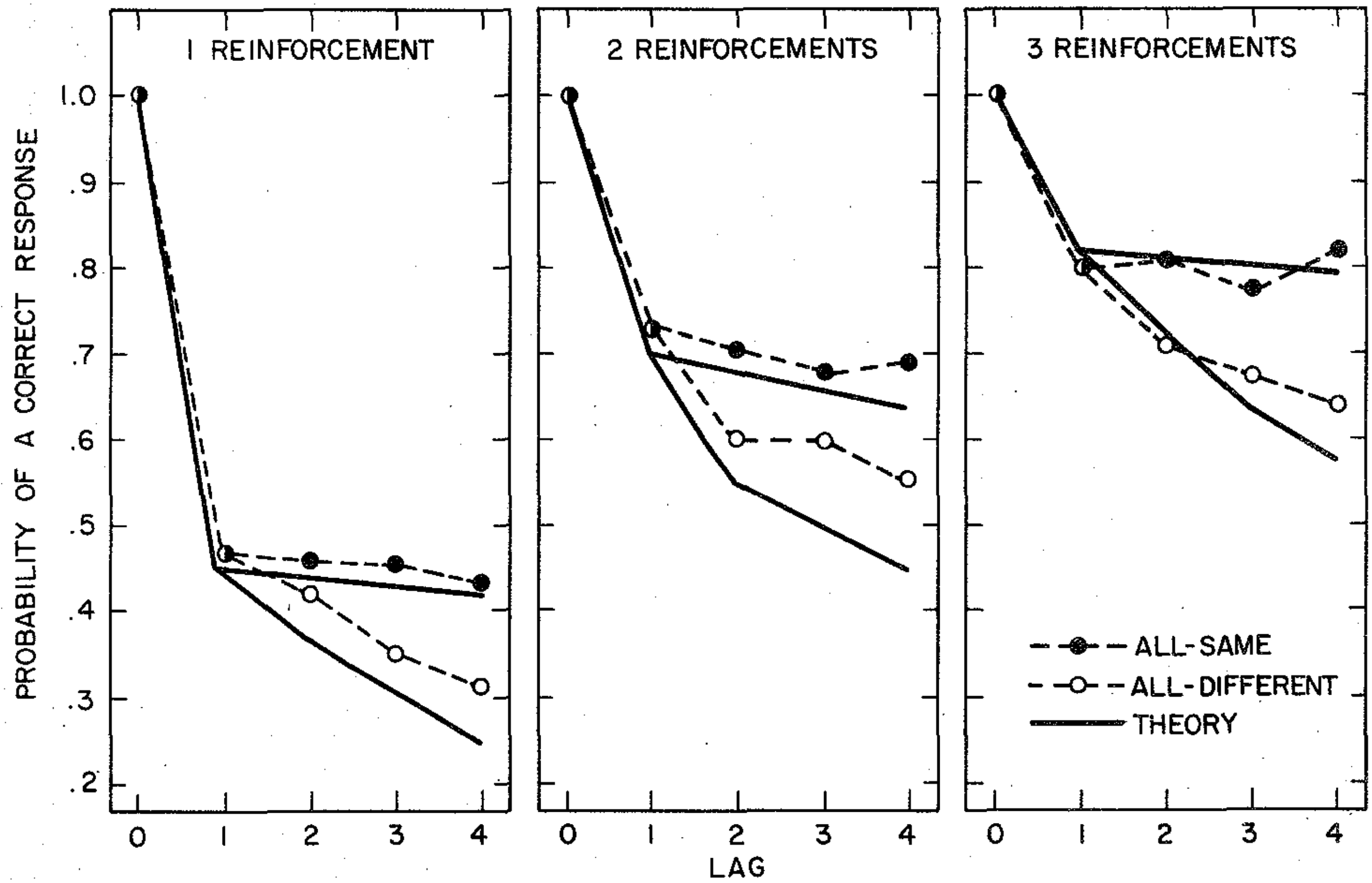


Figure 13 Observed and theoretical probabilities of a correct response as a function of lag for the "all-same" and "all-different" conditions (Experiment 3)

correct response as a function of lag, when the intervening items between study and test all involve different stimuli. Again there are three curves depending on whether the study was the first, second or third reinforcement of the S-R pair. The all-different sequence maximizes the expected number of intervening N-items and therefore the curve should have a much faster drop than the unconditional lag curves in Figure 11. The predictions are shown in the figure as solid lines. The correspondence between predicted and observed values is reasonably good. It is particularly impressive when it is noted that the parameter values used in making the predictions were estimated from the previous data.

We next examine the data displayed in Figure 14. Consider a sequence of consecutive trials all involving the same stimulus, but where the response paired with the stimulus on the study phase of the last trial in the sequence is different from the response on the immediately preceding trial. Then, the theory predicts that the longer this sequence of consecutive trials, the higher will be the probability of a correct response when the last item studied in the sequence is eventually tested. This is so because the probability of the last item entering the buffer increases as the length of the sequence increases: once any item in the sequence enters the buffer, every succeeding one will. The data is shown in Figure 14. What is graphed is the length of the sequence of trials all involving the same stimulus versus the probability of a correct response when the last item studied in the sequence is eventually tested. In this graph we have lumped over all lags at which the eventual test of the last item is made. The predictions generated from the previously estimated parameter values are shown as the smooth line. The predicted values, though not perfect,

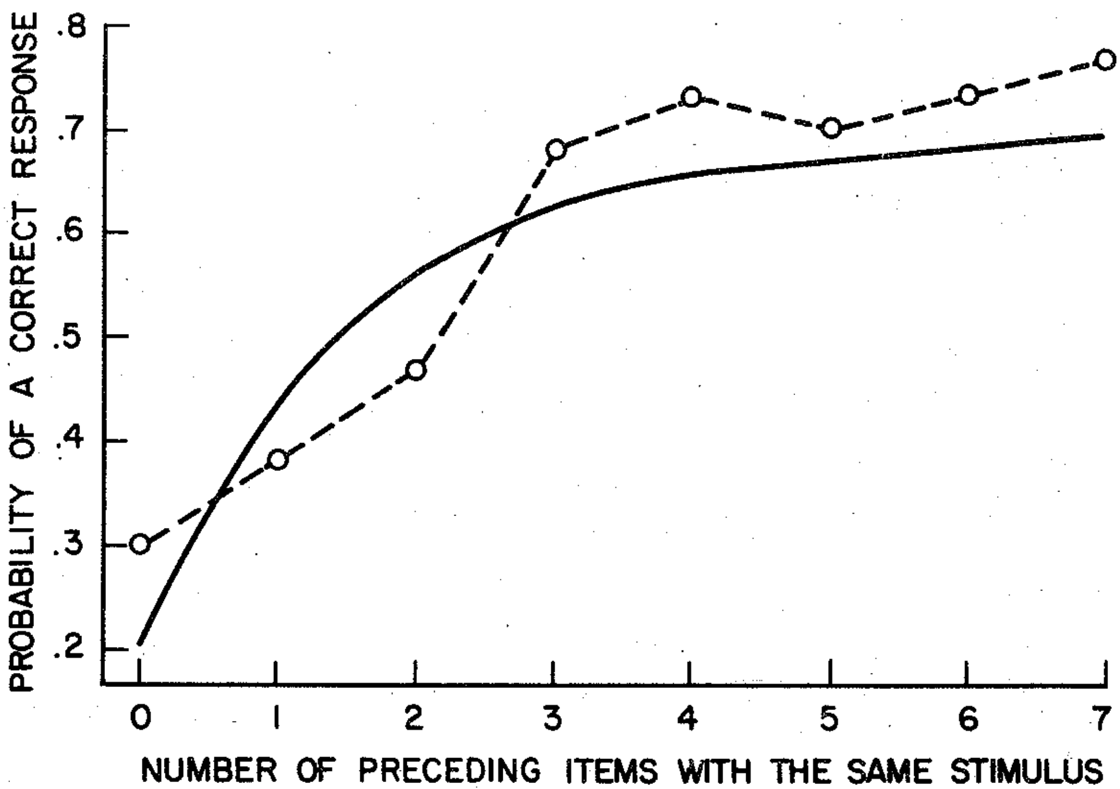
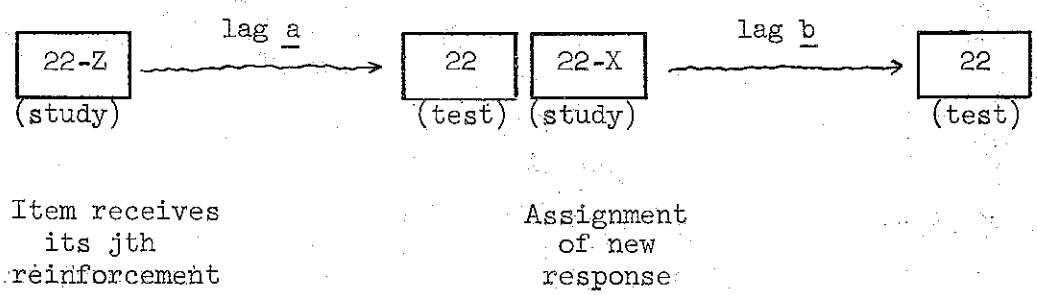


Figure 14 Observed and theoretical probabilities of a correct response as a function of the number of consecutive preceding items using the same stimulus (Experiment 3)

are surprisingly close to the observed proportions correct. It is worth reemphasizing that considerations of negative transfer make this result somewhat unexpected (see page 87).

We next examine another prediction of the theory that ran counter to our initial intuitions. To make matters clear, consider the following diagram:



Item 22-Z is studied for the j^{th} time and then tested at lag _a_; on this trial 22 is paired with a new response X, and tested next at lag _b_. According to the theory, the shorter lag _a_, the better performance should be when the item is tested after lag _b_. This prediction is based on the fact that the more recently a stimulus had appeared, the more likely that it was still in the buffer when the next item using it was presented for study; if the stimulus was in the buffer, then the item using it would automatically enter the buffer. In the present analysis, we examine this effect for three conditions: the preceding item using the stimulus in question could have just received its 1st, 2nd or 3rd reinforcement. Figure 15 presents the appropriate data. In terms of the above diagram, what is plotted is the value of lag _a_ on the abscissa versus the probability of a correct response lumped over all values of lag _b_ on the ordinate; there is a separate curve for j = 1, 2, and 3.

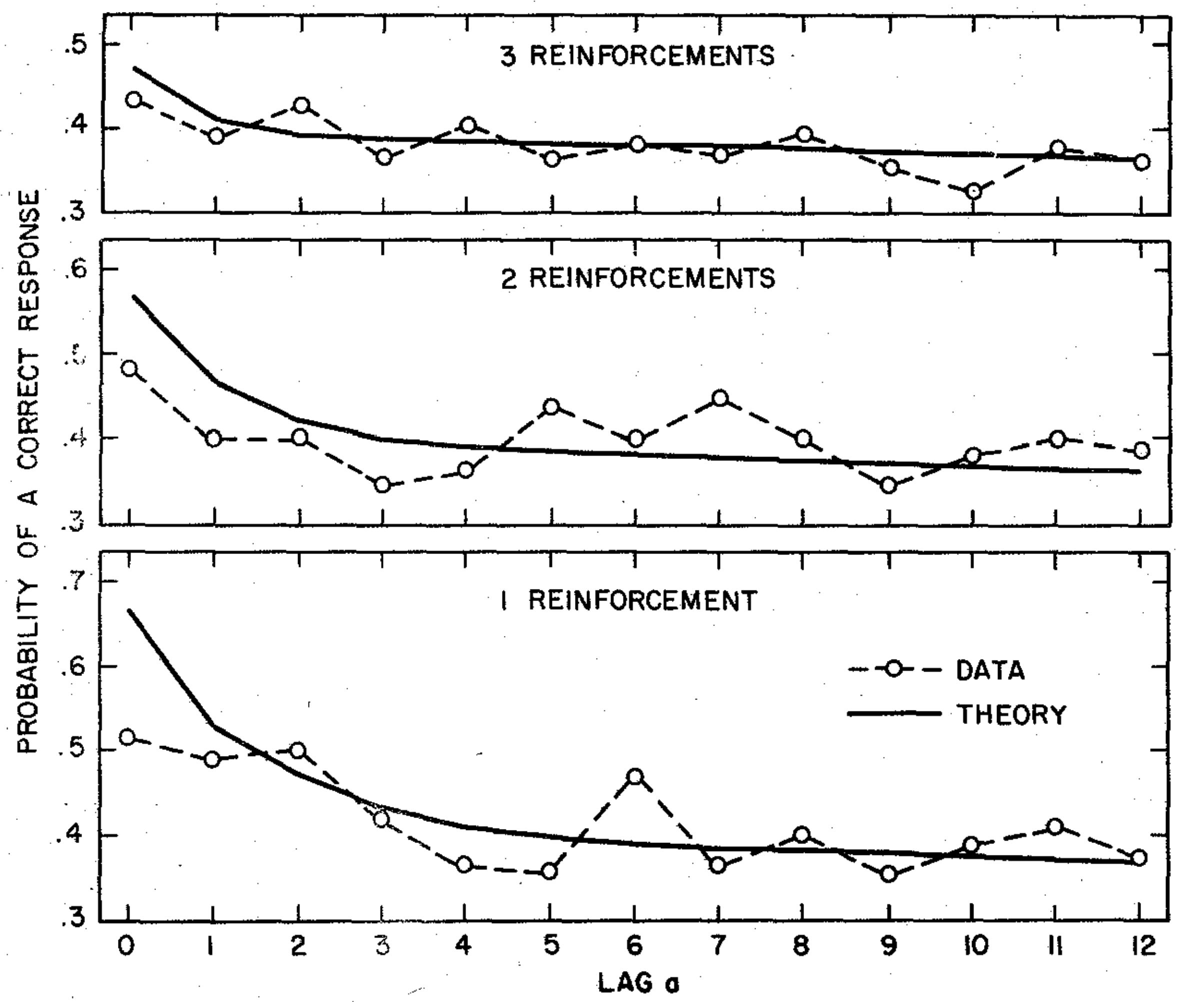


Figure 15 Observed and theoretical probabilities of a correct response as a function of lag <u>a</u> (the lag of the item preceding the item tested, but using the same stimulus) (Experiment 3)

The predicted curves are based upon the previous parameter estimates. The predictions and observations coincide fairly well, but the effect is not as dramatic as one might hope.* One problem is that the predicted decrease is not very large. Considerably stronger effects may be expected if each curve is separated into two components: one where the preceding item was correct at test and the other where the preceding item was not correct. In theory the decrease predicted in Figure 15 is due to a lessened probability of the relevant stimulus being in the buffer as lag a increases. Since an item in the buffer is always responded to correctly, conditionalizing upon correct responses or errors (the center test in the above diagrams) should magnify the effects. To be precise, the decrease will be accentuated for the curve conditional upon correct responses, whereas no decrease at all is predicted for the curve conditional upon errors. If an error is made, the relevant stimulus cannot be in the buffer and hence the new item enters the buffer with probability α independent of lag a. Figure 16 gives the conditional curves and the predictions. The decreasing effect is fairly evident for the "correct" curves; as predicted the "error curves are quite flat over

---

* A curve comparable to the one displayed in Figure 15 for the one-reinforcement condition was obtained from the data of Experiment 1. This curve showed a similar but more pronounced drop and was well predicted by the model.

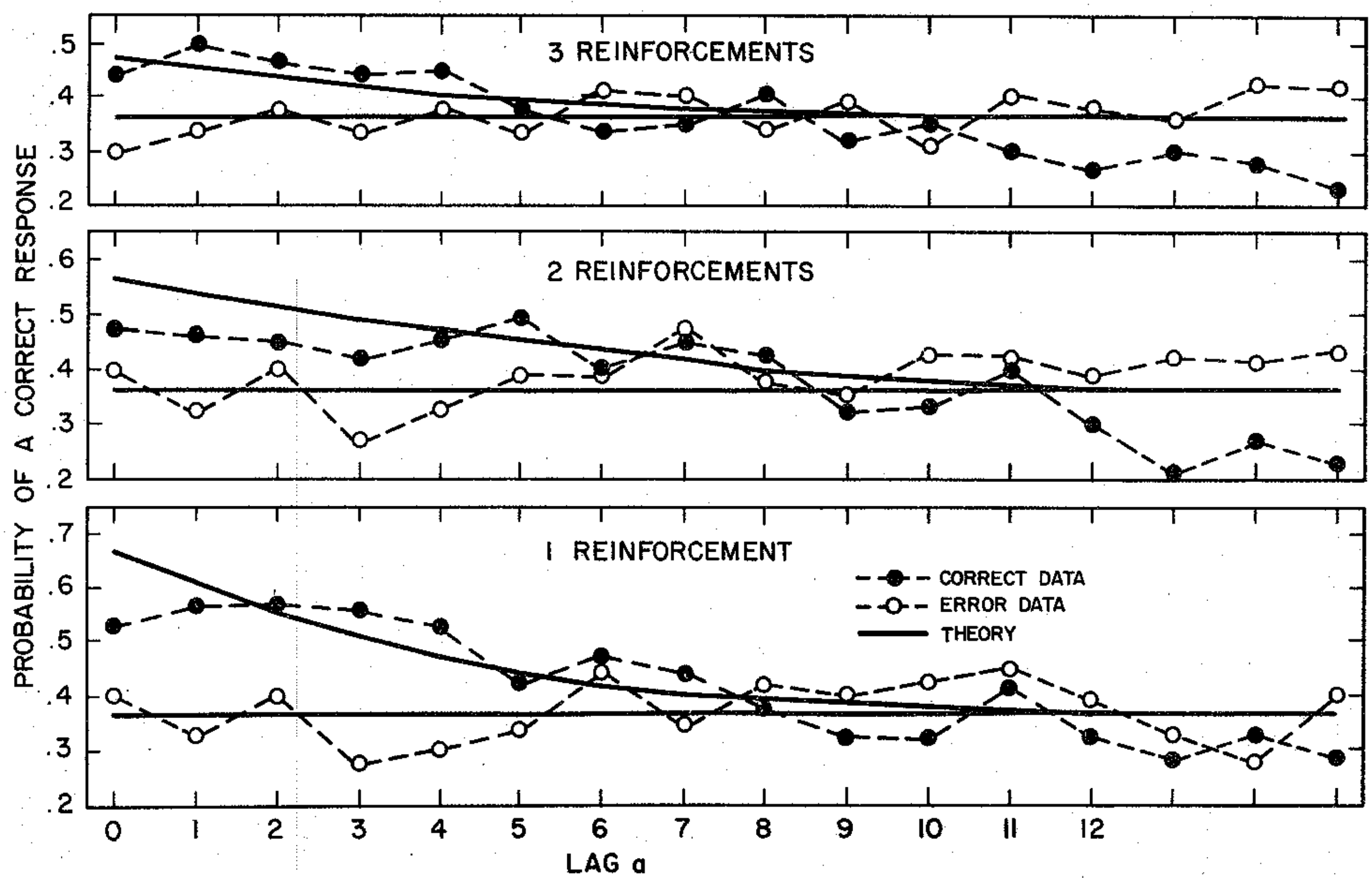


Figure 16 Observed and theoretical probabilities of a correct response as a function of lag a conditionalized on errors or successes on the test at lag a (Experiment 3)

lags.* Conceivably one might argue that the effects are due to item selection: correct responses indicating easier stimuli and incorrect responses indicating more difficult ones. This objection, however, seems contra-indicated in the present case. It is difficult to imagine how item selection could explain the crossing of the correct and error curves found in each of the three diagrams.** Indeed, the model does not explain the crossover -- the model predicts that the two curves should meet. The model is in error at this point because it has not been extended to include negative transfer effects, an extension which would not be difficult to implement. An item responded to correctly at a long lag probably has a strong LTS trace; this strong trace would then interfere with the LTS trace of the new item which, of course, uses the same stimulus. All in all, these curves and predictions may be considered to provide fairly strong support for the details of the model, even to the extent

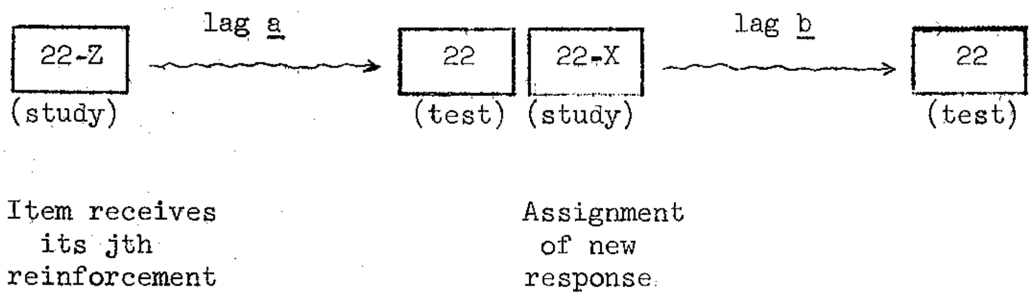
---

* The astute reader will have noticed that the predicted decrease becomes smaller as the number of reinforcements increases. The fact that the data support this prediction is quite interesting, for it sheds light upon the buffer replacement assumptions used in this study. The decreasing effect as reinforcements increase is predicted because the probability of entering the buffer is reduced for an item receiving its third reinforcement; remember, an item recovered from LTS is not entered into the buffer. Thus as reinforcements increase the probability of being in the buffer decreases, and the normally increased probability of being in the buffer as a result of a short lag a is partially counterbalanced.

** Undoubtedly there are some selection effects in the data graphed in Figure 16, but their magnitude is difficult to determine. Thus, these data should be regarded with some wariness.

of illuminating the one aspect omitted, albeit intentionally, from the assumptions.

The aspect left out is, of course, that of LTS response competition, or negative transfer. The model fails to take account of this effect because it fails to keep track of residual LTS strength remaining as a result of the previous items using the same stimulus. This lack is most clearly indicated by the occurrence of intrusion errors; particularly errors which were correct responses on the preceding occurrence of that stimulus. For example, consider the following sequence:



Ittem 22-Z is studied for the j^{th} time and then tested at lag a; on this trial 22 is paired with a new response X and next tested at lag b. By an intrusion error we mean the occurrence of response Z when 22 is tested at the far right of the diagram. The model predicts that these intrusion errors will be at chance level (1/25), independent of lag and number of reinforcements. In fact, these predictions fail. Figure 17 presents the probability of intrusion errors as a function of lag b; where the data have been lumped over all values of lag a, three curves are plotted for $j = 1$, 2 and 3. This failure of the model is not very distressing because it was expected: the model could be extended in a number of obvious ways to take account of competing LTS traces without appreciably

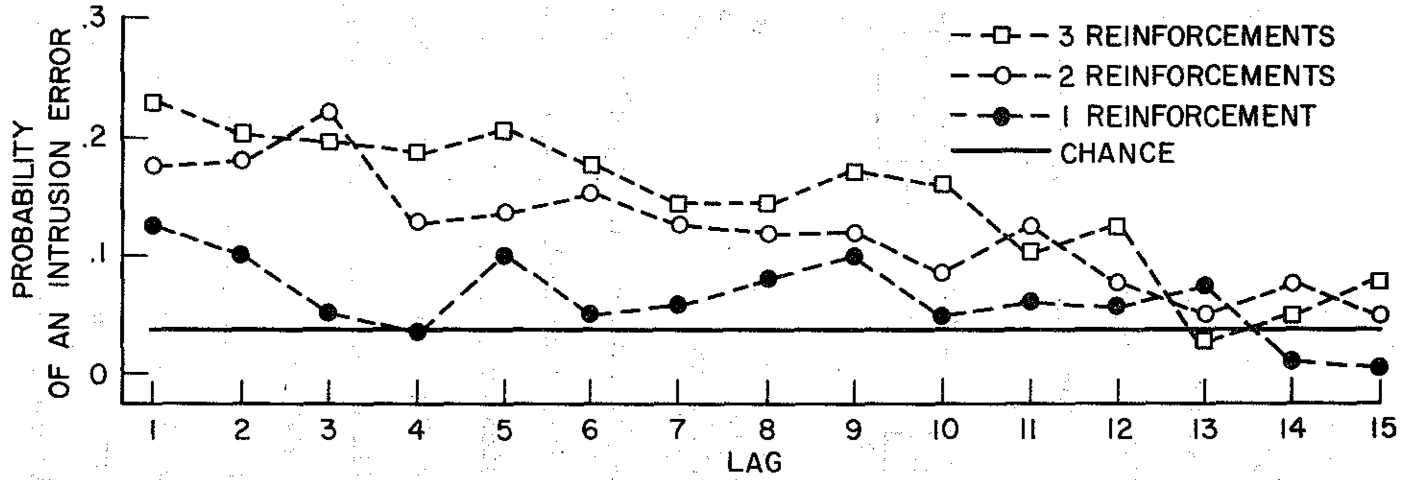


Figure 17 Probability that the correct response for the preceding item using the same stimulus will be given in error to the present item (Experiment 3)

changing any of the predictions so far presented. The extension has not been made because of our interest in this study is centered upon short-term effects.

Judging by the agreement between theory and data for each of the effects examined, the accuracy of the model is extremely good. It is interesting to note that the multiple-reinforcement procedure is not sufficient by itself to cause the subjects to switch their strategies from rehearsal to coding. The major emphasis still appears to be on rehearsal manipulations in STS, a not entirely surprising result since the situation is identical to that used in Experiment 1 except for the number of reinforcements given. The comments previously made concerning the difficulty associated with LTS storage in Experiment 1 apply here also. Because the emphasis is upon short-term mechanisms, this experiment is not to be considered in any strong sense as a bridge to the usual paired-associate learning situation. Nevertheless, a number of long-term effects, such as intrusion errors and interference caused by previously learned items on new items with the same stimulus, demonstrate that LTS mechanisms cannot be ignored in the theory. In Section 5 we consider experiments that are designed to provide a sharper picture of the workings of LTS; experimentally this is accomplished by systematically varying the number of items in LTS through which searches must be made. Before considering this problem, however, there are other features of the STS rehearsal strategy to be explored. We turn next to an experiment in which the probability of entering an item into the buffer is manipulated experimentally.

4.4. Overt vs. Covert Study Procedures (Experiment 4).

The statistics considered in the previous section leave little doubt about the role of O-items, N-items, and the buffer entry parameter α. But one question we have not considered is whether α is amenable to experimental manipulation; if the process is really under the control of the subject, such manipulation would be expected. We now turn to a study by Brelsford and Atkinson (in press) which was designed to answer this question.

In Experiment 1, the proportions of O-items and N-items were varied by changing the size of the stimulus set, and the predicted differences were found. Manipulating α, however, is a somewhat more subtle task since it is the subject's strategy that must be affected. One experimental device which seems likely to increase the probability of an item's entering the buffer is to have the subject recite the item aloud as it is presented for study; this will be referred to as the "overt" study procedure. The "covert" study procedure is simply a replication of the procedure used in Experiment 1 where the subject was not required to recite the item aloud when it was presented for study, but simply told to study it.

Method. The method was identical to that used in Experiment 1 except for the following changes. The size of the stimulus set was fixed at 6 for all subjects and sessions. Each session consisted of 200 trials divided into four 50-trial blocks alternating between the overt and covert conditions. The initial 50 trial block was randomly chosen to be either an overt or a covert condition. The covert condition was identical in all respects to Experiment 1; when the word "study" and an S-R pair appeared on the CRT (the display screen) the subjects were told to silently study the item being presented. In the overt blocks, instead of the word "study" appearing on the CRT during the study portion of a trial, the word "rehearse" appeared. This was a signal for the subject to recite aloud twice the item then being presented for study. This was the only difference from the procedure used during the covert trials. It was hoped that the act of repeating the items aloud would raise the subject's probability of entering the item into his rehearsal buffer.

Results. In order to allow for the subject's acclimation to a change in study conditions, the first 15 trials of each 50-trial block are not included in the data analysis. Figure 18 presents the lag curves for the overt and covert conditions. It is evident that performance is superior in the overt condition. Furthermore, the overt lag curve is S-shaped in form, an effect not observed in earlier curves. Since the parameters of the models will be estimated from these curves, the model is presented before considering additional data.

The model for the covert condition is, of course, identical to that used in the analysis of Experiment 1. It has the four parameters r, α, θ, and τ. Since it was hypothesized that α would be raised in the overt condition, we might try estimating α separately for that condition. This version of the model will not fit the overt data, however, because of the pronounced S-shaped form of the lag curve. Although setting α equal to 1.0 will predict better performance in the overt condition, the lag curve will have the form of an exponentially decreasing function, which is clearly not found in the data. In order

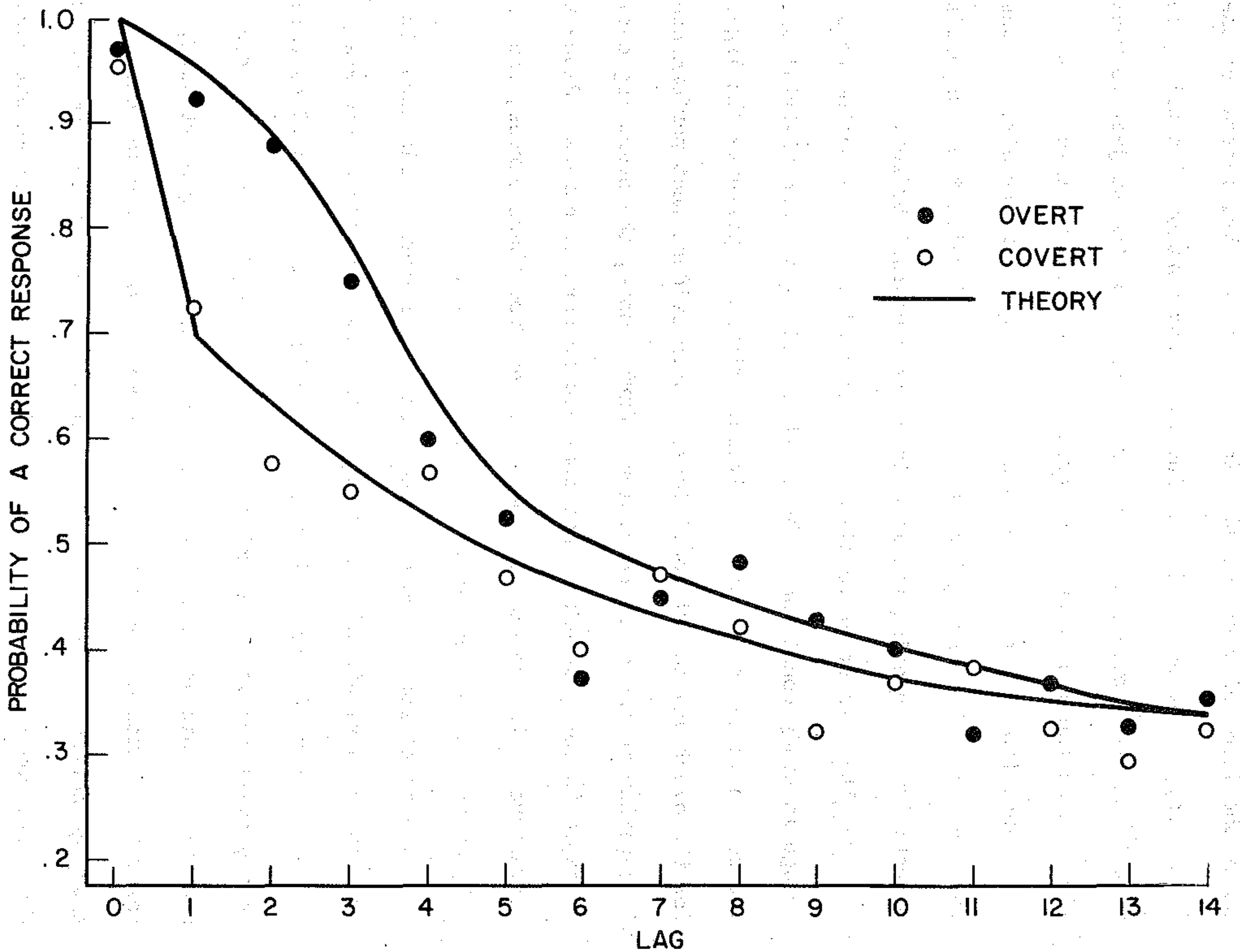


Figure 18 Observed and theoretical probabilities of a correct response as a function of lag (Experiment 4)

to account for the S-shaped curve, we need to assume that in the overt condition the subject tends to knock the oldest items out of the buffer first. In the model for the covert case, an entering N-item is said to knock out at random any item currently in the buffer. It will be assumed for the overt case that an entering N-item tends to replace the oldest item in the buffer; remember O-items are items whose stimulus is currently in the buffer and they automatically replace the item with that stimulus. This probability of knocking the oldest items from the buffer first is specified as follows: if there are r items in the buffer and they are numbered so that item 1 is the oldest and item r is the newest, then the probability that an entering N-item will knock the j^{th} item from the buffer is

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

This equation is derived from the following scheme. The oldest item is knocked out with probability δ. If it is not knocked out, then the next oldest is knocked out with probability δ. The process continues cyclically until an item is finally selected to be knocked out. When δ approaches zero, the knockout probabilities are random, as in the covert case. When δ is greater than zero there will be a tendency for the oldest items to be knocked out of the buffer first; in fact if δ equals one, the oldest item will always be the one knocked out. It should be clear that the higher the value of δ, the greater the S-shaped effect predicted for the lag curve.

The model for the curves in Figure 18 is therefore structured as follows. The parameters r, θ, and τ will be assumed to be the same for the two conditions; the parameters α and δ will be assumed to be affected by the experimental manipulation. To be precise, in the covert case α will be estimated freely and δ will be set equal to zero, which is precisely the model used in Experiment 1. In the overt case, α will be set equal to 1.0, which means that every item enters the buffer, and δ will be estimated freely. The parameter values that provide the best χ^2 fit to the data in Figure 30 were $r = 3$, $\theta = .97$, $\tau = .90$; for the covert condition the estimate of α was .58 (with δ equal to zero) and for the overt condition the estimate of δ was .63 (with α equal to one). The predictions for this set of parameter values are shown in Figure 18 as smooth curves. The improvement in performance from the covert to overt conditions is well predicted; actually it is not obvious that variations in either α or δ should affect the overall level of performance. The principal reason for the improvement is due to the value of α; placing every item into the buffer means that an item entering the buffer will be expected to stay there for a shorter period than if some items did not enter the buffer. This shorter period in the buffer, however, is outweighed by the advantages resulting from the entry of every item in the first place. It is not easy to find statistics, other than the gross form of the lag curve, which reflect changes in δ; thus the assumption that the oldest items are lost first is not easy to verify in a direct way. Nevertheless, it is quite common to find experiments that yield S-shaped recency curves and these results can be fit by assuming that the oldest items in the buffer tend to be knocked out first. Other examples will be presented in Section 5.

A number of additional aspects of the data will now be examined.

First we consider the "all-same" and "all-different" lag curves. Figure 19 gives the "all-same" lag curves for the overt and covert conditions. This curve gives the probability of a correct response for an item when all of the intervening items (between its study and test) have the same stimulus. This curve will be quite flat because the items following the first intervening item tend to be 0-items which will not knock other items from the buffer (for the overt case, <u>every</u> item following the first intervening item is an 0-item, since all items enter the buffer). Figure 19 also presents the "all-different" lag curves. This curve is the probability of making a correct response to a given item when the other items intervening between its study and test all involve different stimuli. The predictions generated by the previous parameter values are given by the smooth curves; they appear to be quite accurate.

We now look for an effect that will be sharply dependent upon the value of α and hence differ for the overt and covert conditions. Such an effect is given in Figure 20; graphed there is the probability of a correct response as a function of the number of immediately preceding items having the same stimulus as the item in question. This is the same statistic that is plotted in Figures 9 and 14; it is not a lag curve because the probability correct is given as an average over all possible lags at which the item was tested. If α is less than one, then the length of the preceding sequence of items with the same stimulus will be an important variable; since any item in the sequence which enters the buffer will cause every succeeding item in the sequence to enter the buffer, the

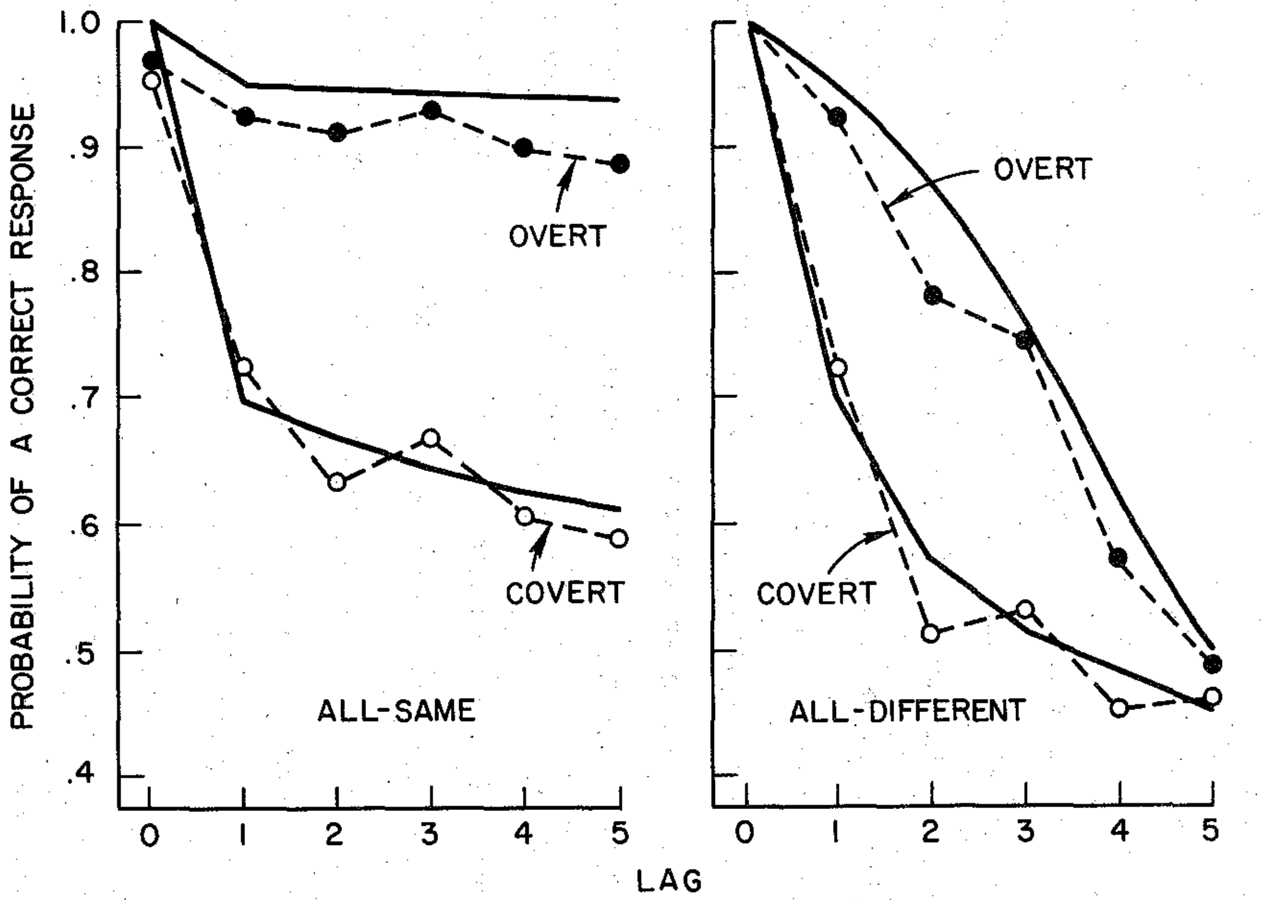


Figure 19 Observed and theoretical probabilities of a correct response as a function of lag for the "all-same" and "all-different" conditions (Experiment 4)

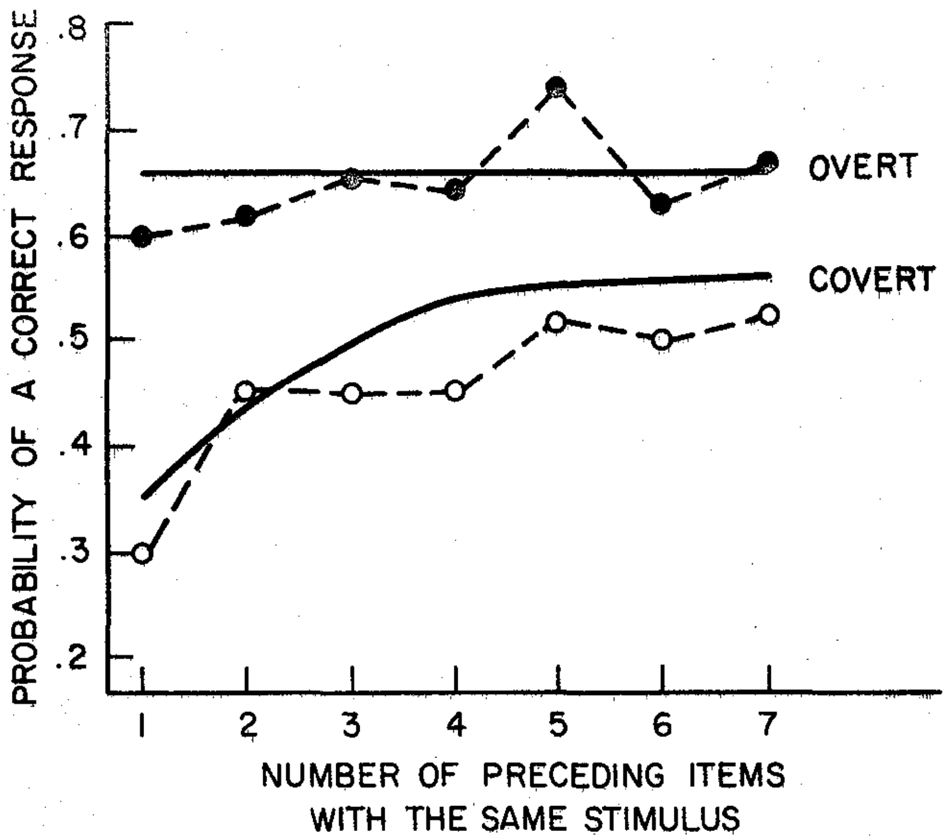


Figure 20 Observed and theoretical probabilities of a correct response as a function of the number of consecutive preceding items all using the same stimulus (Experiment 4)

probability that the item in question enters the buffer will approach one as the length of the preceding sequence of items all using the same stimulus increases. For α equal to one (overt condition), every item enters the buffer and therefore no change would be expected. As indicated in Figure 20, the data and theory are in good agreement. The slight rise in the data points for the overt condition may indicate that an estimate of α a little below 1.0 would improve the predictions, but the fit as it stands seems adequate.

4.5 Additional Variables Related to the Rehearsal Buffer (Experiments 5, 6, and 7).

Known Items and the Buffer (Experiment 5). In this section we shall consider briefly a number of other variables that relate to the rehearsal buffer. The overt manipulation in the preceding section succeeded in raising to near 1.0 the probability of entering an item in the buffer. As an alternative, one would like an experimental manipulation which would cause the entry probability to drop to near zero for some items. W. Thomson at Stanford University has performed an experiment that satisfies this requirement. The experimental manipulation involves interspersing some extremely well-known items among a series of items never seen before. The assumption is that a well-known item will not enter the rehearsal buffer. The experiment was performed using a modification of the "all-different" stimulus procedure employed in Experiment 2. The stimuli were consonant-vowel-consonant trigrams and the responses were the digits 0-9. For each subject two stimuli were chosen at the start of the first session and assigned responses. These S-R pairs never changed throughout the series of sessions. Except for these two items all other items were presented just once. The size of the to-be-remembered set (s) was 6 which included the two "known" items. The presentation schedule was as follows: on each trial with probability .5 one of the two known items would be presented for test and then given yet another study period; otherwise one of the four items in the current to-be-remembered set would be tested and a new stimulus-response pair then presented for study. Thus, the task was like that used in Experiment 2, except that on half the trials the subject was tested on, and then permitted to study, an S-R pair which was thoroughly known. The data from the first session in which the known items were being learned will not be considered.

The simplest assumption regarding the two known items is that their probability of entering the buffer is zero. This assumption is the one used in the multiple-reinforcement study (Experiment 3); namely, that an item successfully recovered from LTS is not entered into the buffer.* In contrast to Experiment 3, in this study it is easy to identify the items that are known since they are experimentally controlled; for this reason we can look at a number of statistics depending upon the likelihood of entering known items into the buffer. The one of particular interest is presented in Figure 21. Graphed there is the unconditional lag curve, the "all-known-intervening" lag curve and the "all-unknown-intervening" lag curve. By known items we mean the two S-R pairs that repeatedly are being studied and tested;

* Underwood and Ekstrand (1967) have found that insertion of known items from a previously learned list into a succeeding list improves performance on the learning of unknown items on the second list, although list length was a confounded variable.

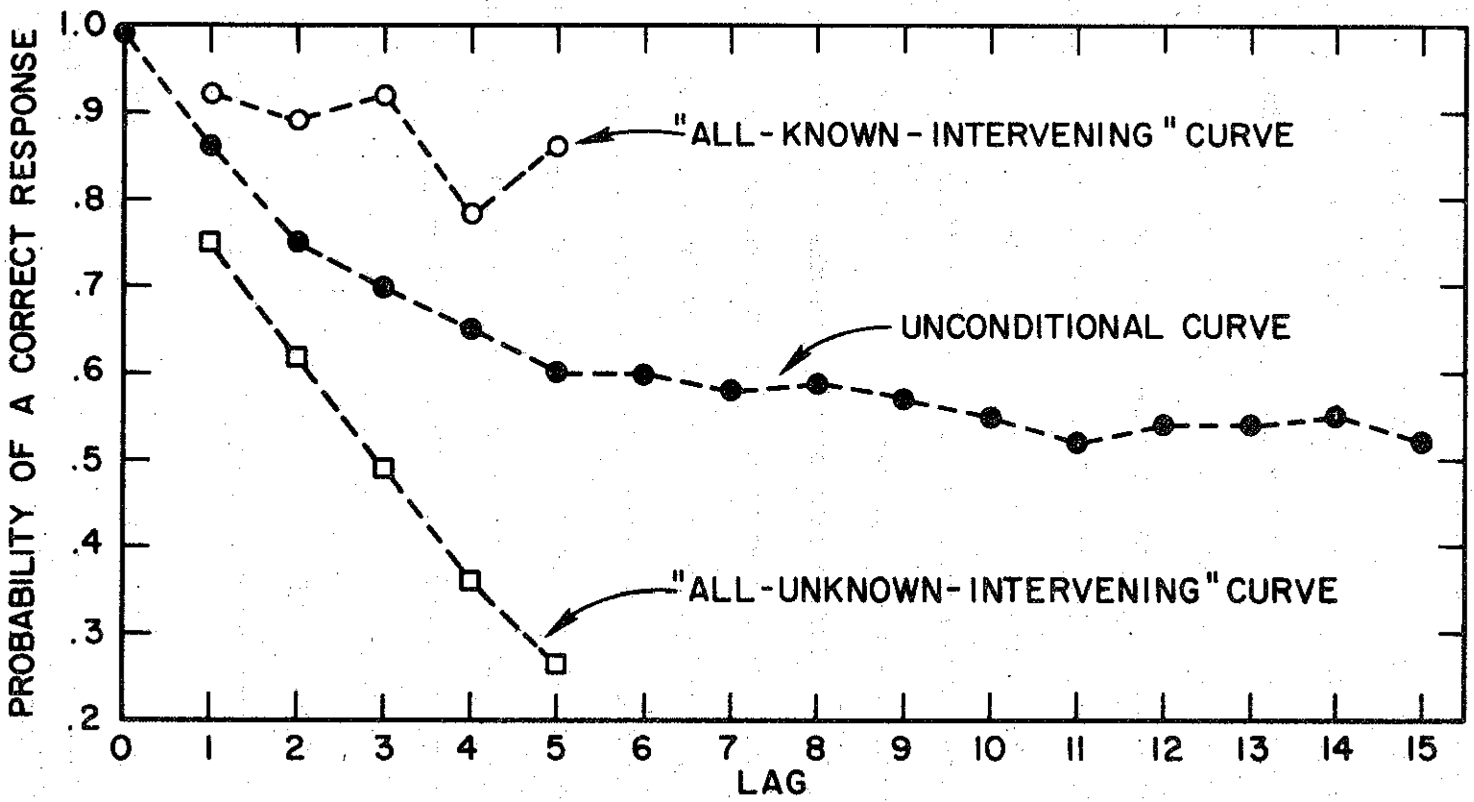


Figure 21 Observed and theoretical probabilities of a correct response as a function of lag, for the overall condition and for the "all-known-intervening" and "all-unknown-intervening" conditions (Experiment 5)

by unknown items we mean those pairs that are studied and tested only once. The unconditional lag curve gives the probability correct for unknown items as a function of lag, independent of the type of items intervening between study and test; of course, the corresponding curve for known items would be perfect at all lags since subjects never make errors on them. The all-known-intervening curve gives the probability correct as a function of lag, when all of the items intervening between study and test are known items. If none of the known items enter the buffer, this curve should be level from lag one on and equal to α, the probability that the item entered the buffer when presented for study. At the opposite extreme is the all-unknown-intervening curve; when all the intervening items are new, the probability of knocking the item of interest from the buffer increases with lag and therefore the curve should decay at a rapid rate. It may be seen that this curve indeed drops at a more rapid rate than the unconditional lag curves. The marked difference between the all-known and all-unknown curves in Figure 21 leads us to conclude that known and unknown items clearly have different probabilities for entering the rehearsal buffer. If the all-known curve were flat after lag 1, then the probability for entering a known item into the buffer would be zero. Another possibility is that α is indeed zero for known items, but that the subject occasionally picks an item from LTS for additional rehearsal when a known item is presented.

Response Time Measures (Experiment 6). We now turn to a consideration of some latency results. Potentially, latencies offer an avenue of analysis that could be more fruitful than the analysis of choice response data; we say this because the latencies should reflect search and retrieval times from both STS and LTS. A detailed latency analysis is beyond the scope of this paper, but one simple result will be considered. Figure 22 presents the average latencies as a function of lag for correct and incorrect responses in a study by Brelsford, Keller, Shiffrin and Atkinson (1966). This experiment employed the same procedure described earlier in our discussion of Experiment 1 except that only 6 rather than 26 responses were used. As in Experiment 1, this study used three different stimulus-set sizes; i.e., s equalled 4, 6 or 8. For each stimulus set in Figure 22 it may be seen that the correct and incorrect latency curves converge at long lags. This convergence would be expected since the probability of a correct response is dropping toward chance at long lags. The theoretical curves are based on an extremely simple latency model which assumes that latencies for responses correctly retrieved from either LTS or STS have a fixed mean value λ, whereas a failure to retrieve and a subsequent guess has a fixed mean value of λ'. Thus error responses always have a mean latency λ'; however, a correct response may occur as a result of a retrieval from memory or a correct guess, and consequently its latency is a weighted average of λ and λ'. We can estimate λ' as the average of the points on the latency lag curve for errors, and λ can be set equal to the latency of a correct response at lag zero since all responses are due to retrievals from memory at this lag. In order to predict the remaining latency data, we make use of the observed probability of a correct response as a function of lag; these values are reported in Brelsford, Keller, Shiffrin and Atkinson (1966). If p_i is the observed probability of a correct response at lag $\underline{i}$, then

$$p_i = x_i + (1-x_i)\frac{1}{6}$$

where x_i is the probability of retrieving the response from memory and $(1-x_i)\frac{1}{6}$ is the probability of making a correct response by guessing. Estimating x_i in this way, we predict that the mean latency of a correct response at lag i is simply $x_i\lambda + (1-x_i)\lambda'$. Using this equation and estimating λ and λ' as indicated above, leads to the theoretical curves displayed in Figure 22. The error latency curve is predicted to be equal to λ' for all lags, whereas the correct latency curve is λ at lag 0 and approaches λ' over lags as the estimate of x_i goes to zero. This latency model is of course oversimplified, and fails to take into account differences in latencies due to retrieval from STS as compared to retrieval from LTS; the results nevertheless indicate that further analyses along these lines may prove fruitful.

Time Estimation (Experiment 7). One factor related to our model that has not been discussed is temporal memory. It seems clear that there is some form of long-term temporal memory; in a negative transfer paradigm, for example, there must be some mechanism by which the subject can distinguish between the most recent response paired with a stimulus versus some other response paired with that stimulus at an earlier time. This temporal memory undoubtedly involves the long-term store; somehow when an event is stored in LTS it also must be given a time tag or stored in such a way that the subject can date the event (albeit imperfectly) at the time of retrieval. In addition to long-term temporal storage, there is evidence that a subject's estimate of elapsed time depends upon an item's length of residence in the buffer. An experiment by R. Freund

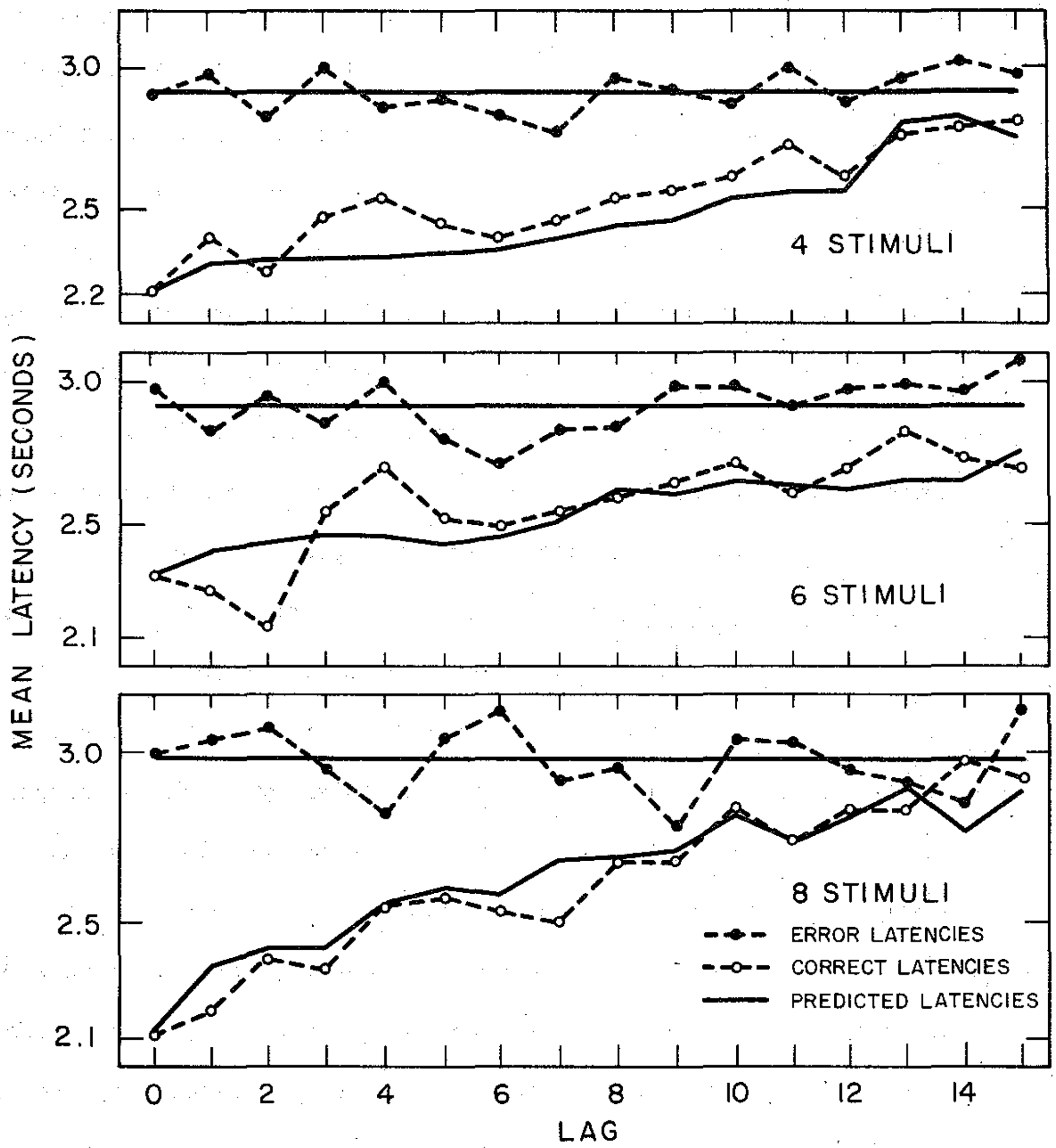


Figure 22 Observed and theoretical mean latencies as a function of lag for correct and incorrect responses (Experiment 6)

and D. Rundus at Stanford University serves to illustrate the dependence of temporal memory upon the buffer.* The study employed essentially the same procedure used in Experiment 2. There was a continuous sequence of test-plus-study trials and the stimuli kept changing throughout each session; each stimulus appeared only once for study and test. The stimuli were consonant-vowel-consonant trigrams and the responses were the 26 letters of the alphabet; the size of the to-be-remembered set of items was fixed at 8. When a stimulus was tested the subject first gave his best guess of the response that had been previously studied with the stimulus and then gave an estimate of the number of trials that intervened between the item's initial study and final test; this estimate could range from 0 to 13; if the subject felt the lag was greater than 13 he responded by pressing a key labeled 14+.

The unconditional lag curve for the probability of a correct response is presented in Figure 23. The solid line represents the predictions that were generated by the model used to fit Experiment 2. The parameter values providing the best fit to the lag curve were $r = 2$, $\alpha = .57$, $\theta = .13$, $\tau = 1.0$. The data of interest is presented in Figure 24. The average lag judgment is plotted as a function of the actual lag. The solid dots are the average lag judgments for those items to which a correct response was given; the open circles are the average lag judgments for those items to which an incorrect response was given. If lag judgments were perfect, they would fall on the 45° diagonal; it may be seen that the correct curve

* This study employs a time-estimation procedure similar to one developed by L. R. Peterson (personal communication).

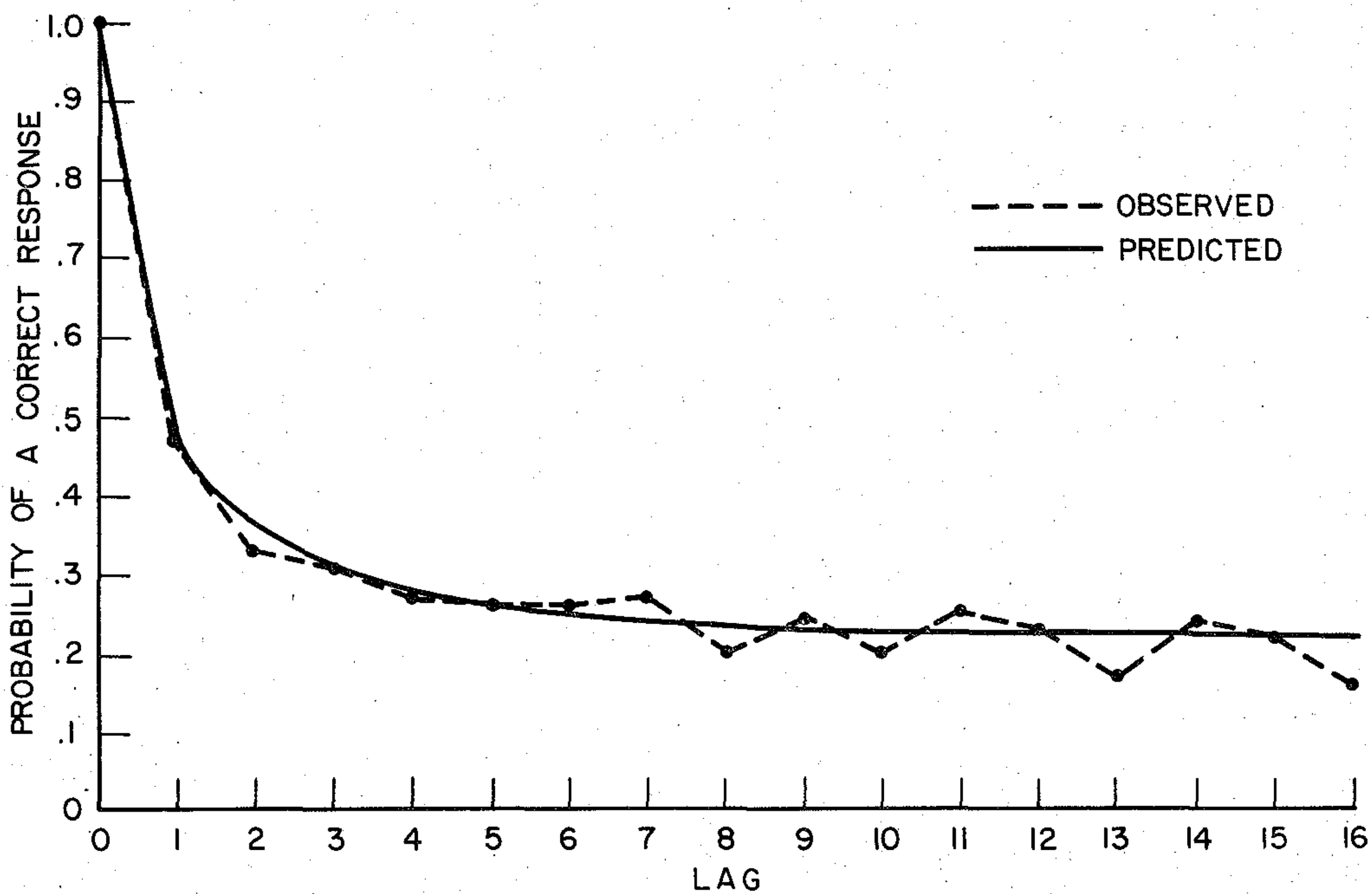


Figure 23 Observed theoretical probabilities of a correct response as a function of lag (Experiment 7)

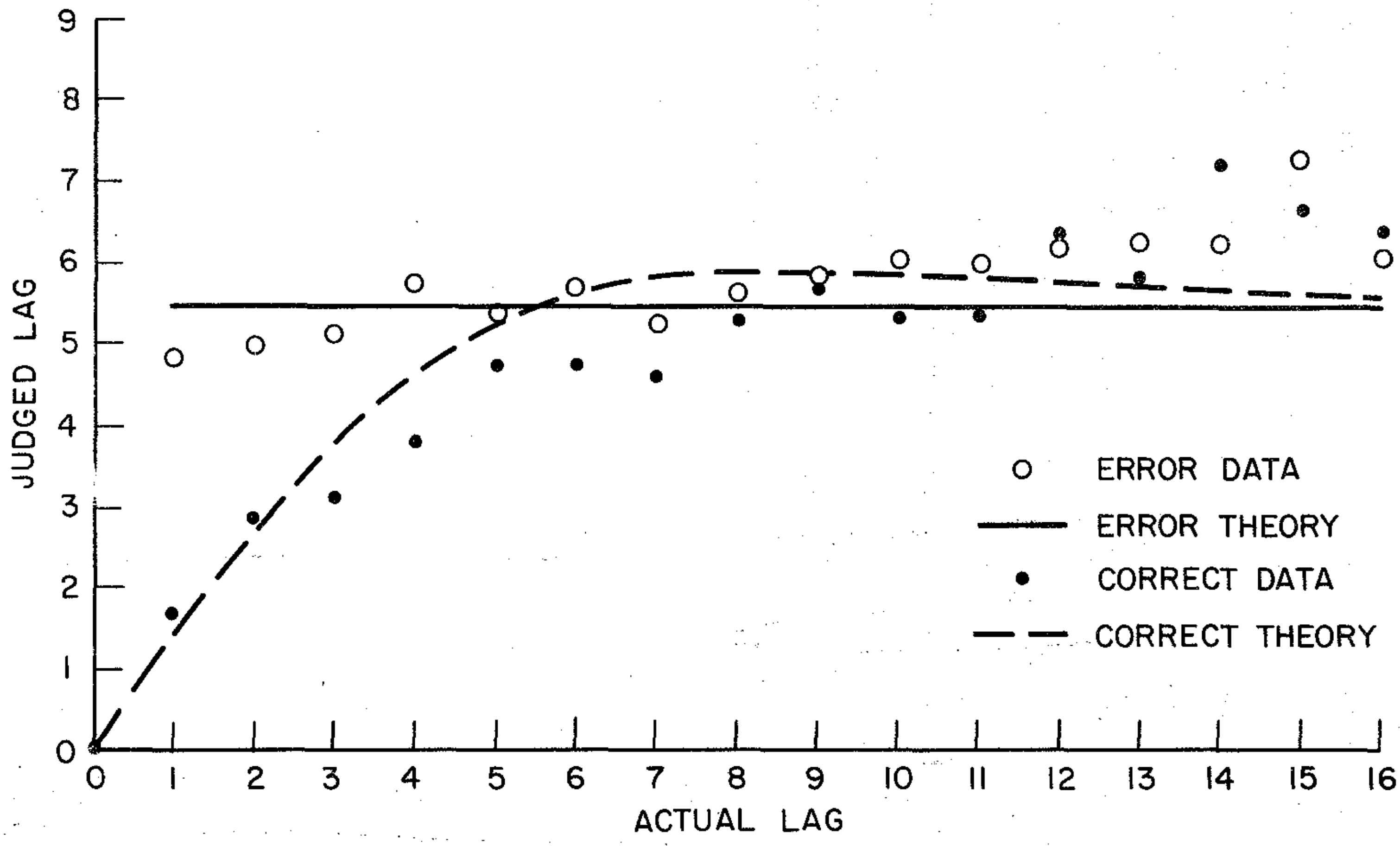


Figure 24 Observed and theoretical mean lag judgments as a function of the actual lag (Experiment 7)

is fairly accurate to about lag 5 and then tails off. The lag judgments associated with incorrect responses seem to be virtually unrelated to the actual lag. This indicates that the retrieval of a correct response and temporal estimation are closely related. An extremely simple model for this data assumes that the mean lag judgment for an item in the buffer is the true lag value; any item not in the buffer is given a lag judgment at random from a distribution that is unrelated to the true lag. The predictions using the above parameter estimates are shown in Figure 24. Freund and Rundus have developed more elaborate models which include both a long- and short-term temporal memory and have obtained quite accurate predictions; but these models will not be examined here. The point we want to make by introducing these data is that temporal memory may be tied to the short-term system even more strongly than to the long-term system.

SECTION 5. EXPERIMENTS CONCERNED WITH LONG-TERM SEARCH AND RETRIEVAL

The major purpose of this section is to examine a series of experiments concerned with search and retrieval processes in LTS. These experiments differ from those of the preceding section in that the memory tasks are not continuous; rather, they involve a series of discrete trials which are meant to be relatively independent from one to the next. On each trial a new list of items is presented sequentially to the subject for study; following the presentation a test is made on some aspect of the list. Using this procedure the size of the list, d, can be systematically manipulated. Variations in list size affect the size of the memory set through which the subject must search when tested, and consequently search and retrieval processes can be examined in more detail than was previously possible. The title of this section is not meant to imply, however, that the short-term processes involved in these experiments are different from those appearing in the continuous-presentation situations; in fact, the models used to describe the experiments of this section will be based upon the same STS rehearsal buffer introduced earlier. The difference is one of emphasis; the long-term processes will be elaborated and explored in greater depth in this section. This exploration of long-term models will by no means be exhaustive, and will be less extensive than that carried out for the short-term processes.

Prior to an examination of particular experiments, a few remarks need to be made about the separability of lists. In any experiment where a series of different lists is presented, we may ask just what information in LTS the subject is searching through at test. The same problem arises, though less seriously, in experiments where the subject is tested on only one list. Clearly the information relevant to the current list of items being tested must be kept separate from the great mass of other information in LTS. This problem is accentuated when individual lists within a session must be kept separated. How this is managed is somewhat of a mystery. One possible explanation would call for a search along a temporal memory dimension: the individual items could be assumed to be temporally ordered, or to have "time tags." It is not enough to propose that search is made through all items indiscriminately and that items recovered from previous lists are recognized as such and not reported; if this were true, the duration and difficulty of the search would increase dramatically over the session. In fact, the usual result is that there is little change in performance over a session except for effects concentrated at the very start. On the other hand, judging from such factors as intrusion errors from previous lists, the subject is not able to restrict his search solely to the current list. In the experiments to follow, we will make the simplifying assumption, without real justification, that the lists are entirely separated in LTS, and that the subject searches only through information relevant to the list currently being tested.

5.1. A Serial Display Procedure Involving Single Tests (Experiment 8).

This experiment involved a long series of discrete trials. On each trial a new display of items was presented to the subject. A display consisted of a random sequence of playing cards; the cards varied only in the color of a small patch on one side; four colors (black, white, blue, and green) were used. 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 it was turned face down on a table 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 table: the card presented first was to the subject's left and the most recently presented card to the right. The trial terminated when the experimenter pointed to one of the cards on the table and the subject attempted to recall the color of that card. The subject was instructed to guess the color if uncertain and to qualify the response with a confidence rating. The confidence ratings were the numerals 1 through 4. The subjects were told to say 1 if they were positive; 2 if they were able to eliminate two of the four possible colors as being incorrect; 3 if one of the four colors could be eliminated as incorrect; and 4 if they had no idea at all as to the correct response.

It is important to note that only one position is tested in a display on each trial. The experiment involved 20 female subjects who participated in five daily sessions, each lasting for approximately one hour. Over the course of the five sessions, a subject was given approximately 400 trials. The display size, d, was varied from trial to trial and took on the following values: d = 3, 4, 5, 6, 7, 8, 11 and 14. Details of the experimental procedure are presented in Phillips, Shiffrin and Atkinson (1967).

Figure 25 presents the probability of a correct response at each serial position for displays of size 5, 6, 7, 8, 11 and 14. For displays of sizes 3 and 4, the probability correct was 1.0 at all positions. The

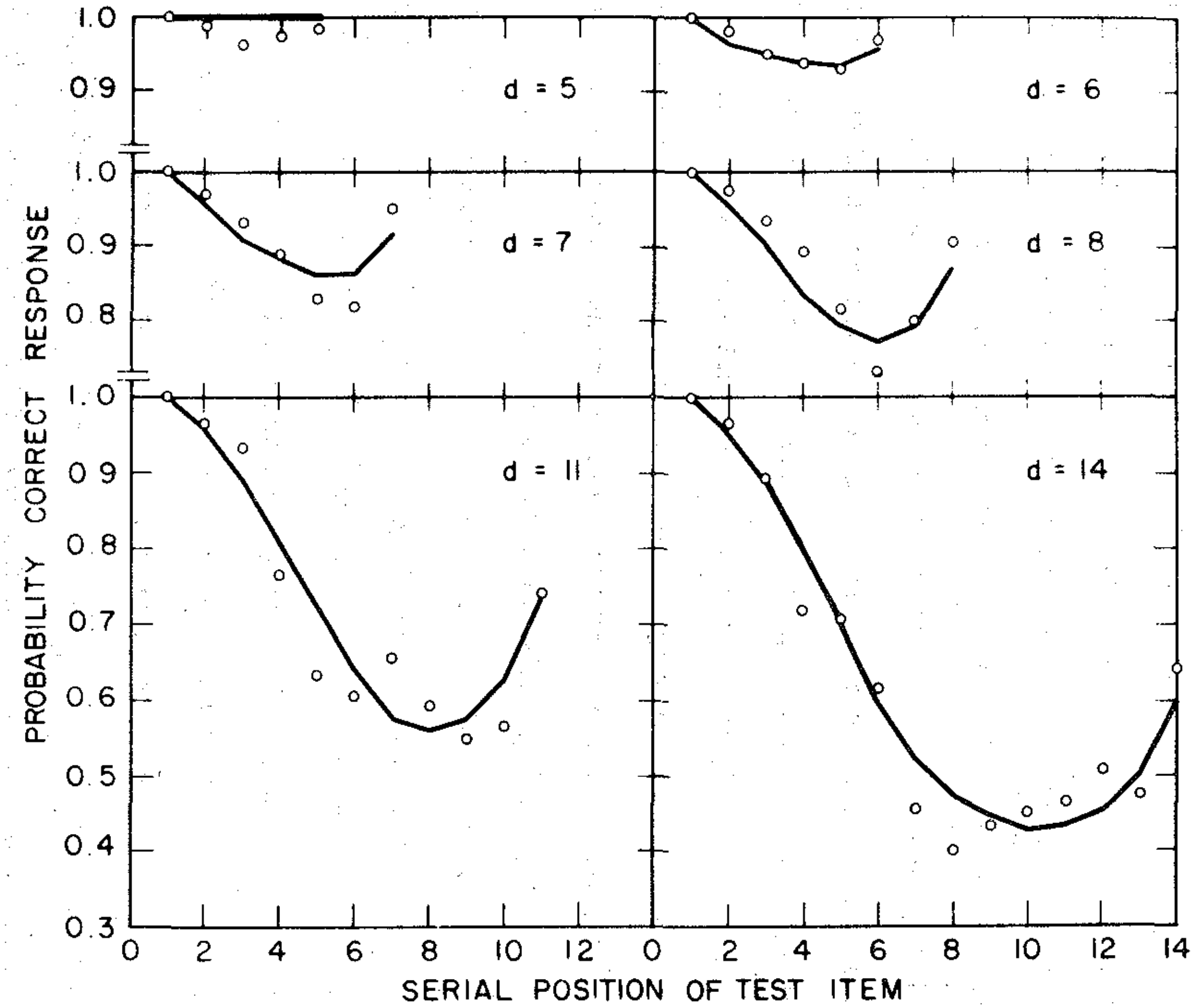


Figure 25 Observed and theoretical probabilities of a correct response as a function of serial position (Experiment 8)

circles in the figure are the observed points; the solid lines are predicted curves which will be explained shortly. The serial positions are numbered so that item 1 designates the last item presented (the newest item), and item d designates the first item presented (the oldest item). The most apparent features of the curves are a fairly marked S-shaped recency portion and a smaller, quite steep primacy portion. For all display sizes, the probability of a correct response is 1.0 at serial position 1.

Theory. We must first decide whether a subject will set up and use a rehearsal buffer in this situation. Despite the fact that the continuous procedure has been dropped, it is still unlikely that the subject will engage in a significant amount of long-term coding. This is true because the task is still one of high "negative transfer"; the stimuli, which are the positions in the display, are constantly being re-paired with new responses as a session continues. Too much LTS encoding would undoubtedly lead to a high degree of interference among lists. It is only for a relatively weak and decaying LTS trace that a temporal search of long-term memory may be expected to keep the various lists separate. This difficulty in LTS transfer leads to the adoption of short-term strategies. Another reason for using a rehearsal buffer in this task depends upon the small list lengths employed; for small list lengths, there is a high probability that the item will be in the buffer at the moment of test. Thus the adoption of a rehearsal buffer is an efficient strategy. There is some question concerning just what the unit of rehearsal is in this situation. For example, the subject could assign numbers to positions in the display and then rehearse the number-color pairs. Most likely, however, the subject uses the fact that the stimuli always remain before her to combine STS rehearsal with some form of visual mnemonic. That is, the unit of rehearsal is the response alone; as the subject rehearses the responses, she "mentally" places each response upon the appropriate card before her. This might therefore be a situation where the a-v-l and visual short-term stores are used in conjunction with each other. In any case, it seems reasonable that the units of rehearsal are the names (or perhaps the abbreviations) of the colors.

We might ask how the buffer will act in this situation. As noted earlier, in reference to the "overt-covert" experiment, the fact that items are read aloud as they are presented will tend to cause the subject to enter each item into the buffer. Furthermore, an S-shaped recency effect would not be unexpected. Indeed, if the units of rehearsal are the responses themselves, then the subject might tend to keep them in consecutive order to ease the visual memory task; if all items enter the buffer and are kept in consecutive order, then the oldest items will tend to be deleted first. That is, when a new item enters the buffer there will be a tendency to eliminate the oldest item from the buffer to make room for it. One other question that should be considered is the size of the buffer the subject would be expected to use in this task. There are a number of reasons why the buffer size should be larger here than in the continuous tasks of Section 4. First, the subject is not continually being interrupted for tests as in the previous studies; more of the subject's attention may therefore be alloted to rehearsal. Second, rehearsal of color names (or their abbreviations) is considerably easier than number-letter combinations. Equivalent to rehearsing "32-G, 45-Q" might be "Black, White, Black, Green" (or even a larger set if abbreviations are used). The magnitude of the difference may not be quite as large as this argument would lead us to expect because undoubtedly some time must be alloted to keeping track of which response goes on which position, but the estimate of the buffer size nevertheless should be larger in this situation than in the continuous tasks.

The STS part of the model for this experiment is similar to that used in the "overt" experiment in Section 4.4 in that every item is entered in the buffer when it is presented. There is one new factor, however, that must be considered. Since each trial starts with the buffer empty, it will be assumed that the first items presented enter the buffer in succession, without knocking any item out, until the buffer is filled. Once the buffer is filled, each item enters the buffer and knocks out one of the items currently there. If the most recently presented item is in slot r of the buffer, and the oldest item is in slot 1, then the probability that the item in slot i of the buffer will be the one eliminated is

$$\frac{\delta(1-\delta)^{i-1}}{1-(1-\delta)^{r}} .$$

This is the same equation that was used to describe the knock-out process for the overt-covert study (Experiment 4). The larger δ, the greater the tendency to delete the oldest item in the buffer when making room for a new one.

The first set of long-term storage and retrieval assumptions that will be considered are essentially identical to those used in the previous sections. Information will be assumed to enter LTS during the entire period an item resides in the buffer at a rate θ per inter-item interval. This process must be qualified with regard to the first few items presented on each trial before the buffer is filled; it is assumed that the subjects divide their attention equally among the items in the buffer. Thus, if the rate of transfer is θ when there is only one item in the buffer, and the buffer size is r, then the rate of transfer will be θ/r when the buffer is filled. That is, since attention must be divided among r items when the buffer is full, each item receives only $1/r^{th}$ as much transfer as when the buffer only holds a single item. In general, information on each item will be transferred to LTS at rate θ/j during the interval in which there are j items in the buffer. The effect of this assumption is that more information is transferred to LTS about the items first presented in a list than about later items that are presented once the buffer is full.

The LTS decay and retrieval processes must now be examined. In earlier experiments we assumed that information decayed solely as a result of the number of items intervening between study and test; in other words, only the retroactive interference effect was considered. Because the previous tasks were continuous, the number of items preceding an item's presentation was effectively infinite in all cases. For this reason the proactive effects were assumed to be constant over conditions and did not need explicit inclusion in the model. In the present experiment the variation in list size makes it clear that proactive interference effects within a trial will be an important variable. The assumption that will be used is perhaps the simplest version of interference theory possible: each preceding and each succeeding item has an equal interfering effect. To be precise, if an amount of information I has been transferred to LTS for a given item, then every other item in the list will interfere with this information to the extent of reducing it by a proportion τ. Thus, if there were d items in the list, the item of interest would have an amount of information in LTS at the time of test equal to $I(\tau^{d-1})$. Clearly the longer the list the greater the interference effect.

The model can now be completed by specifying the response process which works as follows. An item in the buffer at the time of test is responded to correctly. If the item is not in the buffer, then a search is made in LTS. The probability of retrieving the appropriate response is, as in our other models, an exponential function of this information and equals $1 - \exp[-I(\tau^{d-1})]$; if a retrieval is not made, then the subject guesses.

Data Analysis. The parameter values that gave the best fit to the data of Figure 25 using a minimum χ^2 criterion were as follows: $r = 5$, $\delta = .38$, $\theta = 2.0$, and $\tau = .85$.* Remember that r is the buffer size, δ determines the probability of deleting the oldest item in the buffer, θ is the transfer rate to LTS, and τ is the proportional loss of information caused by other items in the list. The theoretical curves generated by these parameter estimates are shown in Figure 30 as solid lines. The predictions are quite accurate as indicated by a χ^2 value of 44.3 based on 42 degrees of freedom. It should be emphasized that the curves in the figure were all fit simultaneously with the same parameter values.

The primacy effect in the curves of Figure 25 is predicted because more information is transferred to LTS for the first items presented on

* For details on the method of parameter estimation see Phillips, Shiffrin and Atkinson (1967).

each trial. There are two reasons for this. First, the transfer rate on any given item is higher the fewer items there are in the buffer; thus the initial items, which enter the buffer before it is filled, accumulate more information in LTS. Second, the initial items cannot be knocked out of the buffer until the buffer is filled; thus the time period that initial items reside in the buffer is longer on the average than the time for later items. The recency effect is predicted because the last items presented in a list tend to be still in the buffer at the time of test; the S-shape arises because the estimate of δ indicates a fairly strong tendency for the oldest items in the rehearsal buffer to be eliminated first when making room for a new item.

Having estimated a set of parameter values that characterizes the data in Figure 25, we now use these estimates to predict the confidence rating data. Actually, it is beyond the scope of this paper to analyze the confidence ratings in detail, but some of these data will be considered in order to illustrate the generality of the model and the stability of the parameter estimates. The data that will be considered are presented in Figure 26; graphed is the probability of giving confidence rating R_1 (most confident) for each list size and each serial position. The observed data is represented by the open circles. It is clear that these results are similar in form to the probability correct curves of Figure 25. The model used to fit these data is quite simple. Any item in the buffer is given an R_1. If the item is not in the buffer, then a search is made of LTS. If the amount of information in LTS on the item is $I(\tau^{d-1})$ then the probability of giving R_1 is an exponential function of that information: namely the function $1 - \exp[-c_1 I(\tau^{d-1})]$, where c_1 is a parameter

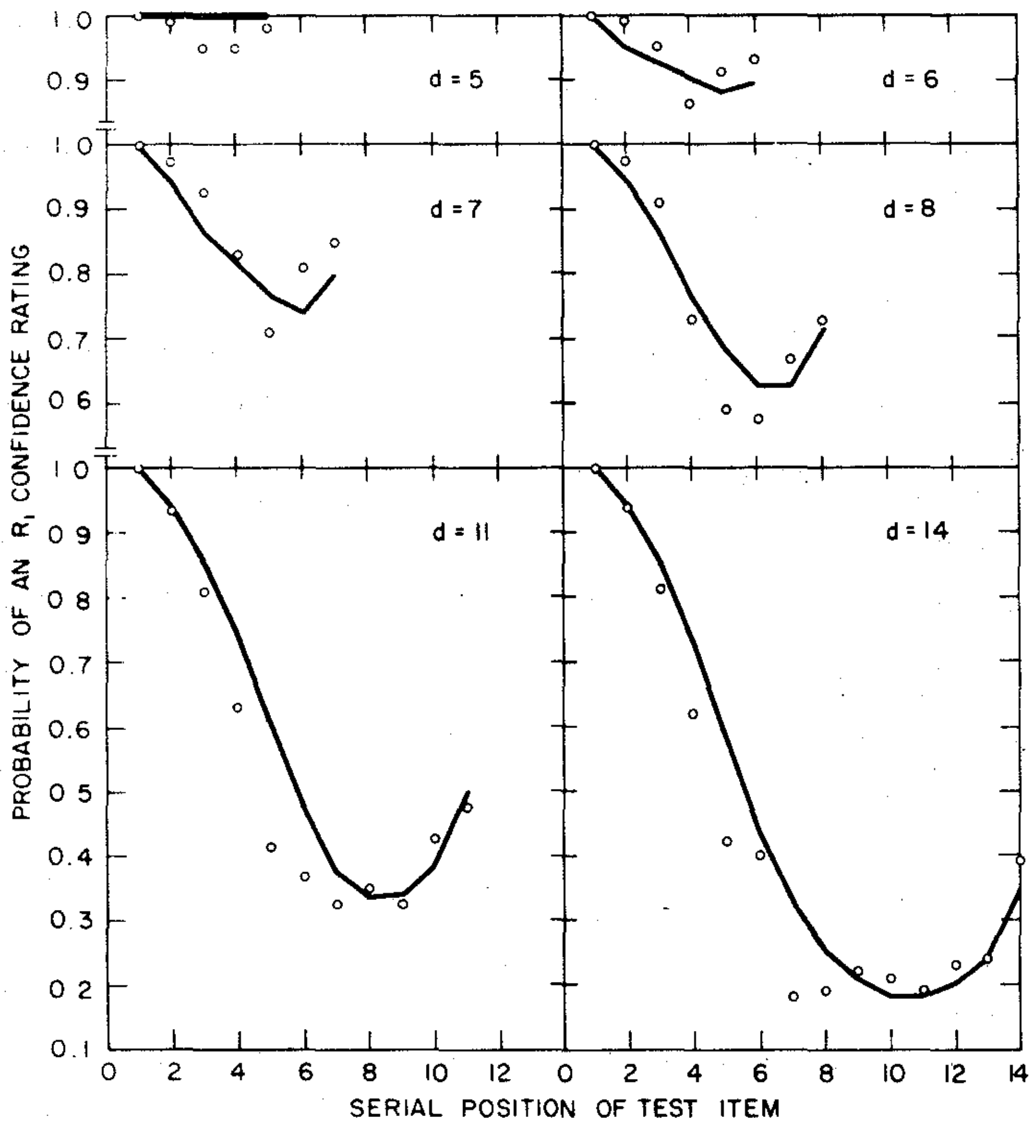


Figure 26 Observed and predicted probabilities of a confidence rating R_1 as a function of serial position (Experiment 8)

determining the subject's tendency to give confidence rating R_1. This assumption is consistent with a number of different viewpoints concerning the subject's generation of confidence ratings. It could be interpreted equally well as an assignment of ratings to the actually perceived amount of information in LTS, or as a proportion of the items that are recovered in an all-or-none fashion.* In any event, the predictions were generated using the previous parameter values plus an estimate of c_1. The predicted curves, with c_1 equal to .66, are shown in Figure 26. The predictions are not as accurate as those in Figure 25; but, considering that only one new parameter was estimated, they are quite good.

Discussion. In developing this model a number of decisions were made somewhat arbitrarily. The choice points involved will now be considered in greater detail. The assumption that the amount of transfer to LTS is dependent upon the number of items currently in the buffer needs elaboration. Certainly if the subject is engaged in coding or other active transfer strategies, the time spent in attending to an item should be directly related to the amount of transfer to LTS. On the other hand, the passive type of transfer which we assume can occur in situations where the subject makes use of a rehearsal buffer may not be related to the time spent in rehearsing an item per se, but rather to the total period the item resides in the buffer. That is, direct attention to an item in STS may not be necessary for some transfer to take place; rather a passive form of transfer may occur as long as the item remains in STS. Thus in situations where the rehearsal buffer is used and active transfer strategies

* The various possibilities may be differentiated through an analysis of conditional probabilities of the ratings given correct and incorrect responses, and through ROC curve (Type II) analyses (Murdock, 1966; Bernbach, 1967 a) but this will not be done here.

such as coding do not occur, it could reasonably be expected that the amount of information transferred to LTS would be related solely to the total time spent in the buffer, and not to the number of items in the buffer at the time. In practice, of course, the actual transfer process may lie somewhere between these two extremes. Note that even if the transfer rate for an item is assumed to be a constant (unrelated to the number of items currently in the buffer) the first items presented for study still would have more information transferred to LTS than later items; this occurs because the items at the start of a list will not be knocked out of the buffer until it is filled and hence will reside in the buffer for a longer time on the average than later items. For this reason, the primacy effect could still be explained. On the other hand the primacy effect will be reduced by the constant transfer assumption; in order to fit the data from the current experiment with this assumption, for example, it would be necessary to adjust the retrieval scheme accordingly. In modeling the free verbal-recall data that follows, a constant transfer assumption is used and accordingly a retrieval scheme is adopted which amplifies more strongly than the present one small differences in LTS strength.

We now consider the decay assumption in greater detail. The assumption is that the information transferred to LTS for a particular item is reduced by a proportion τ for every other item in the list. There are a number of possibilities for the form of this reduction. It could be actual physical interference with the trace, or it could be a reduction in the value of the current information as a result of subsequent incoming information. An example of this latter kind of interference will be helpful.

Suppose, in a memory experiment the first item is GEX-5, and the subject stores "G__-5" in LTS. If tested now on GEX, the subject would give the correct response 5. Suppose a second item GOZ-3 is presented and the subject stores "G__-3" in LTS. If he is now tested on either GEX or GOZ his probability of a correct response will drop to $\frac{1}{2}$. Thus the actual information stored is not affected, but its value is markedly changed.

The assumption that every other item in a list interferes equally is open to question on two counts. First of all, it would be expected that an item about which a large amount of information is transferred would interfere more strongly with other items in LTS than an item about which little information is transferred. Certainly when no transfer occurs for an item, that item cannot interfere with other LTS traces. However, the equal interference assumption used in our analysis may not be a bad approximation. The second failing of the equal interference assumption has to do with separation of items. If the list lengths were very long, it might be expected that the number of items separating any two items would affect their mutual interference; the greater the separation, the less the interference. The list lengths are short enough in the present experiment, however, that the separation is probably not an important factor.

Some Alternative Models. It is worth considering some alternatives to the interference process of the model just presented, henceforth referred to as Model I in this subsection. In particular it is important to demonstrate that the effects of the interference-decay assumption, which could be viewed as a structural feature of memory, can be duplicated by simple search processes. For example, any limited search through the information in LTS will give poorer performance as the amount of that information increases. In order to make the concept of the search process clear, Model II will adopt an all-or-none transfer scheme. That is, a single copy of each item may be transferred to LTS on a probabilistic basis. If a copy is transferred, it is a perfect copy to the extent that it always produces a correct response if it is retrieved from LTS. The short-term features of the model are identical to those of Model I: each item enters the buffer; when the buffer is filled each succeeding item enters the buffer and knocks out an item already there according to the δ-process described earlier.

The transfer assumption for Model II is as follows. If an item is one of the j items in the buffer, then the probability that a copy of that item will be placed in LTS between one item's presentation and the next is $\frac{\theta}{j}$. Therefore, the transfer depends, as in Model I, upon the number of other items currently in the buffer. No more than one copy may be placed in LTS for any one item. The retrieval assumptions are the following. A correct response is given if the item is in the buffer when tested. If it is not in the buffer then a search is made in LTS. If a copy of the item exists in LTS and is found, then a correct response is given; otherwise a random guess is made. As before, we assume that the information pertinent to the current list is distinguishable from that of earlier lists; thus, the search is made only among those copies of items in the current list. The central assumption of Model II is that exactly R selections are made (with replacement) from the copies in LTS; if the tested item has not been found by then, the search ends.

The restriction to a fixed number of searches, R, is perhaps too strong, but can be justified if there is a fixed time period allotted to the subject for responding. It should be clear that for R fixed, the probability of retrieval decreases as the list length increases; the longer the list the more copies in LTS, and the more copies the less the probability of finding a particular copy in R selections. Model II was fit to the data in the same fashion as Model I. The parameter values that gave the best predictions were $r = 5$, $\delta = .39$, $\theta = .72$, and $R = 3.15$. The theoretical curves generated by these parameters are so similar to those for Model I that Figure 25 adequately represents them, and they will not be presented separately. Whereas the χ^2 was 44.3 for Model I, the χ^2 value for Model II was 46.2, both based on 42 degrees of freedom. The similarity of the predictions serves to illustrate the primary point of introducing Model II: effects predicted by search processes and by interference processes are quite similar and consequently they are difficult to separate experimentally.

The search process described above is just one of a variety of such mechanisms. In general there will be a group of possible search mechanisms associated with each transfer and storage assumption; a few of these processes will be examined in the next section on free verbal-recall. Before moving on to these experiments, however, we should like to present briefly a decay and retrieval process combining some of the features of interference and search mechanisms. In this process the interference does not occur until the search begins and is then caused by the search process itself. The model (designated as Model III) is identical in all respects to Model II until the point where the subject begins the search of LTS for the correct copy. The assumption is that the subject samples copies with replacement, as before, but each unsuccessful search may disrupt the sought-after copy with probabilitity R'. The search does not end until the appropriate copy is found or until all copies in LTS have been examined. If the copy does exist in LTS, but is disrupted at any time 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. The parameter values giving the best fit for this model were $r = 5$, $\delta = .38$, $\theta = .80$, and $R' = .25$. The predicted curves are again quite similar to those in Figure 25 and will not be presented. The predictions are not quite as accurate, however, as those of Models I and II, the χ^2 value being 55.0.*

5.2. Free-Verbal-Recall Experiments

The free-verbal-recall situation offers an excellent opportunity for examining retrieval processes, because the nature of the task forces the subject to engage in a lengthy search of LTS. 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 asked to recall as many of the words as possible. Quite often list length has been a variable, and occasionally the presentation time per item has been varied. Deese and Kaufman, for example, used

---

* For a more detailed account of Models I, II and III, and a comparison among models, see Atkinson and Shiffrin (1965).

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 is plotted against the item's position in the list. The Murdock (1962) results are representative and are shown in Figure 27. It should be made clear that the numbering of serial positions for these curves is opposite from the scheme used in the previous section; that is, the first item presented (the oldest item at the time of test) is labeled serial position 1. This numbering procedure will be used throughout this section to conform with the literature on free-verbal-recall; the reader should keep this in mind when comparing results here with those presented elsewhere in the paper. The primacy effect in Figure 27 is the rise on the lefthand portions of the curves and the recency effect is the larger rise on the right hand portions of the curves. The curves are labeled with the list length and the presentation rate per item. Note that the curves are quite similar to those found in Experiment 8 of the previous section; an effect not seen in Experiment 8 (because of the short list lengths used) is the level asymptotic portions of the curves which appear between the primacy and recency effects for the longer lists.

The form of the curves suggests that a buffer process could explain the results, with the words themselves being the units of rehearsal. The recency effect would be due to the probability that an item is still in the buffer at test; this probability goes to near zero after 15 items or so and the recency effect accordingly extends no further than this. The primacy effect would arise because more information accrued in LTS for the first few items presented in the list. Whether a buffer strategy

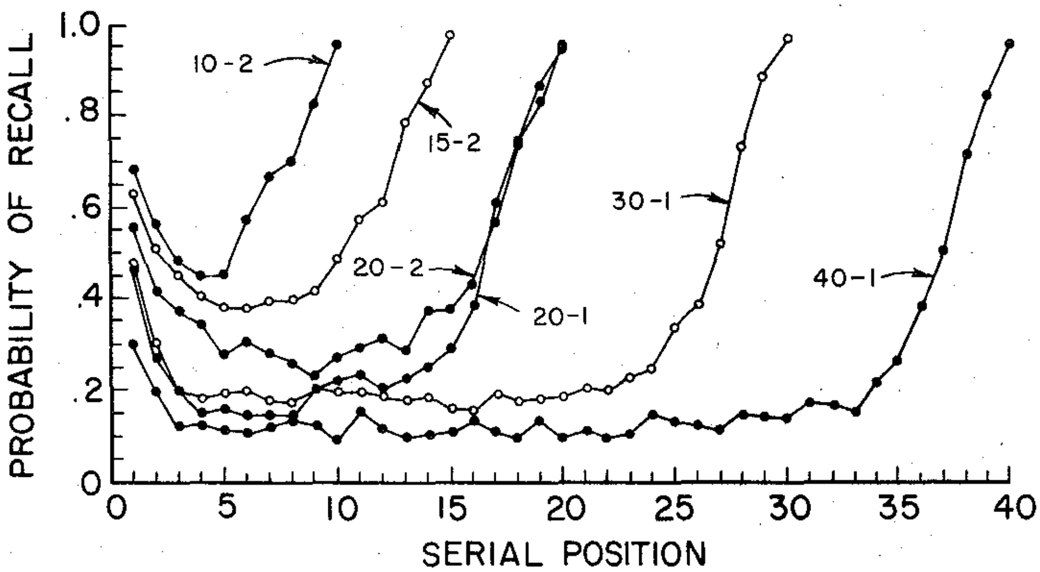


Figure 27 Probability of correct recall as a function of serial position for free verbal recall (after Murdock, 1962)

is reasonable in the free-recall situation, however, is worth further discussion. It can hardly be maintained that high-frequency English words are difficult to code; on the other hand the task is not a paired-associate one and cues must be found with which to connect the words. One possibility is that upon seeing each word the subject generates a number of associates (from LTS) and tries to store the group of words; later during testing a search which retrieves any of the associates might in turn retrieve the desired word. We tend to doubt that this strategy, used by itself, will greatly improve performance.* To the extent that coding occurs, it probably involves connecting words within the presented list to each other. This technique would of course require the consideration of a number of words simultaneously in STS and therefore might be characterized reasonably well by a buffer process. Whether or not coding occurs in the free-recall situation, there are other reasons for expecting the subjects to adopt a buffer strategy. The most important reason is undoubtedly the improvement in performance that a rehearsal buffer will engender. If the capacity of the buffer is, say, 4 or 5 words, then the use of a buffer will assure the subjects of a minimum of four or five items correct on each list (assuming that all of the items may be read out of the buffer correctly). Considering that subjects report on the average only about 8 or 9 items, even for long lists, the items stored in the buffer are an important component of performance.

---

* Cohen (1963) has presented free-recall lists containing closely related categories of words, i.e. North, East, South, West. Indeed, the recovery of one member of a category usually led to the recovery of other members, but the total number of categories recalled did not exceed the number of separate words recalled from non-categorized lists.

It will be assumed, then, that the subjects do adopt a rehearsal strategy. The comparability of the curves in Figure 25 to those in Figure 27 might indicate that a model similar to any of the models presented in the previous section could be applied to the current data. There are, however, important differences between the two experimental paradigms which must be considered: the free-recall situation does not involve pairing a response with a stimulus for each list position, and has the requirement of multiple recall at the time of test. The fact that explicit stimulus cues are not provided for each of the responses desired would be expected to affect the form of the search process. The multiple-response requirement raises more serious problems. In particular, it is possible that each response that is output may interfere with other items not yet recalled. The problem may be most acute for the case of items still in the buffer; Waugh and Norman (1965) have proposed that each response output at the time of test has the same disrupting effect upon other items in the buffer as the arrival of a new item during study. On the other hand, it is not clear whether a response emitted during test disrupts items in LTS. It might be expected that the act of recalling an item from LTS would raise that item's strength in LTS; this increase in strength is probably not associated, however, with the transfer of any new information to LTS. For this reason, other traces will most likely not be interferred with, and it shall be assumed that retrieval of an item from LTS has no effect upon other items in LTS.

Because there is some question concerning the effects of multiple recall upon the contents of the buffer, and because this section is primarily aimed at LTS processes, the part of the free-recall curves which arise from the buffer will not be considered in further analyses. This means that the models in this section will not be concerned with the part of the curve making up the recency effect; since the data in Figure 27 indicates that the recency effect is contained in the last 15 items (to the right in the figure) of each list, these points will be eliminated from the analyses. Unfortunately, the elimination of the last 15 items means that the short list lengths are eliminated entirely. The problem of obtaining data for short list lengths not contaminated by items in the buffer at the time of test has been circumvented experimentally by a variation of the counting-backwards technique. That is, the contents of the buffer can be eliminated experimentally by using an interfering task inserted between the end of the list and the start of recall. We now turn to a consideration of these experiments.

A representative experiment is that by Postman and Phillips (1965). Words were presented at a rate of one per second in all conditions. In one set of conditions three list lengths (10, 20, and 30) were used and recall was tested immediately following presentation. This, of course, is the usual free recall procedure. The serial position curves are shown in the top panel of Figure 28 in the box labeled "0 second." The same list lengths were used for those conditions employing an intervening task; immediately following presentation of the list the subjects were required to count backwards by threes and fours for 30 seconds. Following this intervening task, they were asked to recall the list. The results are shown in the lower panel in Figure 28. If the intervening task did not affect the contents of LTS but did wipe out all items in the buffer, then the recency effects would be expected to disappear with the curves

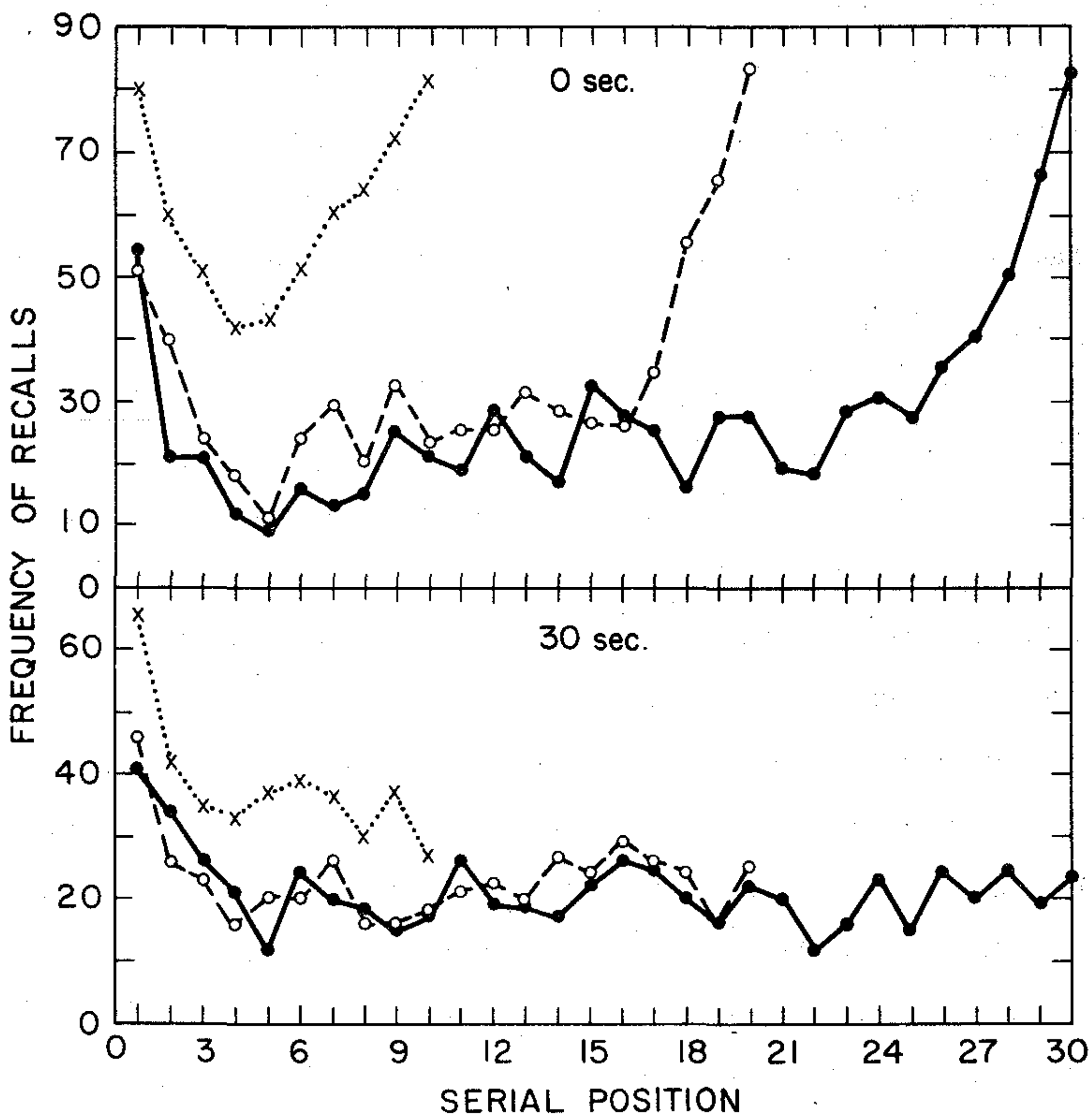


Figure 28 Probability of correct recall as a function of serial position for free verbal recall with test following 0 seconds and 30 seconds of intervening arithmetic (after Postman and Phillips, 1965)

otherwise unchanged. This is exactly what was found. The primacy effects and asymptotic levels remain unchanged while the recency effect disappears. It is clear, then, that normal free recall curves (without intervening arithmetic) from which the last 15 points have been deleted should be identical to curves from experiments using intervening arithmetic. The following data has therefore been accumulated: Murdock's data with the last 15 points of each list deleted; data reported by Deese and Kaufman (1957) using a free-recall paradigm, but again with the last 15 points of each list deleted; the data reported by Postman and Phillips (1965); and some data collected by Shiffrin in which an intervening task was used to eliminate the contents of the buffer.* All of these serial position curves have the same form; they show a primacy effect followed by a level asymptote. For this reason the results have been presented in Table 1. The first three points of each curve, which make up the primacy effect, are given in the table. The level portions of the curves are then averaged and the average shown in the column labeled asymptote. The column labeled "number of points" is the number of points which have been averaged to arrive at the asymptotic level.** The column labeled "list" gives the abbreviation of the experimenter, the list length, and the presentation rate for each of the serial position curves. (M = Murdock, 1962; D = Deese and Kaufman, 1957; P = Postman and Phillips, 1965; S = Shiffrin.)

* The Shiffrin data are reported in more detail in Atkinson and Shiffrin (1965)

**For the Postman-Phillips and Shiffrin lists the number of points at asymptote are simply list length, d, minus 3. For the Murdock and the Deese-Kaufman lists the number of points is d - 15 - 3 because the last 15 points in these lists have been eliminated.

Theoretical Analysis. Having accumulated a fair amount of parametric data in Table 1, we should now like to predict the results. The first model to be considered is extremely simple. Every item presented enters the subject's rehearsal buffer. One by one the initial items fill up the buffer, and thereafter each succeeding item knocks out of the buffer a randomly chosen item. In conditions where arithmetic is used following presentation, it is assumed that the arithmetic operations knock items from the buffer at the same rate as new incoming items. This is only an approximation, but probably not too inaccurate. Information is assumed to be transferred to LTS as long as an item remains in the buffer, in fact as a linear function of the total time spent in the buffer (regardless of the number of other items concurrently in the buffer). If an item remains in the buffer for j seconds an amount of information equal to θ times j is transferred to LTS. Call the amount of information transferred to LTS for an item its strength. When the subject engages in a search of LTS during recall it is assumed that he makes exactly R searches into LTS and then stops his search (the number of searches made might, for example, be determined by the time allowed for recall). On each search into LTS the probability that information concerning a particular item will be found is just the ratio of that item's strength to the sum of the strengths of all items in the list. Thus, items which have a greater LTS strength will be more likely to be found on any one search. The probability that the information in LTS will produce a correct recall, once that information has been found in a search, is assumed to be an exponential function of the strength for that item.

There are just three parameters for this model: r, the buffer size; θ, the parameter determining the rate per second at which information on a given item is transferred to LTS while the item resides in the rehearsal buffer; and R the number of searches made.* The probability of a correct response from the buffer is zero for the results in Table 1 because the contents of the buffer have been emptied experimentally by intervening arithmetic, or because the recency data (which represents recovery from the buffer) has been omitted. The parameters giving the best fit to the data were as follows: $r = 4$, $\theta = .04$, and $R = 34$. The predictions also are presented in Table 1. The predictions are rather remarkable considering that just three parameters have been used to predict the results from four different experiments employing different list lengths and different presentation rates. Some of the points are not predicted exactly but this is largely due to the fact that the data tends to be somewhat erratic; the predictions of the asymptotic values (where a larger amount of data is averaged) is especially accurate.

Some Alternative Models. A number of decisions were made in formulating the free-recall model that need to be examined in greater detail. First consider the effect of an arithmetic task upon items undergoing rehearsal. If the arithmetic caused all rehearsal and long-term storage

---

* It is important to remember that θ for this model is defined as the rate per second of information transfer, and thus the time measures listed in Table 3 need to be taken into account when applying the model. For example, an item that resides in the buffer for three item presentations will have 3θ amount of information in LTS if the presentation rate is one item per second, and 7.5θ if the presentation rate is 2.5 seconds per item.

operations to cease immediately, then the probability of recalling the last item presented should decrease toward chance (since its LTS strength will be negligible, having had no opportunity to accumulate). The serial position curve, however, remains level and does not drop toward the end of the list. One possible explanation is that all transfer to LTS takes place when the item first enters the buffer, rather than over the period the item remains in the buffer; in this case the onset of arithmetic would not affect the formation of traces in LTS. While this assumption could handle the phenomenon under discussion, we prefer to consider the LTS trace as building up during the period the item remains in the buffer. Recall that this latter assumption is borne out by the accuracy of the earlier models and, in particular, the U-shaped functions presented in Figure 12 for the multiple-reinforcement experiment. The explanation of the level serial position curve implied by our model is that the arithmetic operations remove items from the buffer in a manner similar to that of new entering items. Two sources give this assumption credibility. First, Postman and Phillips (1965) found that short periods of arithmetic (15 seconds) would leave some of the recency effect in the serial position curve, suggesting that some items remained in the buffer after brief periods of arithmetic. Second, the data of Waugh and Norman (1965) suggest that output operations during tasks such as arithmetic act upon the short-term store in the same manner as new incoming items.

Another choice point in formulating the model occurred with regard to the amount of LTS transfer for the first items in the list. The assumption used in an earlier model let the amount of transfer depend upon the number of other items concurrently undergoing rehearsal, as if the attention allotted to any given item determines the amount of transfer. An alternative possibility is that the amount of transfer is determined solely by the length of stay in the buffer and is therefore independent of the number of items currently in the buffer. Another assumption resulting in this same independence effect is that the subject allots to items in the buffer only enough attention to keep them "alive"; when the number of items in the buffer is small, the subject presumably uses his spare time for other matters. A free-verbal-recall experiment by Murdock (1965) seems to support a variant of this latter assumption. He had subjects perform a rather easy card-sorting task during the presentation of the list. The serial position curve seemed unaffected except for a slight drop in the primacy effect. This would be understandable if the card-sorting task was easy enough that the buffer was unaffected, but distracting enough that extra attention normally allotted to the first few items in the list (before the buffer is filled) is instead allotted to the card-sorting task. In any case, it is not clear whether the transfer rate should or should not be tied to the number of items concurrently in the buffer. The model that we have proposed for free-recall (henceforth referred to as Model I in this subsection) assumed a constant transfer process; a model using a variable transfer assumption will be considered in a moment.

The search process used in Model I is only one of many possibilities. Suppose, for example, that the strength value for an item represents the number of bits of information stored about that item (where the term "bits" is used in a non-technical sense). A search might then be construed as a

Table 1

Observed and Predicted Serial Position Curves

for Various Free-Verbal-Recall Experiments

List	Point 1		Point 2		Point 3		Asymptote		
	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.	Number of Points
M-20-1	.46	.45	.27	.37	.20	.29	.16	.22	2
M-30-1	.38	.35	.30	.28	.21	.22	.19	.17	12
M-20-2	.55	.61	.42	.51	.37	.41	.31	.32	2
M-40-1	.30	.29	.20	.23	.13	.18	.12	.14	22
M-25-1	.38	.39	.23	.32	.21	.25	.15	.19	7
M-20-2.5	.72	.66	.61	.56	.45	.46	.37	.35	2
D-32-1	.46	.33	.34	.27	.27	.21	.16	.16	14
P-10-1	.66	.62	.42	.52	.35	.42	.34	.32	7
P-20-1	.47	.45	.27	.37	.23	.29	.22	.22	17
P-30-1	.41	.35	.34	.28	.27	.22	.20	.17	27
S-6-1	.71	.74	.50	.64	.57	.52	.42	.40	3
S-6-2	.82	.88	.82	.79	.65	.66	.66	.52	3
S-11-1	.48	.60	.43	.50	.27	.40	.31	.31	8
S-11-2	.72	.76	.55	.66	.52	.54	.47	.42	8
S-17-1	.55	.49	.33	.40	.26	.32	.22	.24	14
S-17-2	.68	.66	.65	.56	.67	.45	.43	.35	14

random choice of one bit from all those bits stored for all the items in the list. The bits of information stored for each item, however, are associated to some degree, so that the choice of one bit results in the uncovering of a proportion of the rest of the information stored for that item. If this proportion is small, then different searches finding bits associated with a particular item will result in essentially independent probabilities of retrieval. This independent retrieval assumption was used in the construction of Model I. On the other hand, finding one bit in a search might result in all the bits stored for that item becoming available at once; a reasonable assumption would be that this information is either sufficient to allow retrieval or not, and a particular item is retrieved the first time it is picked in a search or is never retrieved. This will be called the dependent retrieval assumption.

It is interesting to see how well the alternate assumptions regarding transfer and search discussed in the preceding paragraphs are able to fit the data. For this reason, the following four models are compared:*

Model I: Transfer to LTS is at a constant rate θ regardless of the number of other items concurrently in the buffer, and independent retrieval.

Model II: Transfer to LTS is at a variable rate $\frac{\theta}{j}$ where j is the number of other items currently in the buffer, and independent retrieval.

* These models and the related mathematics are developed in Atkinson and Shiffrin (1965).

Model III: Constant LTS transfer rate, and dependent retrieval.

Model IV: Variable LTS transfer rate, and dependent retrieval.

Model I, of course, is the model already presented for free-verbal-recall.

The four models were all fit to the free-verbal-recall data presented in Table 1, and the best fits, in terms of the sums of the squared deviations, were as follows: Model I: .814; Model II: 2.000; Model III: .925; Model IV: 1.602 (the lowest sum meaning the best predictions). These results are of interest because they demonstrate once again the close interdependence of the search and transfer processes. Neither model employing a variable transfer assumption is a good predictor of the data and it seems clear that a model employing this assumption would require a retrieval process quite different from those already considered in order to fit the data reasonably well.

Perhaps the most interesting facet of Model I is its ability to predict performance as the presentation rate varies. A very simple assumption, that transfer to LTS is a linear function of time spent in the buffer, seems to work quite well. Waugh (1967) has reported a series of studies which casts some light on this assumption; in these studies items were repeated a variable number of times within a single free-recall list. The probability of recall was approximately a linear function of the number of repetitions; this effect is roughly consonant with an assumption of LTS transfer which is linear with time. It should be noted that the presentation rates in the experiments we analyzed do not vary too widely: from 1 to 2.5 seconds per item. The assumption that the subject will adopt a buffer strategy undoubtedly breaks down if a wide enough range in presentation rates is considered. In particular, it can be expected that the subject will make increasing use of coding strategies as the presentation rate decreases. M. Clark and G. Bower (personal communication) for example, have shown that subjects proceeding at their own pace (about 6-12 seconds a word) can learn a list of ten words almost perfectly. This memorization is accomplished by having the subject make up and visualize a story including the words that are presented. It would be expected that very slow presentation rates in free-recall experiments would lead to coding strategies similar to the one above.

One last feature of the models in this section needs further examination. Contrary to our assumption, it is not true that successive lists can be kept completely isolated from each other at the time of test. The demonstration of this fact is the common finding of intrusion errors: items reported during recall which had been presented on a list previous to the one being tested. Occasionally an intrusion error is even reported which had not been reported correctly during the test of its own list. Over a session using many lists, it might be expected that the interference from previous lists would stay at a more or less constant level after the presentation of the first few lists of the session. Nevertheless, the primacy and asymptotic levels of the free-recall serial position curves should drop somewhat over the first few lists. An effect of this sort is reported by Wing and Thompson (1965) who examined serial position curves for the first, second, and third presented lists of a session. This effect is undoubtedly similar to the one reported by Keppel and Underwood (1962); namely, that performance on the task used by Peterson (1959) drops over the first few trials of a session. The effects in both of these experiments may be caused by the increasing difficulty of the search process during test.

5.3. Further Considerations Involving LTS

The models presented in the last section, while concerned with search and retrieval processes, were nevertheless based primarily upon the concept of a rehearsal buffer. This should not be taken as an indication that rehearsal processes are universally encountered in all memory experiments; to the contrary, a number of conditions must exist before they will be brought into play. It would be desirable at this point then to examine some of the factors that cause a subject to use a rehearsal buffer. In addition, we want to consider a number of points of theoretical interest that arise naturally from the framework developed here. These points include possible extensions of the search mechanisms, relationships between search and interference processes, the usefulness of mnemonics, the relationships between recognition and recall, and coding processes that the subject can use as alternatives to rehearsal schemes.

Consider first the possible forms of search mechanisms and the factors affecting them. Before beginning the discussion two components of the search process should be emphasized: the first component involves locating information about an item in LTS, called the "hit" probability; the second component is the retrieval of a correct response once information has been located. The factor determining the form of the search is the nature of the trace in long-term store. The models considered thus far have postulated two different types of traces. One is an all-or-none trace which allows perfect recall following a hit; the other is an unspecified trace which varies in strength. The strength notion has been used most often because it is amenable to a number of possible interpretations: the strength could represent the "force" with which a particular bond has been formed, the number of bits of information which have been stored, or the number of copies of an item placed in memory. It should be emphasized that these different possibilities imply search processes with different properties. For example, if the strength represents the force of a connection then it might be assumed that there is an equal chance of hitting any particular item in a search, but the probability of giving a correct answer following a hit would depend upon the strength. On the other hand, the strength might represent the number of all-or-none copies stored in LTS for an item, each copy resulting in a correct response if hit. In this case, the probability of a hit would depend upon the strength (the number of copies) but any hit would automatically result in a correct answer. A possibility intermediate to these two extremes is that partial copies of information are stored for each item, any one partial copy allowing a correct response with an intermediate probability. In this case, the probability of a hit will depend on the number of partial copies, and the probability of a correct response following a hit will depend on the particular copy that has been found. A different version of this model would assume that all the partial copies for an item become available whenever any one copy is hit; in this version the probability of a correct answer after a hit would depend on the full array of copies stored for that item. In all the search processes where the retrieval probability following a hit is at an intermediate level, one must decide whether successive hits of that item will result in independent retrieval probabilities. It could be assumed, for example, that failure to uncover a correct response the first time an item is hit in the search would mean that the correct response could not be recovered on subsequent hits of that item.* This outline of some selected search processes indicates the variety of possibilities; a variety which makes it extremely difficult to isolate effects due to search processes from those attributable to interference mechanisms.

Other factors affecting the form of the search are at least partially controlled by the subject; a possible example concerns whether or not the searches are made with replacement. Questions of this sort are based upon the fact that all searches are made in a more or less ordered fashion; memory is much too large for a completely random search to be feasible. One ordering which is commonly used involves associations: each item recovered leads to an associate which in turn leads to another associate. The subject presumably exercises control over which associates are chosen at each stage of the search and also injects a new starting item whenever a particular sequence is not proving successful.** An alternative to the associate method is a search along some partially ordered dimension. Examples are easy to find; the subject

* For a discussion of partial and multiple copy models see Atkinson and Shiffrin (1965).

** Associative search schemes have been examined rather extensively using free-recall methods. Clustering has been examined by Deese (1959), Bousfield (1953), Cofer (1966), Tulving (1962), and others; the usual technique is to determine whether or not closely associated words tend to be reported together. The effect certainly exists, but a lack of parametric data makes it difficult to specify the actual search process involved.

could generate letters of the alphabet, considering each in turn as a possible first letter of the desired response. A more general ordered search is one that is made along a temporal dimension; items may be time-tagged or otherwise temporally ordered, and the subject searches only among those items that fall within a particular time span. This hypothesis would explain the fact that performance does not markedly deteriorate even at the end of memory experiments employing many different lists, such as in the free-verbal-recall paradigm. In these cases, the subject is required to respond only with members of the most recent list; if performance is not to degenerate as successive lists are presented, the memory search must be restricted along the temporal dimension to those items recently stored in LTS. Yntema and Trask (1963) have demonstrated that temporal information is available over relatively long time periods (in the form of "time-tags" in their formulation) but the storage of such information is not well understood.

We now turn to a brief discussion of some issues related to interference effects. It is difficult to determine whether time alone can result in long-term interference. Nevertheless, to the extent that subjects engage in a search based upon the temporal order of items, interference due to the passage of time should be expected. Interference due to intervening material may take several forms. First, there may be a reduction in the value of certain information already in LTS as a result of the entry of new information; the loss in this case does not depend on making any previous information less accessible. An example would be if a subject first stores "the stimulus beginning with D has response 3" and later when another stimulus beginning with D is presented, he stores "the stimulus beginning with D has response 1." The probability of a correct response will clearly drop following storage of the second trace even though access to both traces may occur at test. Alternatively, interference effects may involve destruction of particular information through interaction with succeeding input. This possibility is often examined experimentally using a paired-associate paradigm where the same stimulus is assigned different responses at different times. DaPolito (1966) has analyzed performance in such a situation. A stimulus was presented with two different responses at different times, and at test the subject was asked to recall both responses. The results indicated that the probability of recalling the first response, multiplied by the probability of recalling the second response, equals the joint probability that both responses will be given correctly. This result would be expected if there was no interaction of the two traces; it indicates that high strengths of one trace will not automatically result in low strengths on the other. The lack of an interaction in DaPolito's experiment may be due to the fact that subjects knew they would be tested on both responses. It is interesting to note that there are search mechanisms that can explain this independence effect and at the same time interference effects. For example, storage for the two items might be completely independent as suggested by DaPolito's data; however, in the typical recall task the subject may occasionally terminate his search for information about the second response prematurely as a result of finding information on the first response.

Within the context of interference and search processes, it is interesting to speculate about the efficacy of mnemonics and special coding techniques. It was reported, for example, that forming a visual image of the two words in a paired-associate item is a highly effective memory device; that is, one envisages a situation involving the two words. Such a mnemonic gains an immediate advantage through the use of two long-term systems, visual and auditory, rather than one. However, this cannot be the whole explanation. Another possibility is that the image performs the function of a mediator, thereby reducing the set of items to be searched; that is, the stimulus word when presented for test leads naturally to the image which in turn leads to the response. This explanation is probably not relevant in the case of the visual-image mnemonic for the following reason: the technique usually works best if the image is a very strange one. For example, "dog-concrete" could be imaged as a dog buried to the neck in concrete; when "dog" is tested, there is no previously well-learned association that would lead to this image. Another explanation involves the protection of the stored information over time; as opposed to the original word pairs, each image may be stored in LTS as a highly distinct entity. A last possibility is that the amount of information stored is greatly increased through the use of imagery -- many more details exist in the image than in the word pair. Since the image is highly cohesive, the recovery of any information relevant to it would lead to the recovery of the whole image. These hypotheses are of course only speculations. At the present time the relation of the various search schemes and interference processes to mnemonic devices is not well understood. This state of affairs hopefully will change in the near future since more research is being directed toward these areas; mediation, in particular, has been receiving extensive consideration (e.g., Bugelski, 1962; Runquist and Farley, 1964).

Search processes seem at first glance to offer an easy means for the analysis of differences between recognition and recall. One could assume, for example, that in recall the search component which attempts to locate information on a given item in LTS is not part of the recognition process; that is, one might assume that in recognition the relevant information in LTS is always found and retrieval depends solely on matching the stored information against the item presented for test. Our analysis of free-verbal recall depended in part upon the search component to explain the drop in performance as list length increased. Thus if the free recall task were modified so that recognition tests were used, the decrement in performance with list length might not occur. That this will not be the case is indicated by the position-to-color memory study (Experiment 8) in which the number of responses was small enough that the task was essentially one of recognition; despite this fact, the performance dropped as list length increased. One possible explanation would be that search is necessary even for recognition tasks; i.e., if the word "clown" is presented, all previous times that that word had been stored in LTS do not immediately spring to mind. To put this another way, one may be asked if a clown was a character in a particular book and it is necessary to search for the appropriate information, even though the question is one of recognition. On the other hand, we cannot rule out the possibility that part of the decrement in performance in free recall with the increase of list length may be due to search changes, and part to other interference mechanisms. Obviously a great deal of extra information is given to the subject in a recognition test, but the effect of this information upon search and interference mechanisms is not yet clear.

We now turn to a consideration of LTS as it is affected by short-term processes other than the rehearsal buffer. It has been pointed out that the extent and structure of rehearsal depends upon a large number of factors such as the immediacy of test and difficulty of long-term storage. When rehearsal schemes are not used in certain tasks, often it is because long-term coding operations are more efficacious. These coding processes are presumably found in most paired-associate learning paradigms; depending upon conditions, however, the subject will probably divide his attention between coding and rehearsal. Atkinson and Shiffrin (1965) have presented a paired-associate learning model based upon a rehearsal-buffer. Whether a rehearsal strategy would be adopted by the subject in a given paired-associate learning experiment needs to be determined in each case. The answer is probably no for the typical fixed-list learning experiment, because the items are usually amenable to coding, because the test procedure emphasizes the importance of LTS storage, and because short study-test intervals are so infrequent that maintainance of an item in STS is not a particularly effective device. If these conditions are changed, however, then a paired-associate model based upon a rehearsal buffer might prove applicable.

It is important to note the distinction between coding models and rehearsal models. Rehearsal models actually encompass, in a rough sense, virtually all short-term processes. Coding, for example, may be considered as a type of rehearsal involving a single item. The buffer process is a special type of rehearsal in which a fixed number of items are rehearsed for the primary purpose of maintaining them in STS. A pure coding process is one in which only a single item is considered at a time and in which the primary purpose is the generation of a strong LTS trace; almost incidentally, the item being coded will be maintained in STS through the duration of the coding period, but this is not a primary purpose of the process. These various processes, it should be emphasized, are under subject control and are brought into play as he sees fit; consequently there are many variations that the subject can employ under appropriate conditions. One could have a coding model, for example, in which more than one item is being coded at a time, or a combination model in which several items are maintained via rehearsal while one of the items is selected for special coding.

At the other extreme from the buffer strategy, it might be instructive to consider a coding process that acts upon one item at a time. Although such a process can be viewed as a buffer model with a buffer containing only one item, the emphasis will be upon LTS storage rather than upon the maintenance of the item in STS. The simplest case occurs when the presentation rate is fairly slow and the subject attempts to code each item as it is presented for study. However, the case that seems most likely for the typical paired-associate experiment, is that in which not every item is coded, or in which it takes several presentation periods to code a single item. The first case above could be conceptualized as follows: each item is given a coding attempt during its presentation interval, but the probability of finding a code is ξ. The second case is a bit more complex. One version would have a single item maintained in STS over trials until a code is found. It could be supposed that the probability of a code being found during a single presentation interval is ξ; having once coded an item, coding attempts are focused on the next presented item. This model has something in common with the buffer models in that some items will remain in STS over a period of several trials. This will produce a short-term decay effect as the interval between presentation and test is increased.

It is worth considering the form of the usual short-term effects that are found in a paired-associate learning. Figure 29 presents data from a paired-associate experiment by Bjork (1966). Graphed is the probability of a correct response for an item prior to its last error, as a function of the number of other items intervening between its study and subsequent test. The number of intervening items that must occur before this curve reaches the chance level can be taken as a measure of the extent of the short-term effect. It can be seen that the curve does not reach chance level until after about 20 items have been presented. If the coding model mentioned above were applied to this data, a short-term effect would be predicted due to the fact that some items are kept in STS for more than one trial for coding. It hardly seems likely, however, that any item will be kept in STS for 20 trials in an attempt to code it. Considerations of this sort have led a number of workers to consider other sources for the "short-term" effect. One possibility would be that the effect is based in LTS and is due to retroactive interference. A model in which this notion has been formalized was set forth by Restle (1964) and subsequently developed by Greeno (1967). For our purposes Greeno's presentation is more appropriate. He proposes that a particular code may be categorized as "good" or "bad." A good code is permanent and will not be interfered with by the other materials presented in the experiment. A bad code will be retrievable from LTS for a time, but will be subject to interference from succeeding items

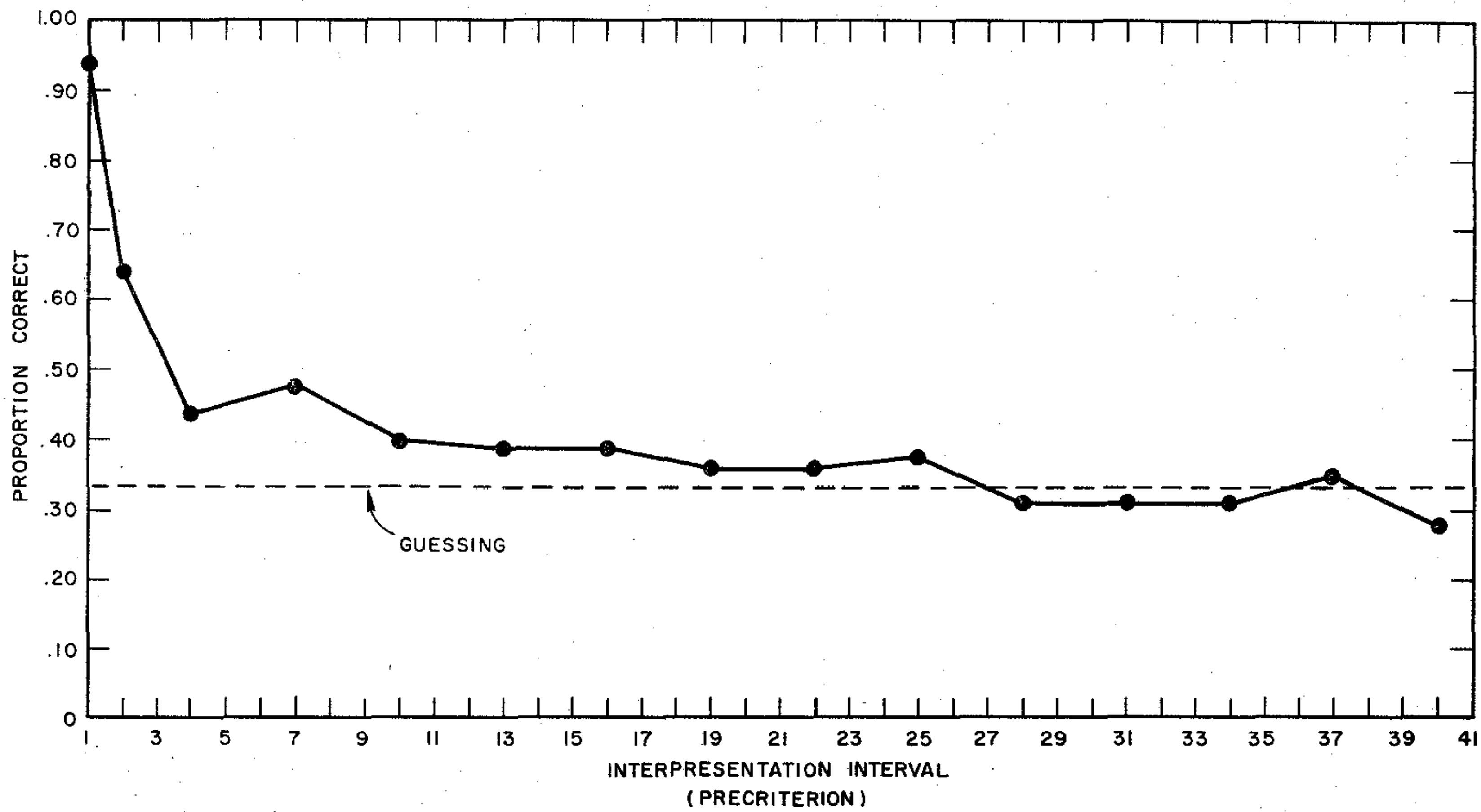


Figure 29 Probability of a correct response prior to the last error as a function of lag (after Bjork, 1966)

and will eventually be useless. Employing this model, the short-term effects displayed in Figure 29 are due to those items that were assigned bad codes (i.e., codes that were effective for only a short period of time). The interesting feature of this model is its inclusion of a short-term memory effect based not upon features of STS, but upon processes in LTS.* One other useful way in which this LTS interference process has been viewed employs Estes' stimulus fluctuation theory (Estes, 1965, a, b). In this view, elements of information in LTS sometime become unavailable; it differs from the above models in that an unavailable element may become available again at a later time. In this sense, fluctuation theory parallels a number of the processes that are expected from search considerations. In any case, the theory has been successfully applied in a variety of situations (Izawa, 1966). There is a great deal more that can be said about paired-associate learning and long-term processes in general, but it beyond the scope of this paper to enter into these matters. We should like to re-emphasize, however, the point that has just been made; namely, that short-term decay effects may arise from processes based in LTS as well as mechanisms in STS; considerable care must be taken in the analysis of each experimental situation in order to make a correct identification of the processes at play.

* It is this short-term effect that is probably captured by the intermediate state in various Markov models for paired-associate learning (Atkinson and Crothers, 1964; Bernbach, 1965; Bjork, 1966; Calfee and Atkinson, 1965; Kintsch, 1965; Young, 1966). Theorists using these models have been somewhat noncommital regarding the psychological rationale for this intermediate state, but the estimated transition probabilities to and from the state suggest to us that it represents effects taking place in LTS.

SECTION 6. CONCLUDING REMARKS

The first three sections of this paper outlined a fairly comprehensive theoretical framework for memory which emphasized the role of control processes -- processes under the voluntary control of the subject such as rehearsal, coding, and search strategies. It was argued that these control processes are such a pervasive and integral component of human memory that a theory which hopes to achieve any degree of generality must take them into account. Our theoretical system has proven productive of experimental ideas. In Sections 4 and 5 a particular realization of the general system involving a rehearsal buffer was applied to data from a variety of experiments. The theoretical predictions were, for the most part, quite accurate, proving satisfactory even when based upon previously estimated parameter values. It was possible to predict data over a range of experimental tasks and a wide variety of independent variables such as stimulus-set size, number of reinforcements, rehearsal procedures, list length, and presentation rate. Perhaps even more impressive are the number of predictions generated by the theory which ran counter to our initial intuitions but were subsequently verified.

It should be emphasized that the specific experimental models we have considered do not represent a general theory of the memory system but rather a subclass of possible models that can be generated by the framework proposed in the first half of the paper. Paired-associate learning, for example, might best be described by models emphasizing control processes other than rehearsal. These models could be formulated in directions suggested by stimulus sampling theory (Estes, 1955a; 1955b; 1967), models stressing cue selection and coding (Restle, 1964; Greeno, 1966), or queuing models (Bower, in press).

Finally, it should be noted that most of the ideas in this paper date back many years to an array of investigators: Broadbent (1957, 1958) and Estes (1967) in particular have influenced the development of our models. The major contribution of this paper probably lies in the organization of results and the analysis of data; in fact, theoretical research could not have been carried out in the manner reported here as little as 12 years ago. Although conceptually the theory is not very difficult to understand, many of our analyses would have proved too complex to investigate without the use of modern, high-speed computers.

References

Atkinson, R. C., Brelsford, J. W., and Shiffrin, R. M. Multi-process models for memory with applications to a continuous presentation task. Journal of Mathematical Psychology, 1967, in press.

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

Atkinson, R. C., and Shiffrin, R. M. Mathematical models for memory and learning. Technical Report 79, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965. (To be published in D. P. Kimble (Ed.), The Proceedings of the Third Conference on Learning, Remembering and Forgetting. New York: New York Academy of Sciences.)

Averbach, E. and Coriell, A. S. Short-term memory in vision. Bell System Technical Journal, 1961, 40, 309-328.

Averbach, E. and Sperling, G. Short term storage of information in vision. In C. Cherry (Ed.), Information Theory, London: Butterworth, 1961, 196-211.

Barbizet, J. Defect of memorizing of hippocampal-mammillary origin: a review. Journal of Neurology, Neurosurgery and Psychiatry, 1963, 26, 127-135.

Battig, W. F. Evidence for coding processes in "rote" paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 171-181.

Bernbach, H. A. A forgetting model for paired-associate learning. Journal of Mathematical Psychology, 1965, 2, 128-144.

Bernbach, H. Decision processes in memory. *Psychological Review*, 1967, in press. (a)

Bernbach, H. A replica model for short-term memory. Technical Report, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1967. (b)

Binford, J. R. and Gettys, C. Nonstationarity in paired-associate learning as indicated by a second guess procedure. *Journal of Mathematical Psychology*, 1965, *2*, 190-195.

Bjork, R. A. Learning and short-term retention of paired-associates in relation to specific sequences of interpresentation intervals. Technical Report No. 106, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1966.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, *49*, 229-240.

Bower, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, *26*, 255-280.

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence and J. T. Spence (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory*. Vol. 1, New York: Academic Press, 1967, in press.

Bower, G. H. Notes on a descriptive theory of memory. In D. P. Kimble (Ed.), *The Proceedings of the Second Conference on Learning, Remembering and Forgetting*. New York: New York Academy of Sciences, in press.

Brelsford, J. W., Jr. and Atkinson, R. C. Multiple reinforcement effects in short-term memory. Technical Report, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1967.

Brelsford, J. W., Jr. and Atkinson, R. C. Short-term memory as a function of rehearsal procedures. Journal of Verbal Learning and Verbal Behavior, in press.

Brelsford, J. W., Jr., Keller, L., Shiffrin, R. M., and Atkinson, R. C. Short-term recall of paired-associates as a function of the number of interpolated pairs. Psychonomic Science, 1966, 4, 73-74.

Broadbent, D. E. The role of auditory localization in attention and memory span. Journal of Experimental Psychology, 1954, 47, 191-196.

Broadbent, D. E. Successive responses to simultaneous stimuli, Quarterly Journal of Experimental Psychology, 1956, 8, 145-152.

Broadbent, D. E. A mechanical model for human attention and immediate memory. Psychological Review, 1957, 64, 205-215.

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

Broadbent, D. E. Flow of information within the organism. Journal of Verbal Learning and Verbal Behavior, 1963, 4 34-39.

Brown, J. Some tests of decay theory of immediate memory. Quarterly Journal of Experimental Psychology, 1958, 10, 12-21.

Brown, R. and McNeill, D. The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 325-337.

Bugelski, B. R. Presentation time, total time, and mediation in paired-associate learning. Journal of Experimental Psychology, 1962, 63, 409-412.

Buschke, H. Retention in immediate memory estimated without retrieval. Science, 1963, 140, 56-57.

Buschke, H. and Lim, H. Temporal and interactional effects in short-term storage. Perception and Psychophysics, 1967, in press.

Calfee, R. C. and Atkinson, R. C. Paired-associate models and the effects of list length. _Journal of Mathematical Psychology_, 1965, _2_, 254-265.

Cofer, C. N. Some evidence for coding processes derived from clustering in free recall. _Journal of Verbal Learning and Verbal Behavior_, 1966, _5_, 188-192.

Cohen, B. H. Recall of categorized word lists. _Journal of Experimental Psychology_, 1963, _66_, 227-234.

Cohen, R. L. and Johansson, B. S. The activity trace in immediate memory; a reevaluation. _Journal of Verbal Learning and Verbal Behavior_, in press.

Conrad, R. Acoustic confusions in immediate memory. _British Journal of Psychology_, 1964, _55_, 1, 75-84.

Conrad, R. and Hille, B. A. The decay theory of immediate memory and paced recall. _Canadian Journal of Psychology_, 1958, _12_, 1-6.

DaPolito, F. J. Proactive effects with independent retrieval of competing responses. Unpublished doctoral dissertation, Indiana University, 1966.

Deese, J. Influence of inter-item associative strength upon immediate free recall. _Psychological Reports_, 1959, _5_, 305-312.

Deese, J. and Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. _Journal of Experimental Psychology_, 1957, _54_, 180-187.

Estes, W. K. Statistical theory of distributional phenomena in learning. _Psychological Review_, 1955, _62_, 369-377. (a)

Estes, W. K. Statistical theory of spontaneous recovery and regression. _Psychological Review_, 1955, _62_, 145-154. (b)

Estes, W. K. A technique for assessing variability of perceptual span. *Proceedings of the National Academy of Sciences*, 1965, *4*, No. 2, 403-407.

Estes, W. K. Learning and memory. In E. F. Beckenbach (Ed.), *Conceptual Bases and Application of the Communications Sciences*, 1967, in press.

Estes, W. K. and Taylor, H. A. A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences*, 1964, *52*, 2, 446-454.

Estes, W. K. and Taylor, H. A. Visual detection in relation to display size and redundancy of critical elements. *Perception and Psychophysics*, 1966, *1*, 9-16.

Estes, W. K. and Wessel, D. L. Reaction time in relation to display size and correctness of response in forced-choice visual signal detection. *Perception and Psychophysics*, 1966, *1*, 369-373.

Freedman, J. L. and Landauer, T. K. Retrieval of long-term memory: "tip-of-the-tongue" phenomenon. *Psychonomic Science*, 1966, *4*, 309-310.

Greeno, J. G. Paired-associate learning with short-term retention: mathematical analysis and data regarding identification of parameters. *Journal of Mathematical Psychology*, 1967, in press.

Harley, W. F., Jr. The effect of monetary incentive in paired-associate learning using a differential method. *Psychonomic Science*, 1965, *2*, 377-378. (a)

Harley, W. F., Jr. The effect of monetary incentive in paired-associate learning using an absolute method. *Psychonomic Science*, 1965, *3*, 141-142. (b)

Hart, J. T., Jr. Recall, recognition, and the memory-monitoring process. Unpublished doctoral dissertation, Stanford University, 1965.

Hebb, D. O. Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), Brain Mechanisms and Learning. London and New York: Oxford University Press, 1961, 37-46.

Hellyer, S. Supplementary report: Frequency of stimulus presentation and short-term decrement in recall. Journal of Experimental Psychology, 1962, 64, 650.

Hintzman, D. L. Classification and aural coding in short-term memory. Psychonomic Science, 1965, 3, 161-162.

Hintzman, D. L. Articulatory coding in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1967, in press.

Izawa, Chizuko. Reinforcement-test sequences in paired-associate learning. Psychological Reports, Monograph Supplement, 3-V18, 1966, 879-919.

Katz, L. A technique for the study of steady-state short-term memory. Psychonomic Science, 1966, 4, 361-362.

Keppel, G. and Underwood, B. J. Proactive inhibition in short-term retention of single items. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 153-161.

Kintsch, W. Habituation of the GSR component of the orienting reflex during paired-associate learning before and after learning has taken place. Journal of Mathematical Psychology, 1965, 2, 330-341.

Lambert, W. E. and Jakobovitz, L. A. Verbal satiation and changes in the intensity of meaning. Journal of Experimental Psychology, 1960, 60, 376-383.

Landauer, T. K. Rate of implicit speech. Perceptual and Motor Skills, 1962, 15, 646-647.

Madsen, M. E. and Drucker, J. M. Immediate memory by missing scan and modified digit span. Psychonomic Science, 1966, 6, 283-284.

Melton, A. W. Implications of short-term memory for a general theory of memory. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 1-21.

Meumann, E. The Psychology of Learning. New York: D. Appleton and Company, 1913.

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

Milner, Brenda. The memory defect in bilateral hippocampal lesions. Psychiatric Research Reports, 1959, 11, 43-58.

Milner, Brenda. Neuropsychological evidence for differing memory processes. Abstract for the symposium on short-term and long-term memory; XVIII International Congress of Psychology, Moscow, August, 1966.

Milner, Brenda. Amnesia following operation on the temporal lobes. In O. L. Zangwill and C. W. M. Whitty (Eds.), Amnesia. London: Butterworths, 1967, in press.

Montague, W. E., Adams, J. A., and Kiess, H. O. Forgetting and natural language mediation. Journal of Experimental Psychology, 1966, 72 829-833.

Murdock, B. B., Jr. The serial position effect of free recall. Journal of Experimental Psychology, 1962, 64, 482-488.

Murdock, B. B., Jr. Effects of a subsidiary task on short term memory. *British Journal of Psychology*, 1965, *56*, 413-419.

Murdock, B. B., Jr. The criterion problem in short-term memory. *Journal of Experimental Psychology*, 1966, *72*, 317-324.

Peterson, L. R. Short-term verbal memory and learning. *Psychological Review*, 1966, *73*, 193-207. (a)

Peterson, L. R. Short-term memory. *Scientific American*, 1966, *215* 90-95. (b)

Peterson, L. R. and Peterson, Margaret. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, *58*, 193-198.

Peterson, L. R., Wampler, R., Kirkpatrick, M., and Saltzman, D. Effect of spacing presentations on retention of a paired-associate over short intervals. *Journal of Experimental Psychology*, 1963, *66*, 206-209.

Phillips, J. L., Shiffrin, R. M., and Atkinson, R. C. The effects of list length on short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1967, in press.

Posner, M. I. Components of skilled performance. *Science*, 1966, *152*, 1712-1718.

Postman, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal Learning and Verbal Behavior*. New York: McGraw-Hill, 1961, 152-179.

Postman, L. Does interference theory predict too much forgetting? *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 40-48.

Postman, L. Short-term memory and incidental learning. In A. W. Milton (Ed.), *Categories of Human Learning*. New York: Academic Press, 1964.

Postman, L. and Phillips, Laura W. Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 1965, *17*, 132-138.

Restle, F. Sources of difficulty in learning paired associates. In R. C. Atkinson (Ed.), *Studies in Mathematical Psychology*. Stanford: Stanford University Press, 1964, 116-172.

Runquist, W. N. and Farley, F. H. The use of mediators in the learning of verbal paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1964, *3*, 280-285.

Sperling, G. The information available in brief visual presentations. *Psychology Monograph*, 1960, *74*, Whole No. 498.

Sperling, G. A model for visual memory tasks. *Human Factors*, 1963, *5*, 19-31.

Sperling, G. Successive approximations to a model for short-term memory. *Proceedings of the XVIII International Congress of Psychology*, Moscow. Amsterdam: North Holland Publishing Company, 1967, in press.

Sternberg, Saul. High speed scanning in human memory. *Science*, 1966, *153*, 652-654.

Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychology Review*, 1962, *69*, 344-354.

Underwood, B. J. and Ekstrund, B. R. Studies of distributed practice: XXIV. Differentiation and proactive inhibition. *Journal of Experimental Psychology*, 1967, in press.

Waugh, Nancy C. Presentation time and free recall. *Journal of Experimental Psychology*, 1967, *73*, 39-44.

Waugh, Nancy C. and Norman, D. A. Primary memory. *Psychological Review*, 1965, *72*, 89-104.

Wickelgren, W. A. Size of rehearsal group and short-term memory. *Journal of Experimental Psychology*, 1965, *68*, 413-419.

Wickelgren, W. A. and Norman, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 1966, *3*, 316-347.

Wing, J. F. and Thomson, Barbara P. Primacy-recency effects in free recall. *The Proceedings of the 1965 Annual Convention of the American Psychological Association*, 1965, 57-58.

Young, J. L. Effects of intervals between reinforcements and test trials in paired-associate learning. Technical Report 101, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1966.

Yntema, D. B., and Mueser, G. A. Remembering the present state of a number of variables. *Journal of Experimental Psychology*, 1960, *60*, 18-22.

Yntema, D. B. and Mueser, G. D. Keeping track of variables that have few or many states. *Journal of Experimental Psychology*, 1962, *63*, 391-395.

Yntema, D. B. and Trask, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 65-74.

SOME SPECULATIONS ON STORAGE AND RETRIEVAL PROCESSES IN LONG-TERM MEMORY

by

R. C. Atkinson and R. M. Shiffrin

TECHNICAL REPORT NO. 127

February 2, 1968

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Some Speculations on Storage and Retrieval Processes in Long-Term Memory

R. C. Atkinson and R. M. Shiffrin

Stanford University
Stanford, California 94305

ABSTRACT

A brief outline of the memory system is followed by somewhat speculative proposals for storage and retrieval processes, with particular care being given to distinguishing structural components from control processes set up and directed by the subject. The memory trace is conceived of as an ensemble of information, possibly stored in many places. For a given set of incoming information, the questions dealt with are whether to store, how to store, and where to store; the last question in particular deals with storage along various dimensions. Retrieval consists of a search along storage dimensions utilizing available cues to limit the search area and provide appropriate entry points. Both storage and retrieval are considered to take place in two steps, one consisting of a highly directed process under control of the subject and the other consisting of a pseudo-random component.

Paper presented at a conference on "Research on Human Decision Making" sponsored by the NASA-Ames Research Center at Moffett Field, California, 31 January - 2 February, 1968.

This paper will take a fairly speculative look at the structure of long-term memory, at the storage and retrieval processes by which information is placed in and recovered from long-term memory, at the joint operation of the short- and long-term stores, and at the control processes governing these various mechanisms. While the discussion will be primarily theoretical with no attempt made to document our assumptions by recourse to the experimental literature, some selected experiments will be brought in as examples. We will begin by outlining the overall conception of the memory system, a conception which emphasizes the importance of control processes. Long-term storage and retrieval will then be discussed in terms of the basic assumption that stored information is not destroyed or erased over time. This assumption may of course be relaxed, but we employ it to demonstrate that forgetting phenomena can be satisfactorily explained by postulating that decrements in performance occur as a result of a decreasingly effective search of long-term memory.

The primary distinction in the overall system is between structural features of memory and control processes (Atkinson and Shiffrin, 1967). Structural features are permanent and include the physical structure and built-in processes that may not be varied. Examples are the various memory stores. Control processes, on the other hand, are selected, constructed, and modified at the option of the subject. The use of a particular control process at some time will depend upon such factors as the nature of the task, the instructions, and the subject's own history. Examples are coding techniques, rehearsal mechanisms, and certain kinds of search processes.

The main structural components of the system are the three major memory stores: the sensory register, the short-term store, and the long-term store. Each of these stores may be further subdivided on the basis of the sensory modality of the stored information; such evidence as is available indicates that memory processes may differ somewhat depending on the sense modality involved (Posner, 1966). The sensory register accepts incoming information and holds it fairly accurately for a very brief period of time; a good example is the brief visual image investigated by Sperling (1960) and others, which decays in several hundred milliseconds. The short-term store (STS) is the subject's working memory in that the various control processes are based in it and directed from it. Information is selectively entered into STS from both the sensory register and the long-term store (LTS) and will decay from this store in about 30 seconds, except for control processes (such as rehearsal) which permit the subject to maintain the information in STS as long as desired. The long-term store is a permanent repository for information, information which is transferred from STS.

PROCESSES IN LONG-TERM MEMORY

The remainder of this paper will deal primarily with LTS, and also with STS in its capacity for handling LTS storage and retrieval. It would now be appropriate to outline our theory of long-term memory and define the most important terms that will be used. Long-term memory processes are first divided into <u>storage</u> and <u>retrieval</u> processes. These two processes are similar in many ways, one mirroring the other. Storage consists of three primary mechanisms: <u>transfer</u>, <u>placement</u> and <u>image</u>

production. The transfer mechanism is based in the short-term store and includes those control processes and mechanisms by which the subject decides what to store, when to store, and how to store information in LTS. The placement mechanism determines where the ensemble of information under consideration will be stored in LTS. It in turn will consist of directed and random components. Having decided finally where to store the ensemble of information, the image production process determines what parts of that ensemble will be permanently stored in that location of LTS. In general, not all the information desired is stored, and conversely, some unwanted information may be stored. The final ensemble of information permanently stored in LTS is called the image. This image is assumed to remain intact over time and during storage of other information. Retrieval, like storage, consists of three primary mechanisms: search, recovery, and response generation. Search is the process by which an image is located in memory, and like placement, consists of directed and random components. Recovery is the process by which some or all of the information in a stored image is recovered and made available to the short-term store, and response generation consists of the processes by which the subject translates recovered information into a specific response. We shall now turn to a detailed consideration of each of the processes outlined above.

Storage: Transfer

Transfer refers to the mechanisms by which information that has entered STS is manipulated there prior to placement in the long-term store. These mechanisms include a number of control-processes having to do with deciding what information to attempt to store, when to attempt the storage operations, and what form of coding or other storage procedure should be employed. Before describing these control processes further, it should be pointed out that transfer involves at least one unvarying structural characteristic: whenever any information resides in the short-term store, some transfer of this information can take place to long-term store. The strongest evidence for this comes from studies of incidental learning (Saltzman and Atkinson, 1954), and from experiments first carried out by Hebb (1961) and Melton (1963). In these latter experiments subjects are given a series of digit spans to perform; for each span the subject is required to repeat back in order a short sequence of digits just presented. Unknown to the subject, a particular sequence is repeated at spaced intervals. Performance on the repeated sequence improves over trials, indicating that information about that sequence is being stored in LTS, even though the nature of the task is such that the subject does not attempt to store information about the individual spans in LTS. This assumption, of course, implies that images are being stored not only during "study" periods, but whenever information is input to the short-term store: during test, during rest periods, during day dreaming, and so forth. (Most laboratory experiments are designed to insure that essentially all storage takes place during study periods, but this is not always the case.)

In many situations, especially the typical experimental paradigms, a large amount of information is being input sequentially to the short-term store. In such a situation, the short-term store will act as a time-sharing system and the subject will select some subset of the presented information for special processing in STS such as rehearsal or coding. The information not given special attention will decay and be lost from STS fairly quickly; LTS storage of this information will therefore be weak and undirected. If information is maintained in STS via simple rehearsal, but no special storage procedure such as coding is used, then the LTS image will be stronger than in the absence of rehearsal, but its placement will be quite undirected and thus the item will be difficult to retrieve at test (see Atkinson and Shiffrin, 1967). The selection of particular items for active attempts at storage will depend upon a number of factors. Items already felt to be retrievable from LTS will be dropped from active consideration; time would be better spent storing new, unknown information. There are many storage strategies the subject can adopt which result in the selection of particular items for processing: for example, in a paired-associate experiment with all responses being either X or Y, the subject might decide to store only the associates with the response X and to guess Y as a response to any unknown stimulus at test. Differential payoffs can also induce selection: items with higher payoffs being selected for storage. This phenomenon is illustrated in studies of reward magnitudes (Hurley, 1965). If two separate lists contain items with different payoffs, performance does not differ between the lists. If items <u>within</u> a list have different payoffs, however, the items worth more are preferentially selected and performance is better for them. Finally, in experiments where no great demand is made on the short-term system, all items can be given special storage procedures even if there is no need to do so.

What to transfer is dependent not only on the items presented for study, but also upon varying strategies the subject may adopt. Thus the subject may attempt to cluster several items currently in STS and store them together. This obviously occurs in serial learning tasks, and often in free-verbal recall. Sometimes all the information in the presented item is not necessary for correct responding; in these cases the subject may decide to store only the relevant characteristics of the input. Most often the subject will select relevant characteristics of the input and then add to this information other information from LTS. In coding a paired-associate for example, the subject may recover a mediator from LTS and then attempt to store the paired-associate plus mediator. Note that the ensemble of information that the subject attempts to store and the ensemble that is actually placed in LTS are by no means identical; the latter may contain a large amount of information that the subject would regard as "incidental" or useless.

How to store the selected information refers largely to the control process adopted. In most cases a consistent strategy will be adopted and used throughout an experiment. These strategies include rehearsal, mnemonics, imagery, and other forms of coding. The level of performance will be greatly affected by the strategy used, the reasons for this becoming evident later in the paper.

<u>Storage</u>: <u>Placement</u>

Placement and search are two processes that have received little systematic consideration in the memory literature but are nevertheless extremely important. Placement refers to where in LTS storage of a particular information ensemble is attempted. By "where" we do not refer to a physical location in the cortex, but to a position in the organization of memory along various informational dimensions.* These dimensions include sensory characteristics of the input (e.g., visual, auditory, or tactile storage), meaningful categorizations such as noun vs verb, or animal vs vegetable, and other characteristics such as the syntactic and temporal aspects of an item. These and other dimensions of storage will be elaborated further in the succeeding discussion.

There are two components to the placement mechanism; these will be called <u>directed</u> and <u>random</u>. Directed refers to that component of the placement mechanism which is specified by the control processes the subject is using, the information ensemble being stored, and the subject's past history of placement. Given these same conditions at a later time, the directed component will direct placement to the same LTS location. Furthermore, the search process during retrieval can follow the directed component to the same area of LTS. The second component of placement is random; it will occur as a result of local factors which change from one moment to the next and can be regarded as essentially random in nature. Thus at certain branches in the placement processes a succeeding storage attempt might select at random a different memory dimension and multiple stored images of the same information

---

*Anatomical evidence such as the Hubel and Wiesel (1962) explorations of information abstraction in the visual cortex of the cat, or the work of Penfield and Roberts (1959), or the older work on motor areas of the cortex, suggests that there may be a topographic placement mechanism. If one is trying to use a visual image to store a noun-noun pair (rather than, say, an auditory-verbal code) it would not be surprising if storage took place roughly in the area of the visual cortex. However, the form of the correspondence of the subject's informational organization of LTS with the physical structure of the nervous system is tangential to the discussion of this paper.

ensemble could result. Furthermore, during retrieval each of the random branches of placement would have to be explored via search in order to locate the stored image.

Note that the directed-random distinction is not the same as the structure-control process distinction; although random placement is not under the control of the subject, part of directed placement is also not under the subject's conscious direction. The directed component has three major determinants that will be considered in turn. The first is the kind of information in the item presented for study (and also in the ensemble selected for storage). Thus presentation in a free-recall task of a card with LION printed on it in black capital letters might lead to placement in locations determined by any or all of the dimensions: black, capitals, letters, words, animals, printed words, and so forth. In this free-recall example, as in other situations, certain storage locations will be more effective than others; storage in an "animal" location is not effective if at test the subject does not recall that he stored any words in the "animal" region. On the other hand, if the task was one of categorized free recall, in which there were a number of animals in the list to be recalled, then placement in an "animal" dimension might be very effective, especially since the first animal word recovered is likely to cause the subject to search in the "animal" region.

The second directed placement determinant is that induced by strategies the subject may select. If the strategy involves the formation of a natural language mediator for a paired-associate, then the informational content and origin of the mediator may indicate placement dimensions for storage of the pair plus mediator, perhaps in the "natural language" area. On the other hand, the formation of a visual image for coding purposes might lead to placement in the "visual area." If a cohesive strategy is used which encompasses many items, (for example, the placing of coded paired associates in the successive rooms of an imaginary house), then the placement of different items might be directed roughly to the same location.

The third placement determinant is that induced by the subject's pre-existing organizational structure and history of placement of similar information in the past. This kind of placement may often occur not under conscious control of the subject, but may nevertheless be consistent over trials. These three determinants of directed placement are necessary in order that the subject may be able to "retrace" his path and find a stored image during retrieval and search.

Either at the will of the subject or not, placement of an information ensemble may occur in more than one location in LTS. For example, the subject may encode an associate in two different ways and then store both resulting codes in each of the two locations defined by the codes. Multiple placement of this kind is said to result in <u>multiple images</u> or <u>multiple copies</u> in LTS. The extent to which multiple placement occurs in the usual experimental tasks is open to question. In some tasks, such as those in which the one-element model has been applied successfully (Bower, 1961), it would appear that a single copy assumption best fits the data. Even in these cases, however, the multiple copy models may be applied if the very first copy stored is always capable of allowing a correct response; in this case the effects of multiple storage are not observable if only correct and incorrect response data are recorded.*

It is too much to ask of a memory system that placement be entirely directed. This would be akin to a library with a complete and accurate filing system, but there are a number of reasons why such a high accuracy system would be unfeasible for the type of memory system outlined here. These reasons include the drastic consequences of small failures in such a system, and considerations of access times. Furthermore, we are assuming that placement and search are parallel processes and there is evidence that search processes at times operate more or less randomly (see Atkinson and Shiffrin, 1965). Consequently we assume that there is a considerable component of placement which is also essentially random. That is, if placement were completely directed, there would be no reason for search to be random to any degree. (We shall consider random search processes later.) Sometimes part of the directed storage may be unavailable during retrieval; that portion of the placement is then essentially random since the subject must initiate a random search to find the right storage location.

Storage: Image Production

An ensemble of information having been placed at some location for storage, the image production process determines what portion of this

*A number of interhemispheric animal studies (Sperry, 1961) have indicated that at least two copies are normally made, one in each hemisphere, but this may not involve placement. Rather, it seems that once an image has been produced, the corpus callosum is involved in an after-the-fact transfer of the image to the other hemisphere.

information is permanently stored as an image there. We cannot say much about this process except that it occurs in some partial or probabilistic manner: at test, subjects can often recall incidental material which is correct but irrelevant, even when the required answer cannot be recalled. Actually it is difficult to separate the effects of image production from those of its retrieval counterpart, recovery. Recovery refers to the extraction of information from a stored image which has been located. A conceivable method for separating these processes is based on the fact that it is sometimes possible to use cueing to elicit from a stored image information not recoverable in a first attempt.

We next consider the contents of the image: the range and form of the stored information. A single image may contain a wide variety of information including characteristics of the item presented for study (its sound, meaning, color, size, shape, position, etc.) and characteristics added by the subject (such as codes, mnemonics, mediators, images, associations, etc.). In addition, an image most probably contains links to other images (other information which was in the short-term store at the same time); these links can be regarded as a set of directions to the locations of related images in LTS. There is some question as to whether temporal information in the form of some sort of internal clock reading may be part of the image. It is our feeling that the ability to make temporal discriminations can be explained on the basis of contextual information and counting processes, rather than on the basis of a clock reading recorded on the image.

We make the assumption that images are essentially permanent; they do not decay or disintegrate over time given an intact, physiologically normal organism. This assumption is made for simplicity. We feel it is possible to propose appropriate search and storage mechanisms that explain decreases in performance over time. Some ways in which this may be done will be suggested when the outline of the system is completed.

Retrieval: Search

At test the subject is given certain cues specifying the nature and form of the required response. Assume that the information necessary to generate a response is not at that time in the short-term store. The subject will then attempt to locate the relevant image, or images, in LTS. This attempt is called the search process. The search will be monitored by the short-term store. That is, at any moment the short-term store will contain a limited amount of information such as the search strategy being employed, part of the information recovered so far in the search, what locations in LTS have been examined already, and some of the links to other images that have been noted in the search but not yet examined. The short-term store will thus act as a "window" upon LTS, allowing the subject to deal sequentially with a manageable amount of information. In addition to the directed search monitored by STS there is a random, diffuse component engendered by the information currently in STS. Thus when, say, the stimulus member of a paired-associate is presented for test, it will enter STS and at once a diffuse search is initiated by this member: as a result a number of images will be activated including many of the associates of this stimulus. There will be feedback such that activated images will be entered into STS, but this must be quite selective since STS has only a limited search capacity. Thus many activated images, possibly including the desired image, may not gain access to STS. As the search continues and new information enters STS, the diffuse pseudo-random search component will be re-elicited by the new STS information. Hopefully, a relevant image will eventually enter STS and be recognized as such.

As the above discussion has tried to indicate, there are directed and random components to the search process. The subject has a considerable amount of control over the directed component and we now consider this in some detail. As was true in placement there are three primary determinants of directed search. Search may first be directed by cues and characteristics of the information presented for test. Thus if "kaq" is presented as a test on a previously studied paired-associate, "kaq-cen," then search might be initiated along dimensions of things sounding like kaq, of words beginning with k, of nonsensical three letter combinations, and so on. On a free-recall test, search might be directed to the "most recent list of items." Secondly, search may be directed by strategies adopted by the subject. Thus a search for natural-language-mediators may be initiated following the presentation of a stimulus member of a paired-associate for test. Or perhaps a search is initiated in the region of visual images containing this stimulus member. One search strategy often used employs ordering of the search. For example, we are likely to do better when asked to name all 50 states if we search memory in an ordered fashion, say alphabetically or geographically, rather than in a haphazard fashion. Thirdly, search may be directed by historical patterns of search behavior that the subject has developed through consistent use.

In any event, to the extent that the subject can remember, he will (or should) attempt to utilize the same directed search strategy as the directed placement used during storage. If the subject stored a paired-associate via a visual image, it would clearly not be effective to search for natural language mediators at test. This provides a strong reason for a subject to utilize a single, consistent storage strategy during training, even though switching coding techniques from item to item might minimize "interference" and confusion.

In carrying out a directed search, information will be recovered from various images and placed in STS. If this information appears to be promising, perhaps in terms of its similarity to the test information, then the search may be continued in the same area and direction, either in terms of the dimensions being searched, or in terms of the links recovered from successive images. Thus the search may be visualized as a branching process with random and directed jumps. At some point it may be decided that a wrong location has been reached (a wrong branch examined); at this time the subject may return to an earlier location or branch if its whereabouts is still held in the short-term monitor. If not, a return may be made to the original test stimulus in order to restart the search.

A decision that is very important in the retrieval process concerns when to terminate an unsuccessful search; after all, the desired information may never have been stored in LTS. A number of termination rules may be adopted. In cases where the response period is restricted, the search may be terminated by the time limit. In other cases, an internal time limit may be set which, if exceeded, terminates the search.

It is likely that this internal time limit will be dependent upon the kind of information actually recovered; if this information seems relevant or close then the search may be extended considerably. Another criterion for termination might be successive search attempts ending at the same unproductive location in LTS. In some cases termination for this reason is used as a positive approach: most of us have sometimes experienced the feeling that "if I only stop thinking about it for a while I'll remember it." In certain tasks other termination rules will sometimes be applicable. In free recall, for example, a series of words is read to the subject who then tries to recall them in any order. During retrieval the subject may find that successive searches result in recovery of words already recalled; in this case a termination rule might be based on the number of successive recoveries of words already recovered.

Of equal importance to the termination rule for an unsuccessful search is the termination rule for a "successful" search. That is, it will often happen that partial or incomplete information is recovered such that the subject is uncertain whether a particular response is appropriate. Similarly, some portion of the response might be recovered and a decision must be made whether to continue the search for the remainder, or to guess based on the partial information. Decisions in this case are probably based on available response time, payoffs for correct or fast responses, probability of correct guessing, and so forth. Termination criterion of this sort are closely related to the response production process which will be considered shortly.

Retrieval: Recovery

Once an image has been located, it is appropriate to ask what information contained in the image will be entered into the short-term store. This process is called recovery. To an extent, recovery of part or all of the stored information will be probabilistic, depending upon such factors as the current noise level in the system. Furthermore, as noted earlier, since the short-term monitor is limited and selective not all recoverable information will be entered into STS. This problem will tend to arise in fast large-scale random searches, in which large amounts of information may be activated with relatively little of this information being relevant. Thus in any particular situation the recovery of all the information in a stored image is by no means certain. The recovery process could conceivably be isolated from the others outlined so far by utilizing various cueing conditions at test to try and make more and more of the stored information available.

Retrieval: Response Generation

Having terminated the search and recovered information from LTS, the subject is faced with the task of translating this information into the desired response. Actually, a fair amount of experimental work has examined this aspect of retrieval and our remarks here will not be particularly novel. It should be pointed out first that when we speak of recovery of information we do not imply that this information will be verbalizable or directly available in the conscious experience of the subject. In some cases partial information may result in nothing more concrete than a feeling of familiarity on the part of the subject. Thus, in many cases this aspect of the subject's performance might be well represented by a decision-theoretic model in which the subject is attempting to filter information through a noisy background (e.g., see Wickelgren and Norman, 1966; Bernbach, 1967; Kintsch, 1967). A good part of the response generation process consists of what can be called the guessing strategy. In general, guessing refers to the subject's selection of a response on the basis of partial information. There are a large number of guessing strategies that can be adopted and they will not be considered in detail here. It should be realized, however, that the probability of a correct response may not always be related in an obvious way to the amount of information recovered; guessing strategies can complicate matters. For example, in a paired-associate experiment where a list of stimuli is mapped on to two responses X and Y, the subject may store only information about stimuli with response X and then always guess response Y when a stimulus is tested for which no information can be retrieved. In this case, no information will be recovered about Y pairs, but they will always be responded to correctly. This serves to emphasize again the importance of control processes in even the simplest experiments.

DISCUSSION

We have now traced information from its presentation through storage, retrieval and output. We have not described ways in which performance will decline with time and intervening items. One way in which this can occur involves the storage of an increasing number of images, without a corresponding increase in the accuracy of the placement and search processes. In order to illustrate this point, and also indicate how the system may be applied in an actual situation, we may consider free-verbal recall. A number of lists of words are read to a subject. Following each list the subject attempts to recall as many of the words in the preceding list as possible, in any order. Two results of interest here are the facts that there are almost no intrusions from preceding lists, and that performance decreases as list length increases (Murdock, 1962). These effects are found even if short-term storage is obliterated (Postman and Phillips, 1965; Atkinson and Shiffrin, 1965), so we shall consider this experiment only from the point of view of LTS. One interpretation of the lack of intrusions would hold that the placement process directs information about successive lists to separate locations in LTS, and at test a directed search is made only of the most recent location. Let us assume that within a list, information about individual words is stored in a non-directed fashion in that list location. Call the amount of information stored for the i^{th} word, S_i. Then the amount of information stored altogether in the most recent list location will be $\Sigma S_i = S$. At test the search process is immediately directed to the most recent list location, but the search is random within that area. Assume that $\underline{n}$ random searches are made in this area during the time allotted for responding. By random search we mean that the probability of finding an image relevant to word i on a search will be S_i/S. The probability of recovering information from that image and then generating the correct word will depend of course upon the amount of information, S_i. Suppose that performance is the result of $\underline{n}$ independent random searches of this kind. What then will happen to performance as list length increases? S_i will remain the same but

$\Sigma S_i = S$ will increase. Since the probability of "hitting" any image on a search is S_i/S, this probability will decrease with an increase in list length. Thus decreases in performance with increasing list length can be explained with reference to problems inherent in the storage and retrieval processes, without the necessity of assuming loss of information from stored images.

This free-recall model has been applied successfully to a large amount of data (Atkinson and Shiffrin, 1967). The model is particularly interesting because it utilizes all three retrieval processes outlined in this paper. The directed search refers to location of the most recent list. A random search is then made within that list location. Images identified in the search may or may not have information recovered from them. The amount of information recovered then determines the probability of correct response generation.

The free-recall model is one possible application of the system described in this paper. Despite its relative success, the assumption that placement is random within a list location is probably only roughly correct at best. Certainly most subjects tie together some of the words within a list (Mandler, 1967; Tulving, 1962). Furthermore, the search itself may not be nearly as random as was assumed. A situation in which these possibilities are accentuated is that of categorized free recall (Cohen, 1963). In this type of experiment a number of the words within a single list fall into well-known categories (e.g., months of the year, numbers from 0 - 9, kinds of monkeys, etc.). In this case we would probably expect both placement and search to be directed down to the level of the category, rather than the level of the list. A model which seems to work well for this type of task assumes that the initial search is random within a list location, but once one member of a category is reported a directed search is made through the other members of the category, with any presented item in the category having a constant probability c of being recovered.

Another question we might consider in our framework is the source of differences in performance between recognition and recall procedures. One primary source arises in the response generation process: the recovery of partial information in the search will lead to better performance in recognition than in recall. For example, being able to recover the first letter of a response may guarantee perfect performance on a recognition test, but virtually chance responding for recall. Another source found in paired-associate tasks is related to the search process: recall provides only one member of the pair, and location of the stored image must be based on cues provided by this single member. In recognition, however, both a stimulus and a response member are presented and search for the relevant image in LTS may be based on cues provided by either or both members. Finally, another source of difference between performance in recall and recognition may be found in the storage process: expectation of a recognition test may allow easier storage than expectation of a recall test. That is, less detailed information would need to be stored about an item if the tests were recognition rather than recall. This might permit storage of items that would otherwise have been ejected from STS for lack of time to deal with them. One test of storage versus retrieval effects was carried out by Freund, Brelsford, and Atkinson (1967). At study a paired-associate item was presented and the subject was told he was either going to be tested by recall, by recognition, or he was not told which form of test would be used. Comparison of performance for the four types of items (told recall-tested recall, told recognition-tested recognition, not told-tested recall, or not told-tested recognition) allows storage and retrieval effects to be separated. Using this design it was established that differences between recognition and recall depended on differences in retrieval and not on storage. However, it seems clear that the results depended upon the specific stimulus materials used; with appropriate stimulus materials storage differences might also be detected.

It is sometimes implicitly assumed by memory theorists that recognition tests (yes-no or old-new tests in the simplest cases) eliminate retrieval effects and that differences between the various recognition procedures may therefore be attributed to storage. This assumption would be most parsimonious if true, but there is insufficient evidence to justify it. From our viewpoint there is reason to assume that retrieval effects are not eliminated by using recognition tests. In some recognition tasks it is clear that search effects are present. For example, if a paired associate is presented and the subject is asked whether the correct response is being displayed with the stimulus, one procedure the subject will use is to search memory, find the correct response, and compare it with the one presented. Thus, even in the simplest cases it is likely that recognition involves a variety of retrieval and search processes. In this regard we can point to several factors which might favor recall over recognition tests. The recognition condition may cause a premature termination of the search process because the subject thinks he can correctly identify a given response, while an extended search would recover the correct one. In a recognition task where an incorrect response alternative is displayed, the incorrect alternative may initiate inappropriate search patterns that consume time and otherwise hinder performance.

The above discussions illustrate one of the benefits of introducing a highly structured, albeit speculative, long-term memory system. Such a system can be quite productive of alternative explanations for a wide range of memory phenomena that less structured systems may not deal with effectively. This in turn leads to experiments designed to determine which explanations are applicable in which situations. It is unfortunately beyond the scope of this paper to apply the system to the many experimental results in long-term memory. Nevertheless, we hope that it has been of some value to outline the theoretical system. Parts of the theory have been incorporated in models for a variety of experiments (Atkinson and Shiffrin, 1965, 1967) but the overall framework has not previously been elaborated.

In this paper no attempt was made to compare our system with extant theories of long-term memory. Most of the current theories have been presented at a somewhat more general level than was used here, and the present system may therefore be liberally interpreted as an extension and elaboration of certain ideas already in the literature.

References

Atkinson, R. C., and Shiffrin, R. M. Mathematical models for memory and learning. Technical Report 79, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965. (To be published in D. P. Kimble (Ed.), _Proceedings of the third conference on learning, remembering and forgetting_. New York: New York Academy of Science.)

Atkinson, R. C., and Shiffrin, R. M. Human memory: A proposed system and its control processes. Technical Report 110, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1967. (To be published in K. W. Spence and J. T. Spence (Eds.), _The psychology of learning and motivation: Advances in research and theory_, Vol. 2. New York: Academic Press.)

Bernbach, H. A. Decision processes in memory. _Psychological Review_, 1967, _74_, 462-480.

Bower, G. H. Applications of a model to paired-associate learning. _Psychometrika_, 1961, _26_, 255-280.

Cohen, B. H. Recall of categorized word lists. _Journal of Experimental Psychology_, 1963, _66_, 227-234.

Freund, R., Brelsford, J. W., Jr., and Atkinson, R. C. Recognition versus recall: Storage or retrieval differences? Paper presented at the Psychonomic Society Annual Meeting, Chicago, October, 1967. ~~(Mimeographed copies available.)~~

Harley, W. F., Jr. The effect of monetary incentive in paired-associate learning using an absolute method. _Psychonomic Science_, 1965, _3_ 141-142.

Hebb, D. O. Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), _Brain mechanisms and learning_. London and New York: Oxford University Press, 1961. Pp. 37-46.

Hubel, D. H., and Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. _Journal of Physiology_, 1962, _160_, 106-154.

Kintsch, W. Memory and decision aspects of recognition learning. _Psychological Review_, 1967, _74_, 496-504.

Mandler, G. Organization and memory. In K. W. Spence and J. T. Spence (Eds.), _The psychology of learning and motivation: Advances in research and theory_, Vol. 1. New York: Academic Press, 1967. Pp. 328-372.

Melton, A. W. Implications of short-term memory for a general theory of memory. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 1-21.

Murdoch, B. B., Jr. The serial position effect of free recall. Journal of Experimental Psychology, 1962, 64, 482-488.

Penfield, W., and Roberts, L. Speech and brain mechanisms. Princeton, N. J.: Princeton University Press, 1959.

Posner, M. I. Components of skilled performance. Science, 1966, 152, 1712-1718.

Postman, L. and Phillips, L. W. Short-term temporal changes in free recall. Quarterly Journal of Experimental Psychology, 1965, 17, 132-138.

Saltzman, I. J., and Atkinson, R. L. Comparison of incidental and intentional learning after different numbers of stimulus presentations, American Journal of Psychology, 1954, 67, 521-524.

Sperling, G. The information available in brief visual presentations. Psychology Monograph, 1960, 74 (Whole No. 498).

Sperry, R. W. Cerebral organization and behavior. Science, 1961, 133, 1749-1757.

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

Wickelgren, W. A., and Norman, D. A. Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 1966, 3, 316-347.

SEARCH AND RETRIEVAL PROCESSES IN LONG-TERM MEMORY

by

Richard M. Shiffrin

TECHNICAL REPORT NO. 137

August 15, 1968

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

ACKNOWLEDGMENTS

I am indebted to Professor Richard C. Atkinson, chairman of my dissertation committee, for his frequent advice and encouragement during the course of this dissertation. Thanks are due to Professor William K. Estes and Professor Gordon H. Bower, who also served on my dissertation committee, and to Professor Walter Kintsch, for their comments at various stages of this paper. Michael S. Humphreys was instrumental in the design of Experiment I, and has engaged in valuable conversations with the writer. Helen Butler and her staff have been invaluable in the final preparation of the paper, particularly Maria Jedd who painstakingly prepared the many figures. Thanks are due to Carolyn Young who typed the major portion of the final manuscript. Finally, thanks are especially due to my wife Susan, for typing numerous rough drafts and a portion of the final draft, for providing cheerful encouragement, and for relieving me of many concerns extraneous to this project.

The research reported here was accomplished while I held a National Science Foundation Terminal Graduate Fellowship. Support for the research was provided, in part, by the National Aeronautics and Space Administration Grant NGR-05-020-244.

TABLE OF CONTENTS

		Page
Acknowledgments		iii
List of Tables		vi
List of Figures		vii
Introduction		1
Chapter		
I	A Theory of Storage and Retrieval in Long-Term Memory	3
	The Memory System	3
	Storage and Retrieval	5
	Storage	8
	Retrieval	13
	Search	13
	Recovery	17
	Response Generation	17
	Search Termination	18
	Applications and Extensions	19
	Recognition and Recall	19
	Ranking	20
	Second-Guessing	20
	Interference Phenomena	22
	Latencies	23
II	The Experiments: Design, Procedure, and Results	25
	Experiment I	25
	Design Justification	25
	Design	28
	Subjects	31
	Apparatus	31
	Stimuli and Responses	31
	Instructions	32
	Procedure	33
	Results of Experiment I	35
	Ranking Performance vs. Second-Guessing Performance	37
	Learning and Forgetting	41
	The Effects of Intervening Items	48
	Summary	52
	Experiment II	53
	Design Justification	53
	Subjects	57
	Stimuli and Responses	57
	Instructions	59

TABLE OF CONTENTS (cont.)

Chapter		Page
	Procedure	60
	Results of Experiment II	62
	First-Response Data	64
	Second-Guess Data	70
	Latencies	75
	Conclusions	82
III	Theoretical Analysis: A Storage and Retrieval Model	84
	Experiment I	84
	The Short-Term System	84
	Storage	86
	Retrieval	88
	Review of the Model	93
	Mathematical Analysis	95
	Predictions of the Model	98
	Experiment II	101
	Mathematical Analysis	104
	Predictions of the Model	106
	Extensions of the Model	114
	Concluding Discussion	115
Appendix 1		119
Appendix 2		121
Appendix 3		123
Appendix 4		126
References		128

List of Tables

Table		Page
II-1	Sequence of Lags for Item-Types of Experiment I	29
II-2	Mean Probability Correct for Subjects of Experiment I . .	36
II-3	Mean Probability Correct for Successive Days of Experiment I	36
II-4	Probability Correct as a Function of the Average State of Knowledge Concerning the Items Making up the Preceding Lag	36
II-5	Item-Types for Experiment II	58
II-6	Mean Probability Correct for Subjects of Experiment II .	63
II-7	Mean Probability Correct for Successive Days of Experiment II	63
II-8	First-Guess Intrusions	69
II-9	Second-Guess Intrusions	73
II-10	Second-Guesses Following Old-Intrusions as First-Guesses	76
II-11	Mean Latency of Correct First Responses	77
II-12	Mean Latency of First-Guess Old-Intrusions	78
II-13	Mean Latency of First-Guess New-Intrusions	79
II-14	Mean Latency for Correct Second-Guesses Following Old-Intrusions	80
II-15	Mean Latency for Correct Second-Guesses Following New-Intrusions	81

List of Figures

Figure		Page
I-1	An LTS Search in a Continuous Memory Task	15
II-1	Conditional Probabilities of Second-Guessing and Second-Ranking	39
II-2	Probability Correct as a Function of Lag, for Rankings and Rerankings	43
II-3	Probability of Correct First Ranking as a Function of Number of Presentations, for Item-Types 1-6	44
II-4	Probability of Correct First Ranking as a Function of Number of Presentations, for Item-Types 7-13	45
II-5	Probability of Correct First-Reranking as a Function of Number of Presentations, for Item-Types 1-6 . . .	46
II-6	Probability of Correct First-Reranking as a Function of Number of Presentations, for Item-Types 7-13 . . .	47
II-7	Probability of First-Guess Correct Responses and First-Guess Intrusions, for the Major Item-Types . .	65
II-8	Probability of Correct First-Guesses as a Function of Lag, for the Matrix Item-Types	66
II-9	Probability of Intrusions for New Items, as a Function of Duration of Session	71
II-10	Probability of Second-Guess Correct Responses and Second-Guess Intrusions, for the Major Item-Types . .	72

INTRODUCTION

In the past fifteen years, there has been an increasing interest in theories of human memory that consider storage and retrieval to be probabilistic processes that may vary randomly from one moment to the next. These theories for the most part can be regarded as variants of Stimulus Sampling Theory (Estes, 1959; Atkinson and Estes, 1963), and stimulus fluctuation theory (Estes, 1955a,b). A fairly large number of memory variables have been analyzed by quantitative, mathematical models within this framework. Heretofore these models have tended to be quite restrictive, their range of application being limited to a small number of variations within simple situations. In addition, these models have been concerned primarily with the memory acquisition process rather than the memory loss process. This report attempts to extend this earlier work by introducing a theory which can deal quantitatively and simultaneously with many of the variables previously examined individually, and which will deal as extensively with forgetting as learning. The theory is formulated in the spirit of Stimulus Sampling Theory, but due to the complexity of the data examined, is not a direct extension of the earlier models which have largely taken the mathematical form of multi-state Markov models.

The theory is conceived of as a quantitative alternative to primarily qualitative theories such as "two-factor theory" (Postman, 1961), although the variables dealt with in the two cases do not entirely overlap. The direct antecedents of the present work are the theoretical papers of Atkinson and Shiffrin (1965, 1968) and Shiffrin and Atkinson

(1968). As a result, the theory is primarily concerned with an elaboration of a complex search and retrieval process from long-term memory.

Chapter I of the present report outlines the general framework of the theory. Chapter II describes and presents the results of two experiments designed to provide a wide range of data to test a quantitative version of the overall framework. The first experiment is concerned with the probabilistic nature of retrieval, and forgetting of individual items. The second experiment is concerned with intrusion phenomena in responding, and with interference phenomena following the altering of the response assigned with a stimulus. A number of other variables which are examined will be described in the text. Chapter III presents a specific quantitative model based on the theory of Chapter I, and applies it to the results of the two experiments.

CHAPTER I

A THEORY OF STORAGE AND RETRIEVAL IN LONG-TERM MEMORY

This chapter begins with a brief survey of the human memory system, largely following the format of Atkinson and Shiffrin (1965, 1968). The report will then turn to a detailed discussion of a theory of storage and retrieval for long-term memory. Although the system is meant to be quite general, the theory will be described as it applies to a continuous paired-associate learning task. Such a task consists of a series of anticipation trials. On each trial a stimulus is presented for test and then paired with a response for study. The task is called continuous because new stimuli are continually being introduced at randomly spaced intervals. The theory is described in relation to this task because it is the one utilized in the experiments described in Chapter II.

The Memory System

It has proved of value (Atkinson and Shiffrin, 1968) to dichotomize memory processes on a dimension of subject control. Thus, on the one hand, there are "structural processes" which are permanent, unvarying features of the memory system, features which may not be modified at the will of the subject. On the other hand are "control processes" which are selected, constructed, and used at the option of the subject, and may vary greatly from one task to another. This distinction was set forth in great detail in the report cited, and will not be belabored here. In the remaining portions of this chapter it will be clear that most of the processes discussed, from storage mechanisms to search schemes, are under subject control to one degree or another. Except where special emphasis is required, the distinction between structural and control processes will not be stated explicitly.

The three major components of the memory system are the "sensory register," the "short-term store" (STS), and the "long-term store" (LTS). The sensory register accepts incoming sensory information and holds it very briefly while it is given minimal processing and then transferred to STS. If a large amount of information is presented quickly, then only a portion of this information can be transmitted to STS, and the precise characteristics of the sensory register will become quite important. In the experiments to be considered in this report, however, the presentation rates are slow enough, and the information quantities are small enough, that the information presented can be assumed to transit the sensory register and enter STS essentially intact. In the following, then, discussion of the sensory register will be omitted.

The short-term store is the subject's working memory; it is used for the momentary holding of information utilized by control processes such as the storage mechanisms and search schemes. Information will decay and be lost from this store within about 30 seconds or less if unattended, but may be maintained there indefinitely by rehearsal. In some situations, such as those discussed in Section 4 of Atkinson and Shiffrin (1968), the primary function of STS is one of memory -- that is, information will be maintained there via rehearsal from the time of presentation until the moment of test. The situations in which STS assumes this function are ones in which the study-test intervals are short, interference is high, and long-term learning is difficult. In other situations, such as the ones examined in this report, the memory function of STS is utilized in a different manner; STS is used for the temporary holding of information needed for long-term processing. Thus information needed for coding and search schemes is temporarily stored in STS. Although STS is utilized for the transient handling of information, it is not utilized for maintenance of the information until the moment of test.

The long-term store is a permanent repository for information. It will be assumed that information once stored is never thereafter lost or eliminated from LTS, but the subject's ability to retrieve this information will vary considerably with such variables as time and the amount of intervening, interfering material. The interaction between STS and LTS, in terms of the mechanisms and stages of storage and retrieval, is the main concern of this chapter. We turn to these considerations directly.

Storage and Retrieval

The discussion here follows the terminology of Shiffrin and Atkinson (1968). Storage refers to the set of processes by which information initially placed in STS is examined, altered, coded, and permanently placed in LTS. Retrieval refers to the inverse operations by which desired information is sought for, recovered, and emitted at test. It is convenient to subdivide both storage and retrieval into three components. The components of storage are "transfer," "placement," and "image-production." The transfer mechanism includes those control processes by which the subject decides what to store, when to store, and how to store information in LTS. The placement mechanism determined the LTS location in which an ensemble of information under consideration will be stored. Image-production is the process by which a portion of the information ensemble presented for storage will achieve permanent status in LTS. The components of retrieval are "search," "recovery," and "response-generation." Search is the mechanism by which an image is located in memory. Recovery is the mechanism by which some or all of the information in a stored image is recovered and made available to the short-term store. Response generation consists of the processes by which the subject translates recovered information into a specific response.

Before detailing the above processes, there are several general comments to be made about LTS as a whole. First, the use of the term "location" is not meant to imply necessarily a specific cortical area; rather, an LTS location is a psychological construct used to denote closeness of storage. The closer the location of two stored images, the more likely the examination of one will occur jointly with the examination of the other. Thus to say an image is stored in a single LTS location is to imply that the information in the image will tend to be recovered together. Second, a number of different terms will be used to denote an ensemble of information stored in some LTS location: ensemble of information, image, and code will be used interchangeably.

Finally, the structure of LTS may be clarified by an analogy with computer memories. A location-addressable memory is the normal computer memory; if the system is given a memory location, it will return with the contents of that location. A content-addressable memory is constructed so that the system may be given the contents of a word and will return with all the memory locations containing those contents. A location-addressable memory must be programmed before this is possible: an exhaustive search is made of all memory locations and the locations of all matches recorded. There are two primary methods for construction of content-addressable memories. In one, a fast parallel search is made of all locations simultaneously, with a buffer recording the locations of matches. In the other, the contents themselves contain the information necessary to identify the location where those contents are stored. This latter possibility can occur if the information is originally stored in accord with some precise plan based on the contents, as in some form of library shelving system. When followed at test, this storage plan will lead to the appropriate storage location. For example, a library with a shelving system based on the contents of books would store a book on the waterproofing techniques for twelfth century Egyptian rivercraft in a very precise location. When a user later desires a book with these contents, the librarian simply follows the shelving plan used for storage and directly reaches the storage location. This type of memory will be termed self-addressing. The point of view adopted in this report is that LTS is largely a self-addressing memory. That is, to a fair degree of accuracy, presented information will lead at once to a number of restricted locations where that information is likely to be stored. To give this discussion concrete form consider an experiment in which a series of consonant trigrams are presented and the subject's task is to tell whether each one has been presented previously or not. Suppose JFK is presented. In a location-addressable memory an exhaustive search would be carried out comparing JFK with each stored code. In a content-addressable memory of the first type, a parallel search is carried out which gives the locations of codes containing JFK. We assume, however, that LTS is self-addressing; hence a search is at once made of those locations where JFK is momentarily most likely to be stored. These locations are defined by a number of fairly restricted areas. The long-term store is assumed to be only partially self-addressing in that a search must next be initiated within each probable area to determine whether the desired information is indeed present. We now turn to a detailed discussion of storage and retrieval.

Storage

It is convenient to discuss the three components of the storage process in an order opposite to that normally obtaining. Thus we consider first the image-production mechanism. Image-production refers to the process by which some portion of an ensemble of information directed to some LTS location is permanently fixed there. The subject can control this mechanism in two primary ways. In the first, the subject may control the number of presentations of the information ensemble, more repetitions resulting in a larger proportion of information stored in the final image. In the second, the duration of the period of presentation may be controlled by the subject -- the longer the period during which the information resides in STS, the larger the proportion of information stored. Apart from these means, image production is beyond the control of the subject. In many applications it will simply be assumed that a random proportion of the presented information will be permanently stored.

No distinction will be made in this report between the quality and quantity of stored information; rather each image, or portion of an image, will be described by a strength measure which lumps both quality and quantity. The strength of an image will be a number between 0 and ∞, the higher the number the greater the strength. In the paired-associate situation, it is necessary to consider three strength measures, one describing stimulus related information, one describing response related information, and one describing stimulus-response associative information. This varied information may or may not be stored in the same LTS location. Specifically, it will be assumed that the stimulus information stored will have a strength distribution $F_s(I)$, the response information will have a strength distribution $F_r(I)$, and the associative information will have a strength distribution $F_a(I)$. (It should be apparent that these measures may be partially independent from each other. For a given stimulus-response pair, the subject may store information solely concerned with the stimulus, solely concerned with the response, or partially concerned with their association; these measures may even be stored in separate locations.) The form of the three distributions above will vary according to the experimental task and the techniques of storage adopted by the subject, but in general will have some spread. For example, a "good" stimulus-response pair is one that will typically result in a larger amount of stored information than a "bad" pair.

The placement process determines where information shall be stored. As pointed out previously, LTS is assumed to be largely a self-addressing memory; hence the information stored will partially direct itself to its own storage location. Thus a visual image of a cowboy will be stored in the appropriate region of the visual area of LTS. From a different point of view, it may be seen that placement will be determined by the form of the code adopted by the subject. A visual code will result in a different storage location than an auditory code. A mediator may establish its own storage location; for example, the pair QWZ - 64 may be stored via use of the mediator "the 64,000 dollar question," and the location used may be in the "television-quiz-show" region of LTS. In a paired-associate task, (when inter-pair organizational schemes are not feasible, as in continuous paradigms), the placement method yielding the best performance is one in which the location of storage is as unique as possible while simultaneously being recoverable at test. Since the stimulus is presented at test, it is most efficient to store in a location determined by stimulus information. Experiments demonstrating the relative efficacy of, say, visual imagery instructions as opposed to no instructions, demonstrate that subjects are not often aware of the most effective placement techniques to be utilized. Considerable subject differences are often found in long-term memory experiments for this reason.

The transfer process consists of subject decisions and strategies detailing what to store, when to store, and how to store information currently available in STS. It is a rather important process in most experiments because of the high degree of control that the subject exerts over it. When to store is the first decision that must be made. Consider a new paired-associate that has not been seen previously; the subject must decide whether to attempt to encode this pair. If the study time is long enough, and if the presented information is simple enough, then a coding attempt may always be made. In most experiments, however, these conditions are not met, and the subject will not find it feasible to attempt to encode every item. In this event, the decision to encode will be based upon momentary factors such as the expected ease of encoding, the time available for encoding, the importance of the item, the extent to which the item fits into previously utilized storage schemata, and so forth. In continuous experiments with homogenous items, these factors will vary randomly from trial to trial and we may assume that α, the probability of attempting to store a new item, is a parameter of a random process, and identical for each new item presented. The same holds for a previously presented item about which no information can currently be retrieved from LTS. In this latter case, however, the image stored will be in a different location than the unretrievable previous image; thus an item may have two or more codes stored in LTS over a period of reinforcements. At a subsequent test the information in each of these codes will have some chance of retrieval. If an item is currently retrievable from LTS when presented for study, then the subject has several options. When sufficient time is available for study, the subject may decide to store a new code in a new location. With less time available, information may merely be added to the current code. In complex tasks with short study periods the subject may be satisfied with simply tagging the current code with temporal information that will update it to the present time.

When a stimulus that has previously been presented with one response, called R1, is presented for study with a new response, called R2, several mechanisms may come into play. Either instructional set or individual initiative may lead a subject to add the information encoding the R2 response to the code for the R1 response (if this code is present in LTS and currently retrievable); this mechanism can be called "linking" or "mediating." Mediating is especially useful if a future test will require that both the R1 and R2 responses be given. In other situations, especially those where the subject is instructed to "forget" the R1 pairing when the R2 pairing is presented, the R2 pairing may be coded in independent fashion and stored in a new location. As was the case for a new item, it is assumed that the probability of attempting to code is a parameter α_0, which may be different than α. Note that there is no assurance that α or α_0 will not change from one reinforcement to the next. Especially in list structured experiments, there may be increasing incentive for coding unretrievable items as learning proceeds. However, in the continuous tasks we shall be discussing, it is not unreasonable to expect this probability to remain constant over successive reinforcements.

Each of the components of the storage process are accomplished by the subject via one action: the generation and maintenance in STS of the information intended for storage. It is assumed that information is transferred to LTS from STS during the period that the information resides in STS.*

---

*Throughout this paper, transfer of information is not meant to imply that the information is removed from one location and placed in another. Rather, transfer implies the copying of information from a location without affecting it in any way.

Retrieval

When a test occurs the subject will first search STS and then LTS for the desired information. The STS search is assumed to be a relatively fast and accurate process compared with the LTS search. In the following, we shall consider only the case where the desired information is not found in STS, and the retrieval process will be considered solely as it applies to LTS. LTS retrieval is assumed to take place as follows. The search process generates an image to be examined. The recovery process makes some of the information contained in this image available to STS. Finally, response-production consists of decisions concerning whether to output a response found, whether to cease searching, or whether to continue the search by examining another image. The search continues until it terminates of its own accord, or until an external time limit of the experimental procedure has expired. Retrieval is best described as a rather complex sequential search scheme.

Search. Because memory is assumed to be partially self-addressing, a stimulus presented for test will at once lead to a number of likely LTS locations where information about that stimulus may be stored. In certain cases the stimulus will have some characteristic so salient that a storage location is defined uniquely and precisely. This location will then be examined. If the experiment is such that certain stimuli presented for test may be new (not presented previously), and if no stored information is found in the location indicated, the subject may decide that the stimulus is new, and cease further search. There will be a bias mechanism determining how much information must be present for the search to continue. In most cases, the information required will be extremely minimal, since the coded image itself may be stored in a location other than the one indicated by the salient stimulus characteristic.

Regardless of the salience of the stimulus characteristics, the images or codes examined will initially be determined by stimulus information $[F_s(I)]$. That is, the locations in memory to be examined will be roughly indicated by information contained in the stimulus presented. Within the regions thus indicated, an image will be chosen for examination partly on the basis of recency (temporal information stored), partly on the basis of its strength, and partly on the basis of chance. Once the search has begun successive images examined will depend not only upon stimulus information, but also upon associative information recovered during the search. In a continuous paired-associate task the conception of the search may be simplified somewhat, as illustrated in Figure I-1. We first define a "subset" of codes in LTS which will eventually be examined if the search does not terminate via a response recovery and output. This subset will be termed the "examination-subset." It is then possible to consider the order of search through this subset. Figure I-1 portrays this process. The stimulus of the paired-associate labeled number <u>18</u>, on the far left, has just been presented for test, on trial 70. The second row from the bottom in the Figure gives the sequence of presentations preceding this test. The third row from the bottom gives the images stored in LTS for each item presented, where the height of the bar gives the strength of the code stored (lumping stimulus, associative, and response information.) The fourth row from the bottom gives those codes that are in the examination-

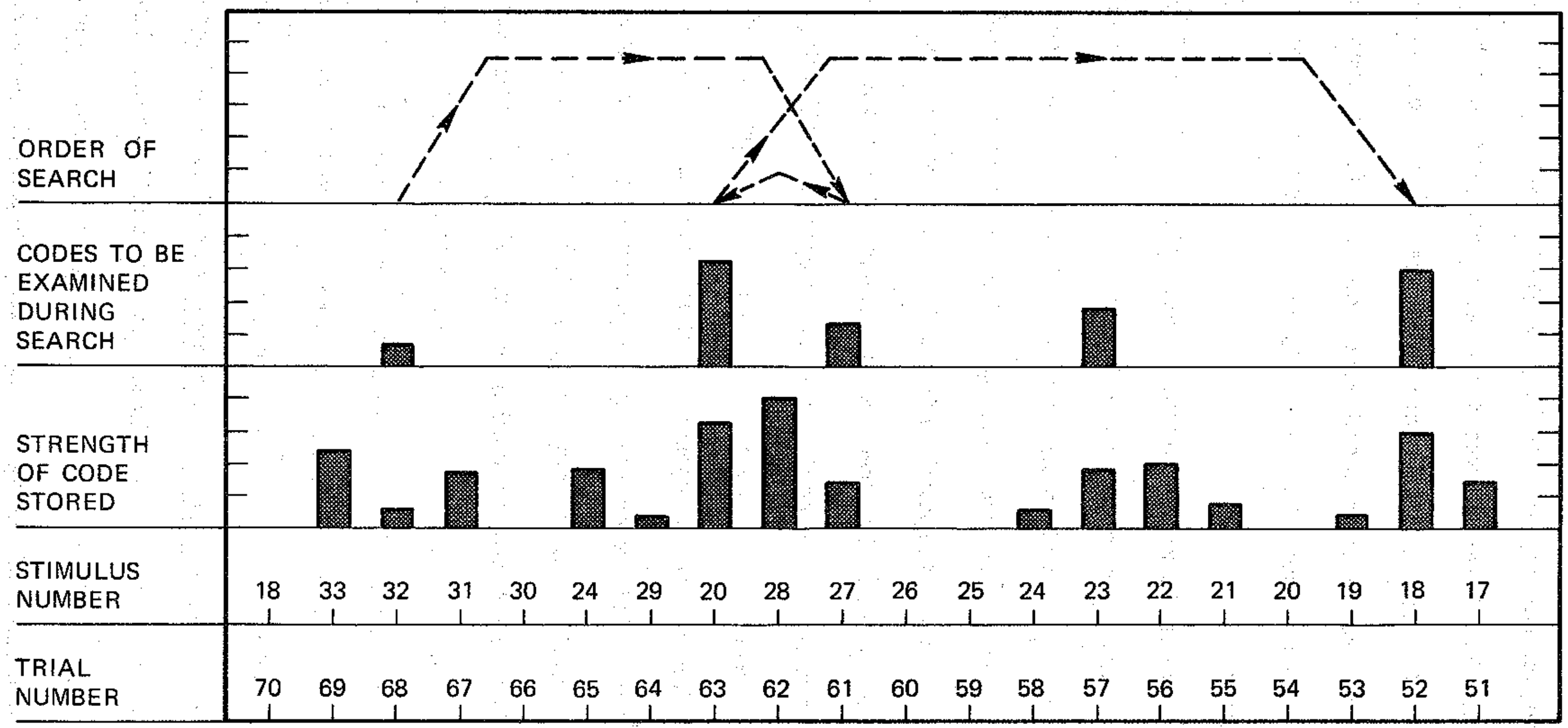


Figure I-1. An LTS Search in a Continuous Memory Task.

subset. The arrows on the top of the Figure give the order of search through the subset. Thus item 32 was first examined and rejected, then item 27, then item 20. Finally, the code for item 18 was examined, the response coded there was recovered and accepted, and the search ended with a correct response. Note that item 23 was not examined because the search terminated.

In continuous tasks it may be assumed generally that the order of search through the subset of codes is a function both of the "age" and strength of the codes involved, where age is related to the number of items that have intervened between storage of a code and the present test. It seems clear that temporal information must be an important determiner of search order. In free recall tasks, for example, successive series of items are presented to the subject. Following each series, the subject attempts to output the members of the series. The important finding for present purposes is that intrusions from one series in the responses for a following series are extremely rare; apparently subjects can order their search temporally so that only the members of the most recent list are examined during retrieval. The question of the degree to which search order depends upon temporal factors will be examined in Chapters II and III, and will not be discussed here.

There are several factors which help determine which codes will be in the examination-subset. Denote the image which encodes the pair currently being tested as a c-code. A c-code should have a higher probability of being in this subset the higher its strength (primarily the amount of its stimulus information). Other images, denoted i-codes, should have a probability of being in the subset which is greater, the greater the degree of generalization between its stimulus information and the stimulus being tested. In general, however, i-codes will have a much smaller probability of being in the subset than a c-code of equal strength. As a result, the total number of codes making up the subset of codes to be examined may be fairly small.

Recovery. Recovery refers to the extraction of information from the image under examination. The recovery of a desired complex of information, if this information is actually encoded in the image under examination, should be a monotonic function of the strength of the image. A number of decisions are dependent upon the outcome of the recovery process. Stimulus information recovered is largely responsible for accepting or rejecting the image as containing the desired response. That is, regardless of response information recovered, if the stimulus information is discrepant with the stimulus being tested, then the search will skip by this image and continue elsewhere. Response information recovered allows the subject to emit the encoded response. Associative information recovered will often serve the purpose of directing the search to a different LTS location where an image encoding the response may be stored.

Response Generation. Following recovery of information from an image, a decision process must be utilized to decide whether to emit a response, and if so, what response. It will normally be the case that the stimulus information recovered from a c-code will be congruent with the stimulus being tested, and a decision will then be made to attempt to output the response if at all possible. Whether a response can be emitted will depend upon the response information recovered. In cases where the response set is well delineated, a criterion is assumed to be set which will monitor the sensitivity of the output process. If the criterion is set quite low, then many responses will be emitted, but they will often be wrong. If the criterion is set quite high, few responses will be given, but these will almost always be correct. For i-codes the probability of emitting a response will be considerably lower than for c-codes; this occurs because output may be suppressed when the recovered stimulus information does not match the stimulus being tested. Thus a response will be emitted after examination of an i-code considerably less often than after examination of a c-code. In some applications (as in Chapter III) the recovery and response generation processes will be lumped for simplicity into a single process. In this event the probability of output of the response encoded will be a function of the strength for c-codes. For i-codes the strength will be multiplied by a generalization parameter less than one; the resultant quantity will be termed the "effective strength" of the i-code. The probability of output will then be the same function as for c-codes, but the function will be based upon the effective strength of the i-code. This scheme will be discussed fully in Chapter III.

Search Termination. Depending upon the task, a variety of mechanisms help determine when the search ceases. If the test interval is quite short, then the search may continue until a response is output or time runs out. Furthermore, if the test interval is short, the subject may output the first likely response recovered in the search. When longer response periods are available, then the search might be allowed to continue until a number of likely responses are recovered; these responses will then be evaluated and a first choice chosen for output. When sufficient time is available, the subject may adopt one of a number of sophisticated termination schemes. These were discussed in Atkinson and Shiffrin (1965) and will not be discussed further here.

Applications and Extensions

We shall next consider applications of the theory to a variety of manipulations which may be carried out in the context of a continuous paired-associate design. Primarily we shall discuss those variations which were actually employed in the experiments presented in Chapter II.

Recognition and Recall. In a recognition test, a specific item is presented and the subject must attempt to ascertain whether this item has been presented previously in the session or not. It has sometimes been assumed that use of such a test will eliminate search from the retrieval process, but this is not necessarily correct. Characteristics of the item presented will lead the subject to examine some restricted LTS region for relevant information. The more salient are these characteristics, the more restricted will be the region indicated, and the smaller will be the search needed to locate the desired information. In general, however, some search will be required. When a stimulus is presented in a recall test where the number of responses is large, a considerably more extensive search is required. This occurs because stimulus information alone is required for the recognition phase, but the response may be encoded in quite another LTS location than that indicated by any salient stimulus characteristics. In a continuous paired-associate task with recall tests, recognition is still an important process; for example, the subject may recognize that a stimulus presented for test is new and has not been previously presented; upon such a recognition, the search will cease. When the task is such that the subject may either refrain from responding or emit a response, then wrong responses actually emitted are called intrusions. Due to the recognition process, the intrusion rate for new items being tested may be considerably lower than that for previously presented items.

Ranking. The task may require the subject to rank a series of responses in the order of their perceived likelihood of being correct. When the retrieval scheme is such that the search ceases when the first likely response is recovered, then the response ranked first will often be correct. However, responses ranked after the first will be correct only to the degree expected by pure guessing. If on the other hand, enough time is available for several likely responses to be recovered and considered, then responses ranked after the first will be correct at an above chance level. The degree to which the rankings after the first will be above chance will depend upon the decision process used to choose between likely responses, and also the coding schemes used.

Second-Guessing. Second-guessing refers to a procedure in which the subject is told whether his first response is wrong; if it is wrong he is then allowed to make an additional response, called the second-guess. First consider the case where a search procedure is used that would not result in an above chance ranking effect, i.e., the first likely response recovered in the search is output. When informed of an incorrect response, the subject will initiate another search of LTS. Performance on the second-guess will be partly determined by the degree of dependence of the second search upon the original search. If the second search is completely dependent, both in terms of the items making up the examination subset and also the order of search, then a correct second-guess can be made only in those instances where the wrong first response was an intrusion emitted before the c-code was examined in the original search. In these instances, the second search may continue beyond the point of the intrusion and thereafter result in a correct recovery. On the other hand, if the searches are completely independent, then correct recoveries can be made during the second search in cases where the c-code was present in LTS but not in the examination subset during the original search. In this event, the c-code might be in the examination subset during the second search. These considerations are complicated slightly if the original search was of the type which recovers several likely response alternatives, ranks them, and outputs the most likely. In this case, it is possible for the subject to forego a second search entirely and simply give the response ranked second most likely during the original search. If a second search is nevertheless engaged in, then the final response given must be the result of a decision process involving all the likely response alternatives recovered during both searches.

Regardless of the form of the second-guess search, there is no guarantee that the parameters of this search will be the same as on the original search. In particular, it would be natural for the subject to lower his criterion for output of recovered responses, since the original error indicates that the state of knowledge regarding the correct answer may be quite weak.

Interference Phenomena. Interference refers to a paradigm in which the first response paired with a stimulus (R1) is changed to a different response (R2); a subsequent test for R1 is called a retroactive interference condition, while a subsequent test for R2 is called a proactive interference condition. Although considerable work on interference phenomena has taken place within designs employing repeated presentations of whole lists of paired-associates, it is currently uncertain what form these phenomena will take in a continuous task. This entire question will be discussed more fully in subsequent chapters of this report. For the present we should merely like to point out that the theory can predict either proactive or retroactive interference effects. That is, learning of the R1 response may hinder recall of the R2 response, or vice versa. The predictions will depend upon the precise form of the assumptions regarding order of search and the addition of information to codes currently stored in LTS. For example, if search order is strictly temporal and proceeds starting with the most recent item, and if the original response code is older than the new response code, then no proactive effect will be expected. This prediction results from the following argument. In those cases where both the old and new codes for a stimulus are simultaneously in the examination subset, the new response code will always be examined prior to the older response code. Hence the probability correct will not be affected by the presence or absence of the older code.* On the other hand, a strong retroactive

*This is not quite true, but approximately so. Recovering the R1 response and emitting it will insure that an error is made. On the other hand, a different type of intrusion, or a pure guess, will be correct at the chance level. Thus the above argument is true when the chance level is zero, and is almost true when the chance level is quite low.

effect will be expected in this case, at least if the search terminates at the R2 code an appreciable proportion of the time.

To the degree that the strictly temporal search order assumption is relaxed, a proactive effect will be expected. However, if information is added to the R1 code that the response has been changed, then the search will bypass that code and continue; thus the proactive effect will be dependent on the information added to the R1 code when the response is changed. These same factors apply to retroactive interference. This discussion should make it clear that the theory has a good deal of freedom with regard to interference predictions. Experiment II in the next chapter examines proactive interference, and further discussion is reserved until that point.

<u>Latencies</u>. The recovery of a response from STS is assumed to be associated with a very short latency. The latency associated with a response recovery from LTS is assumed to be monotonically related to the number of codes examined before the response is given, the more codes examined, the slower the response. For the present discussion, components of response time associated with the decision processes involved in retrieval will be ignored. This rather simple conception of latencies leads to a large number of predictions. The latency of pure guesses should be quite long, since guesses occur only at the conclusion of an unsuccessful LTS search. The latency of intrusions will depend upon the order of search, but will probably be somewhat larger than correct response latencies. The latency of a correct response is expected to increase as the length of the period since the previous presentation increases, since a greater number of codes will tend to be

examined prior to the c-code as this period increases. The correct response latency will be expected to decrease as the number of reinforcements increases, since the c-code will tend to be stronger, and codes of greater strength will tend to be examined earlier in the search. This list of predictions may be extended in a natural fashion to change-of-response conditions, and to second-guess conditions, but further discussion will be reserved until the latency data of Experiment II is examined.

CHAPTER II

THE EXPERIMENTS: DESIGN, PROCEDURE, AND RESULTS

The two experiments of the present study were designed to investigate various facets of search and retrieval from long-term memory, and to provide a source of quantitative data against which a specific version of the theory outlined in Chapter I could be tested. Although both experiments utilized a continuous paired-associate design, the differences between them were considerable and their procedures will be described separately. The experiments are referred to as continuous because a particular item may have had its first presentation on any trial of the experiment, appeared a few times at varying intervals, and then been discarded. Each trial of the experiments consisted of a test phase followed by a study phase. During the test phase a stimulus was presented alone and the subject was then tested in some detail concerning his knowledge of the correct response. During the study phase, the stimulus just tested was presented with a response to be remembered. In what follows, we use the term lag to refer to the number of trials intervening between two successive presentations of a particular stimulus.

Experiment I

Design Justification. Experiment I was designed with several objectives in mind. A primary aim was the independent establishment of the imperfect-search characteristics of memory retrieval in the paired-associate situation. In order to accomplish this, a design was utilized which would separate two components of "second-guessing" performance: the partial-information component and the imperfect-search component.

A number of paired-associate experiments have shown that performance on a second response (following information that a first response was incorrect) may be well above chance level (Bower, 1967; Binford and Gettys, 1965); other experiments have shown that ranking of responses in their order of being correct can result in rankings beyond the first choice which are also above the chance level (Bower, 1967). These findings can be explained by either of two models: in the first, retrieval from memory results in recovery of partial information about more than one response; in the second, retrieval results in recovery of information about only one response; but if it's an error, a second search of memory results in recovery of new information about some other response. These models are separated in Experiment I by utilizing both rankings and second-guesses on each test trial.

The second major objective of Experiment I was the examination of changes in retrieval of individual items from memory, in a steady-state situation. Forgetting, particularly, needs extensive examination in a continuous task, since almost all the research on long-term forgetting has utilized a list-structure design. In such a design performance changes are measured for whole lists, and then inferred for individual items, but this inference lacks validation. For this reason, list structure is eliminated in Experiment I by using a continuous task: new items are continually being introduced, and old items eliminated.

A third objective of Experiment I was the demonstration that a class of previously used models for paired-associate learning suffered from certain deficiencies, deficiencies not present in the theory of Chapter I (henceforth called LTS theory). The design of Experiment I is similar to those used by Bjork (1966) and Rumelhart (1967). Each of these workers used a model to describe their data which has been called the GFT. The GFT model is basically a three state Markov model with a long term absorbing state (L). The probability that an item will be in L increases as the number of presentations of the item increases. Once an item enters L, a correct response will always be given and the item cannot thereafter leave L. Thus the GFT implies that the probability correct following a given sequence of reinforcements cannot be lower than a certain minimum, regardless of the lag of the current test; the minimum is determined by the probability that the item is in the state L at the time of test, which is not affected by the previous lag. These predictions are quite at odds with LTS theory: as long as new items are continually being introduced, LTS theory predicts that the probability correct should decrease toward chance as the lag increases. It is not surprising that the Bjork data was handled well by the GFT, because the design used did not allow for the continual introduction of new items; rather the design basically utilized a list structure, so that all items late in the session had been presented many times before. In such a situation LTS theory predicts that all items will become permanently learned, much as if an absorbing state was present; the prediction is based on many factors, which are described in Shiffrin and Atkinson (1968). Thus either GFT or LTS theory will provide an adequate description of list-structured designs. The Rumelhart study, on the other hand, used a design in which new items are continually being introduced; nevertheless the GFT model fit the data quite adequately. We propose that the GFT model proved adequate only because the range of lags examined was quite restricted, never being larger than 32. It should be possible to demonstrate that the GFT model is inadequate if a large enough range of lags is examined. For example, if the probability correct at very long lags tends toward chance, then a model in which an appreciable number of items enter an absorbing state will not be appropriate. For these reasons, the range of lags examined in Experiment I is very large, ranging from 0 to about 225.

Design. A daily session for each subject consisted of a series of 440 trials, each made up of a test phase followed by a study phase. On each trial a stimulus, possibly one not presented previously, was chosen according to a prearranged schedule and presented for test. Following the test phase that same stimulus is presented with a correct response during the study phase. The sequence in which the stimuli are presented for test and study are the same for every subject and every session; Appendix 1 gives the actual sequence used. In the Appendix, the sequence of trials is given in terms of the stimulus number. For a given subject and session each stimulus number represents some randomly chosen stimulus (actually a consonant trigram). Thus the sequence of trials remained fixed, but the actual stimuli and responses were changed from session to session.

A particular stimulus could be presented for a maximum of eight trials (eight reinforcements), at varying lags. Table II-1 gives the sequence of lags associated with each "item-type," where a stimulus of item-type *i* is presented at successive lags according to the *i*th row of the table. The first column in Table II-1 gives the item-type. The next seven columns give the successive lags at which items of each type

TABLE II - 1

SEQUENCE OF LAGS FOR ITEM-TYPES
OF EXPERIMENT I

Item-type	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Number of Sequences
1	1	1	1	1	1	16	100	6
2	1	1	1	1	1	100	100	2
3	6	6	6	6	6	16	100	6
4	6	6	6	6	6	100	100	3
5	10	10	10	10	10	16	100	7
6	10	10	10	10	10	100	100	4
7	25	25	25	25	25	25	25	7
8	50	50	50	50	50	50	50	4
9	100	100	100	100				8
10	0	100	100	100	100			4
11	1	100	100	100	100			4
12	10	100	100	100	100			5
13	∼225							6

are presented. The final column gives the number of stimuli of each item-type that are presented during each experimental session. As indicated in the table, the lags vary from 0 to about 225. The different stimuli of a given item-type are given first presentations which are spaced fairly evenly throughout each experimental session; the exact presentation schedule is presented in Appendix 1.

Four responses are used in Experiment I. When a stimulus is presented for test the subject responds by ranking the four responses in the order of their likelihood of being correct, using a random ranking if he does not know the correct answer. If the response ranked first is incorrect, then the subject is informed of this fact and he proceeds to rerank the three remaining alternatives, not necessarily in the same order as on the first ranking, and again guessing if the answer is not known. In order to make subsequent discussions clear, we adopt the following terminology. The subject's first four responses on a test trial are referred to as the "ranking." The second group of three responses (when given by the subject) is referred to as the "reranking." There is a further breakdown depending on the order of response. Thus the first response given on the test trial is called the "first-ranking," the second is called the "second-ranking," etc. The first response of the reranking (when the subject engages in reranking) is termed the "first-reranking" and so forth. It should be noted that the ranking responses in this experiment are akin to the responses given in the typical ranking experiment in the literature. Similarly, the first-ranking and first-reranking responses in this experiment are akin to the responses given in the typical second-guessing experiment.

Subjects. The subjects were ten students from Stanford University who received $2.00 an hour for their services. Each subject participated in a minimum of 8 and a maximum of 11 experimental sessions. The sessions were conducted on weekday evenings and took approximately 1-1/4 hours each. The subjects were procured without regard for sex through the student employment service.

Apparatus. The experiment was conducted in the Computer-Based Learning Laboratory at Stanford University. The control functions were performed by computer programs running in a modified PDP-1 computer manufactured by the Digital Equipment Corporation, and under control of a time-sharing system. The subject was seated at a cathode-ray-tube display terminal; there were five terminals each located in a separate 7 × 8-ft. sound-shielded, airconditioned room. Stimuli and other information were displayed on the face of the cathode ray tube (CRT); responses were made on an electric typewriter keyboard located immediately below the lower edge of the CRT.

Stimuli and Responses. The stimuli were 990 consonant trigrams (CCC's) made up of all possible 3 letter permutations of the following consonants: B,D,F,G,J,K,P,Q,X,W, and Z. Thus a typical stimulus was JXQ. Ninety stimuli were randomly selected for use during each session, with the restriction that any stimulus used in a session could not be used in any succeeding session for that subject. Thus a subject could not take part in more than 11 sessions.

Four responses were used: the numbers 1,2,3, and 4. Thus the guessing probability of a correct first-ranking was 1/4 and the guessing probability of a correct first-reranking was 1/3.

Instructions

When a subject arrived for the first session he was given a sheet of instructions to read, as follows:

> "This is an experiment to test your memory. You will be sitting in a soundproof booth facing a T.V. screen with a typewriter keyboard below it. Each day take the same booth as the previous day. To start the session, type the semicolon (;). The experiment will then begin.
>
> You will be required to remember the response members of a number of paired-associates, each consisting of a nonsense-syllable paired with a number as a response. The responses will always be either 1,2,3, or 4. Each paired-associate will be presented a number of times during a session and you should try to learn it. Each trial will consist of a test followed by a study. On a test, the word "test" appears on the top of the screen, and then below it appears a nonsense-syllable. Below the syllable will appear the term "rank answers." You will try to remember the response paired with the syllable presented for test. To respond, type the number you think most likely to be the correct response; then type the second most likely number; then the third most likely, then the least likely. That is, you will rank the responses 1-4 in order of their likelihood of being correct. As you type these 4 responses, they will appear on the screen, your first choice being on the left. If you are satisfied with your answers, then type a carriage-return (CR). If not satisfied at any point, and you wish to change your ranking, type E and the screen will clear and you may type in a new ranking. If you make a typing mistake, the screen will clear your responses at once: in this case, type them in again.
>
> When you rank the responses and type a carriage-return, the computer will check to see whether your first ranked response was correct. If it was correct, you will go on to a study trial on the syllable you were just tested on. If your first rank was incorrect, then you will get one more chance: the words "wrong. rerank answers" will appear on the screen. You will then rerank the three remaining answers in the order of their likelihood of being correct. That is, the first number typed is the first choice, etc. These "reranks" do not have to correspond to the first rerankings. If your first ranking was incorrect, search your memory again, and then make your best possible choices. As you type in your reranks they will appear on the screen. If you are satisfied with your three choices, then type a carriage return and the test trial will be terminated. The syllable you were tested on will then be presented with the

correct response for 2 seconds of study. Then after a short delay, the next test trial will begin.

Take the time you need to respond during test trials, but attempt to respond as quickly as possible without lowering your performance.

Your task is to learn and remember as many pairings as possible and to demonstrate this learning during the test phases of the trials. Feel free to use any codes or mnemonics you can devise in order to learn the pairs.

The way the experiment is being run, a syllable will first be presented for test on a trial, and then for study. Thus, especially at the start of a session, you will be tested on syllables whose response you have not yet seen. In this case, simply rank the responses randomly, i.e., guess. When guessing, do not always type in the answers in the same way - try to guess randomly. Furthermore, even if you feel you know the answer, do not always type in the remaining answers in the same order. Try to type these answers randomly also. Any questions? The experimenter will now review these instructions with you verbally."

The experimenter reviewed the instructions with the subjects and then introduced them to the computer and its operation. The entire first session was used to familiarize the subject with the apparatus and instructions, and to give him practice at the task.

Procedure

Each session consisted of a sequence of 439 trials, a trial being defined as a test followed by a study. Each trial involved a fixed series of events. (1) The word TEST appeared on the upper face of the CRT. Beneath the word TEST a specifically determined member of the stimulus set appeared, the stimulus member indicated by the presentation schedule given in Appendix 1. Below the stimulus appeared the words RANK ANSWERS. The subject then ranked the four responses by typing them in order on the keyboard, the most probably correct answer first, and so forth. The answers appeared on the CRT as they were typed. After ranking the four responses the subject typed a carriage-return and the rankings were evaluated by the computer. Previous to this point, the subject could begin his rankings anew by typing E. If the first-ranked response was wrong (even for stimuli never seen before) then the words WRONG. RERANK ANSWERS appeared on the CRT below the original rankings, which remained on the CRT. The subject then reranked the three remaining answers under the same conditions that pertained to the original rankings. The rankings and rerankings were self-paced, but instructions were used which insured that the subject took about 6-7 seconds for responding, on the average. (2) The CRT was cleared and a blank screen appeared for 1/4 second. (3) The word STUDY appeared at the top of the CRT. Beneath the word STUDY appeared the stimulus just tested along with the correct response. The correct pairing remained on the CRT for 2 seconds. (4) The CRT was blanked for 3/4 seconds. Then the next trial began. As indicated above, a complete trial took about 10 seconds or less and thus a session lasted about 1 hour and 15 minutes.

At the start of each session, the computer randomly assigned each subject 90 stimuli he had not seen in previous sessions. Each stimulus was then randomly assigned one of the four responses as the correct pairing to be used throughout that session. It should be noted again that the sequence of trials was the same for every subject-session, but the actual stimuli and responses differed. The first 12 trials of each session consisted of 10 filler items; these appeared seldom thereafter. From the 13th trial on, almost all trials were instances of one or another of the 13 item-types listed in Table II-1. These item-types were spaced roughly uniformly through the remaining 427 trials.

Altogether 83 subject-sessions of data were collected following the initial practice session. Because of computer stoppage or other extraneous reasons, only 58 sessions were entirely completed, but the remaining sessions were at worst within 10 or 20 trials of completion. The data collected on each trial consisted of the stimulus tested and its correct response, and the rankings and rerankings given by the subject. Latencies were not recorded. At the conclusion of the experiment, each subject filled out a written questionnaire.

Results of Experiment I

Table II-2 presents the summary results for each of the 10 subjects in the experiment. Tabled is the probability of a correct first-ranking lumped over all trials and sessions. The results are listed in order of increasing probability correct. It is evident that there are appreciable subject differences in overall ability in this task. Nevertheless, in order to gain precision of estimates, the remaining data are presented in a form lumped over all subjects. This should not overly distort the observed effects, since a consideration of the data to follow, where the number of observations permitted a subject by subject breakdown, consistently showed that the same qualitative effects hold for individuals as for the average data. Possible selection effects introduced by averaging will be discussed in Chapter III.

Table II-3 gives the probability of a correct first-ranking over successive days of the experiment (the practice session is not included). It is clear that no trend over days is present in the table. Apparently, proactive interference from session to session was minimal. The data to follow will be lumped over all sessions, excluding the practice session.

TABLE II - 2

MEAN PROBABILITY CORRECT FOR SUBJECTS OF EXPERIMENT I

Subject Number	10	7	4	2	9	1	6	3	8	5
Probability of Correct First-ranking	.45	.47	.51	.52	.54	.56	.59	.68	.69	.77

TABLE II - 3

MEAN PROBABILITY CORRECT FOR SUCCESSIVE DAYS OF EXPERIMENT I

Day Number	1	2	3	4	5	6	7	8	9	10
Probability of Correct First-ranking	.58	.55	.58	.62	.61	.55	.56	.63	.54	.60

TABLE II - 4

PROBABILITY CORRECT AS A FUNCTION OF THE AVERAGE STATE OF KNOWLEDGE CONCERNING THE ITEMS MAKING UP THE PRECEDING LAG

	Low K Group Pr(C)	High K Group Pr(C)
Lag 1, Reinforcement 1:	.70	.75
Lag 6, R1:	.54	.61
Lag 10, R1:	.54	.57
Lag 25, R1:	.43	.52
Lag 50, R1:	.35	.43
Lag 100, R1:	.31	.39
Lag 1, Reinforcement 2:	.85	.88
Lag 6, R2:	.70	.76
Lag 10, R2:	.67	.68
Lag 25, R2:	.54	.57
Lag 50, R2:	.37	.43
Lag 100, R2:	.47	.46

Ranking Performance vs. Second-Guessing Performance. As stated earlier, a number of previous experiments have found that responses ranked after the first choice are correct at an above chance level. A hypothesis which can explain this finding holds that the subject sometimes retrieves from memory information which indicates the possible correctness of two or more responses. The subject examines this ambiguous information and then produces his rankings as the result of some type of decision process. Thus the correct response is sometimes ranked second rather than first, and the above finding is observed. Other experiments in the literature demonstrate that second-guesses, after the subject is told the first-guess is wrong, can result in performance well above chance levels. The hypothesis proposed above can also be utilized to explain this result: the subject engages in implicit ranking on the first guess and gives the response implicitly ranked first; if he makes an error, he then outputs the response he had previously ranked second. It is possible, however, that a substantial portion of the second-guessing effect may be explained by an alternative hypothesis: the subject makes his first guess on the basis of information available at the time; upon knowledge of an error he then engages in an additional search of memory. This second search sometimes results in retrieval of information not previously available to the subject, information which may then be used to respond correctly. This hypothesis is quite different from the first in its emphasis of the essentially probabilistic nature of the memory retrieval process.

The present experiment provides a means of separating these hypotheses. The essential statistic examines those instances where the response ranked first is wrong, but where the response reranked first is not the response ranked second. For these instances, a probability of correct first-reranking above the level expected by chance guessing implies that the second hypothesis is operative in the experiment. A convenient way to begin an analysis of the data is presented in Figure II-1. On the abscissa is the probability of a correct first-ranking divided into successive intervals which are marked on the graph. These intervals start at .30 since no item-type had a probability of correct first-ranking on any test after the first reinforcement which was below .30. For each interval we consider all trials in the sequence of 440 on which the probability of correct first-ranking lies in the interval. For these trials we graph (1) the probability that the second-ranked answer is correct and (2) the probability that the first reranked answer is correct. Both probabilities are plotted conditional upon a first-ranking error; thus the chance level for both probabilities is .33. In what follows we will refer to the first-reranking as second-guessing.

From the upper curve in Figure II-1 it is evident that a substantial amount of correct second-guessing has taken place. On the other hand, the lower curve indicates that virtually no initial ranking effect took place. The probability of correct second-ranking is barely above the chance level, the mean for all trials except those on which new stimuli are presented being .352. This probability is significantly above chance since it is based upon approximately 7000 observations, but it is obvious that the magnitude of the ranking effect is small compared with that of second-guessing. This result suggests that the second hypothesis presented above is appropriate for this experiment. That is, since the

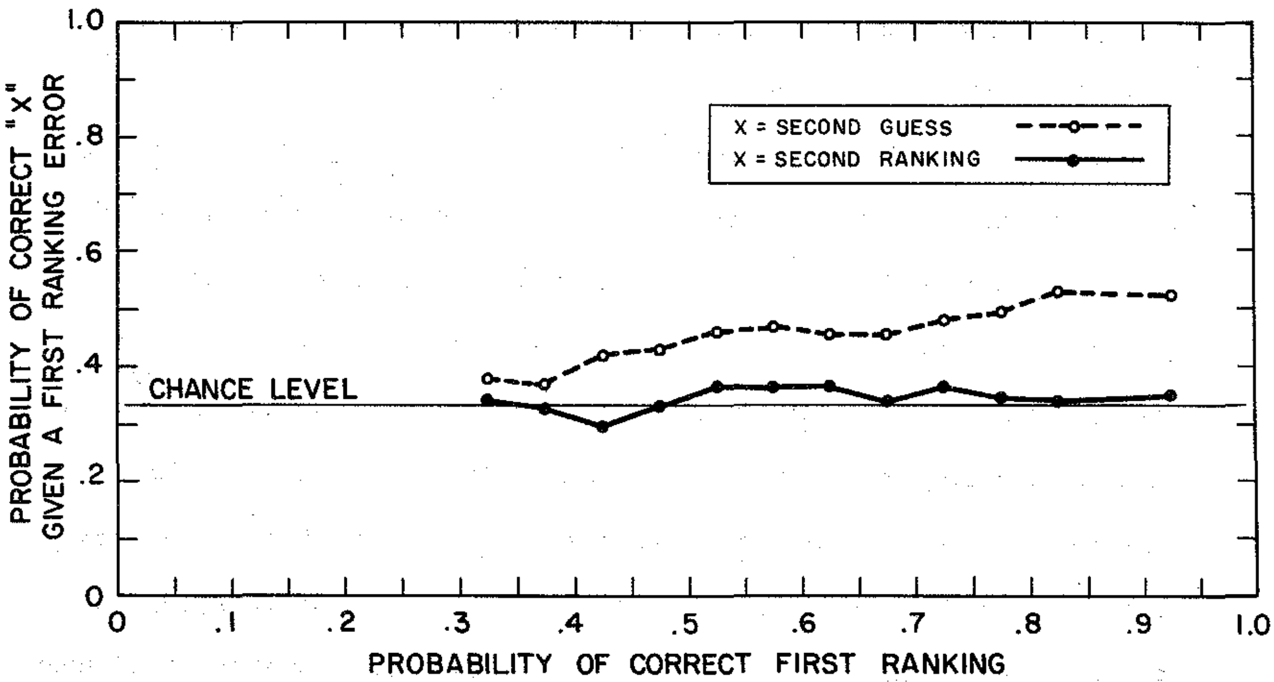


Figure II-1. Conditional Probabilities of Second-Guessing and Second-Ranking.

ranking effect was near chance, the majority of correct second-guesses were responses that were not ranked-second during initial ranking. Thus the subjects were utilizing information during second-guessing that was not utilized during initial ranking. A straightforward interpretation holds that after the error feedback a search was initiated which occasionally resulted in the correct response being found.*

It is most likely that the failure to find a large second-ranking effect was due to the instructions regarding response rate. Although responding was self-paced, the subjects were instructed to respond quickly enough to finish in an hour and a quarter, and had to respond rapidly as a result. Under these conditions, the subjects would be led to adopt a memory-search strategy which would output the first likely response alternative located in the search. If responding rates were lower, the subjects could adopt a strategy in which the memory-search continued until all likely alternatives could be recovered and evaluated. In this case a second-ranking effect would very likely result.

The failure to find a substantial ranking effect might lead us to expect that the reranking effect would also be minimal. This was indeed the case; rerankings after the first were correct with a conditional

---

*It conceivably could be argued that the subjects "knew" during their initial rankings the information they later used to second-guess, but nevertheless ignored it while making the rankings. This seems doubtful, especially if one takes the subjects own written comments into account: in several instances the subjects stated the second hypothesis almost verbatim on their final questionnaire. In any event, if the need arose, it is not difficult to formulate experiments to clear up this possible ambiguity, perhaps by giving positive payoffs for correct second-rankings.

probability of .498, almost exactly the level expected by chance. As a result, the remaining data analysis is considerably simplified. Only the first-ranking and first-reranking results will be considered and will be referred to as first-guessing and second-guessing respectively.

Learning and Forgetting. The title of this section should not be misconstrued: by learning and forgetting is meant only increases and decreases in retrieval. As indicated in Chapter I, our theoretical approach does not allow for the disappearance of stored information from memory, and the use of the term forgetting should not be taken to mean such.

In the following data the number of observations at each point may be found approximately by reference to Table II-1: for each item-type, multiply the entry in the column headed "NUMBER OF SEQUENCES" by 80, the approximate number of subject sessions. Figure II-2 presents the lag curves for first reinforcement items. The top panel presents the probability of a correct first-guess following an item's first reinforcement at a lag marked on the abscissa. The lower panel presents the probability of a correct second-guess conditionalized upon an error on the first guess. The observed data are plotted as open circles connected by dashed lines. The predictions are based on the model presented in Chapter III and may be ignored for the present. As might be expected in a continuous task, the lag curve decreases toward chance as the number of intervening items increases, albeit quite slowly. The chance level in the top panel is .25, and in the bottom panel is .33. The second-guessing curve is of interest because of its relatively small variance over the range of lags shown, and because of its maximum at about a lag of 10 or thereabout.

Discussion of the second-guessing data is reserved for the next chapter. The first-guess curve is most important because it demonstrates that the probability of a correct response tends toward chance as the lag increases. Thus the GFT model, or any model with a long-term absorbing state, will not provide an appropriate description of the data.

Figures II-3 and II-4 present the "learning" curves for each of the item-types in the experiment. The probability of a correct first-guess is plotted as a function of the number of presentations, for each item-type. The lag between successive presentations is listed in each graph as a small number placed between successive points on the predicted curve. In the two figures, the chance level is .25. Figures II-5 and II-6 present the same curves for second-guessing. These figures present the probability of a correct second-guess conditionalized upon a first-guess error; thus the chance level is .33. In each of these last four figures, all curves begin at the chance level, since on the first presentation the subject has not previously seen the item being tested. In Figure II-5 several observed points have been deleted from the Type 1 and Type 2 graphs. The number of observations at these points was below 30 (because the probability of a correct first-guess was so high).

Several characteristics of these data should be noted at this time. First, as found by previous workers (Greeno, 1964; Peterson, Hillner, and Saltzman, 1962; Rumelhart, 1967), a distributed practice effect occurred. Consider item-types 10, 11, and 12 in Figure II-4. As the first lag was varied from 0 to 1 to 10, the probability correct after a subsequent lag of 100 rose from .37 to .44 to .49; i.e., the longer

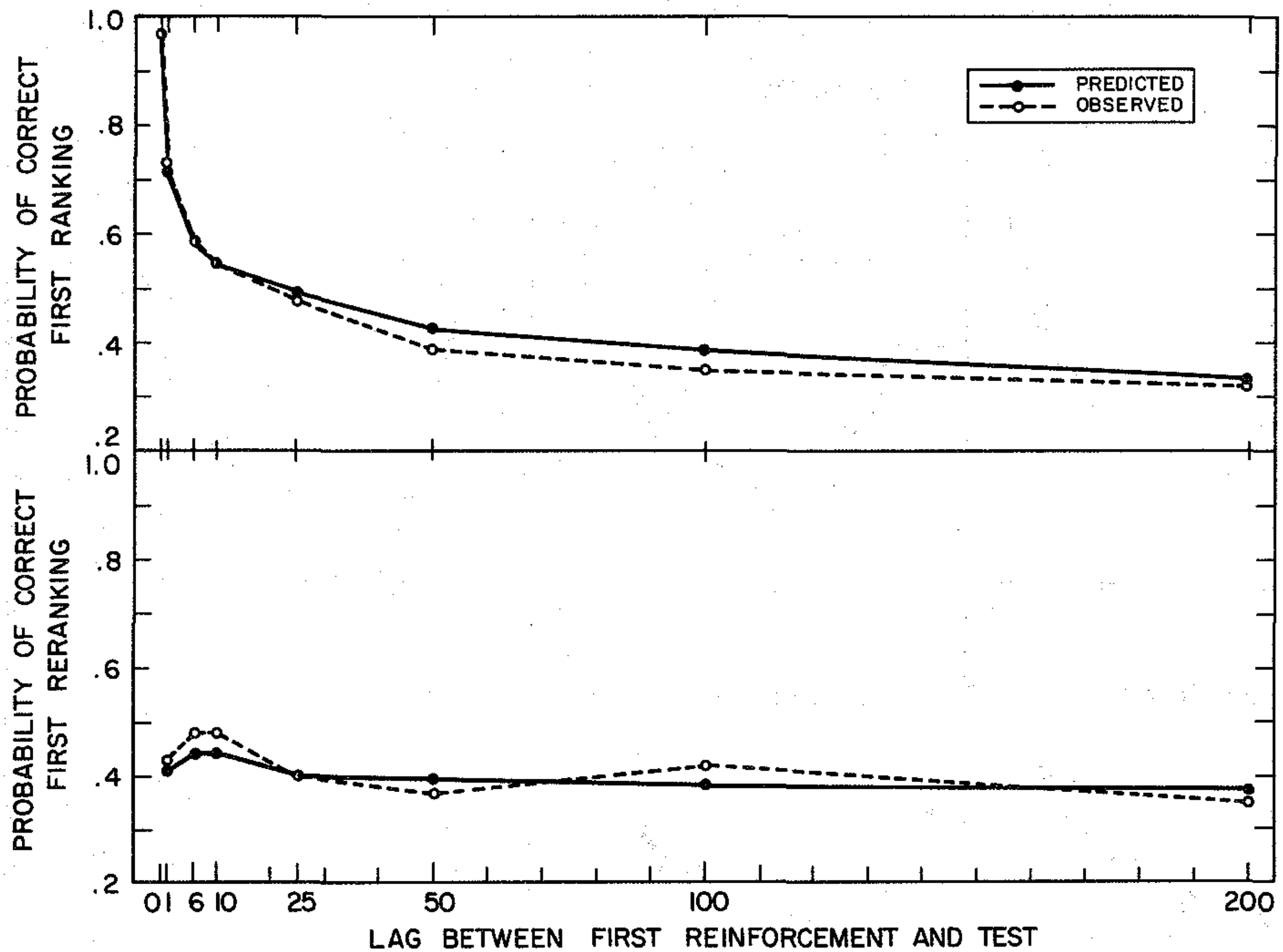


Figure II-2. Probability Correct as a Function of Lag, for Rankings and Rerankings.

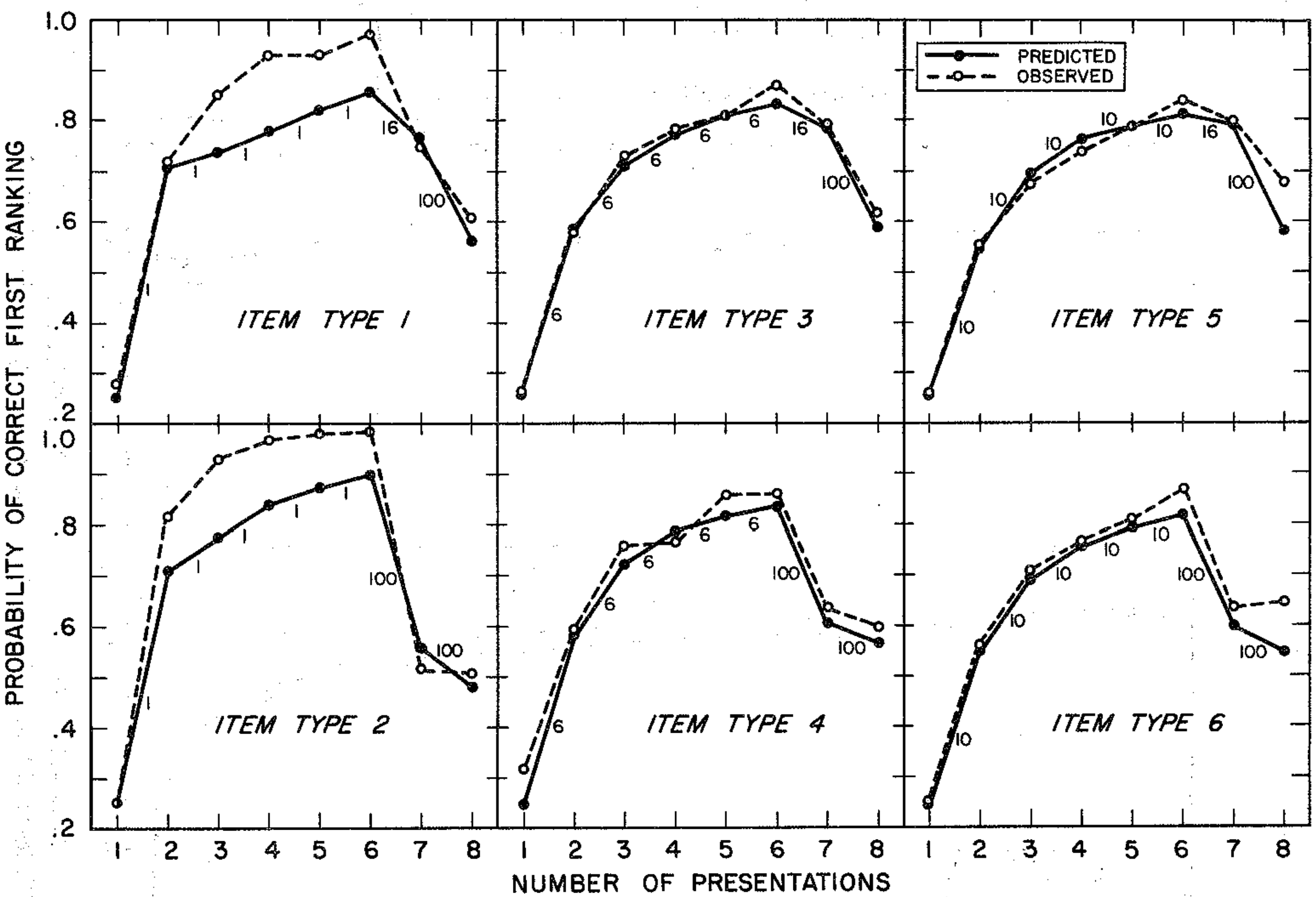


Figure II-3. Probability of Correct First Ranking as a Function of Number of Presentations, for Item-Types 1-6.

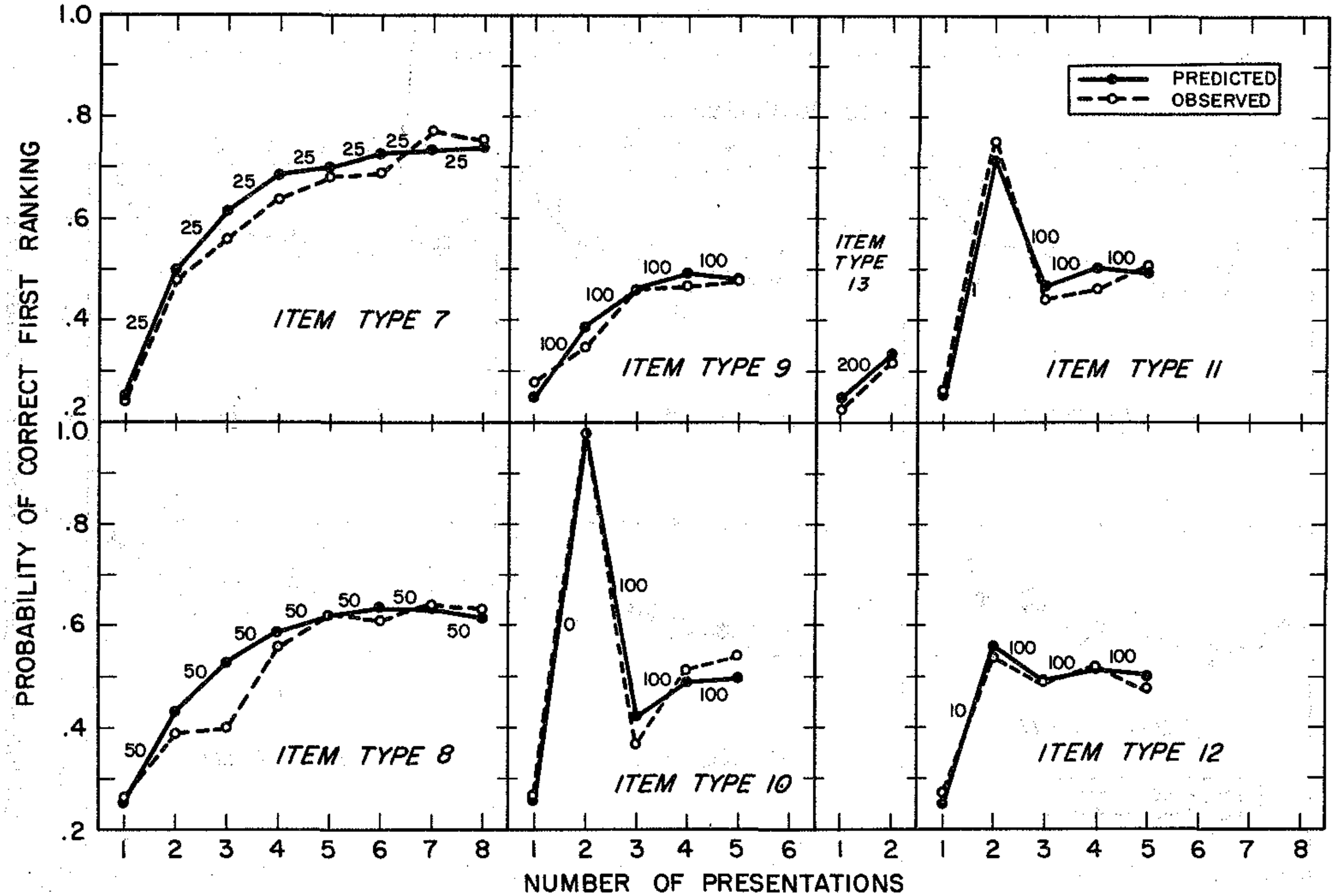


Figure II-4. Probability of Correct First Ranking as a Function of Number of Presentations, for Item-Types 7-13.

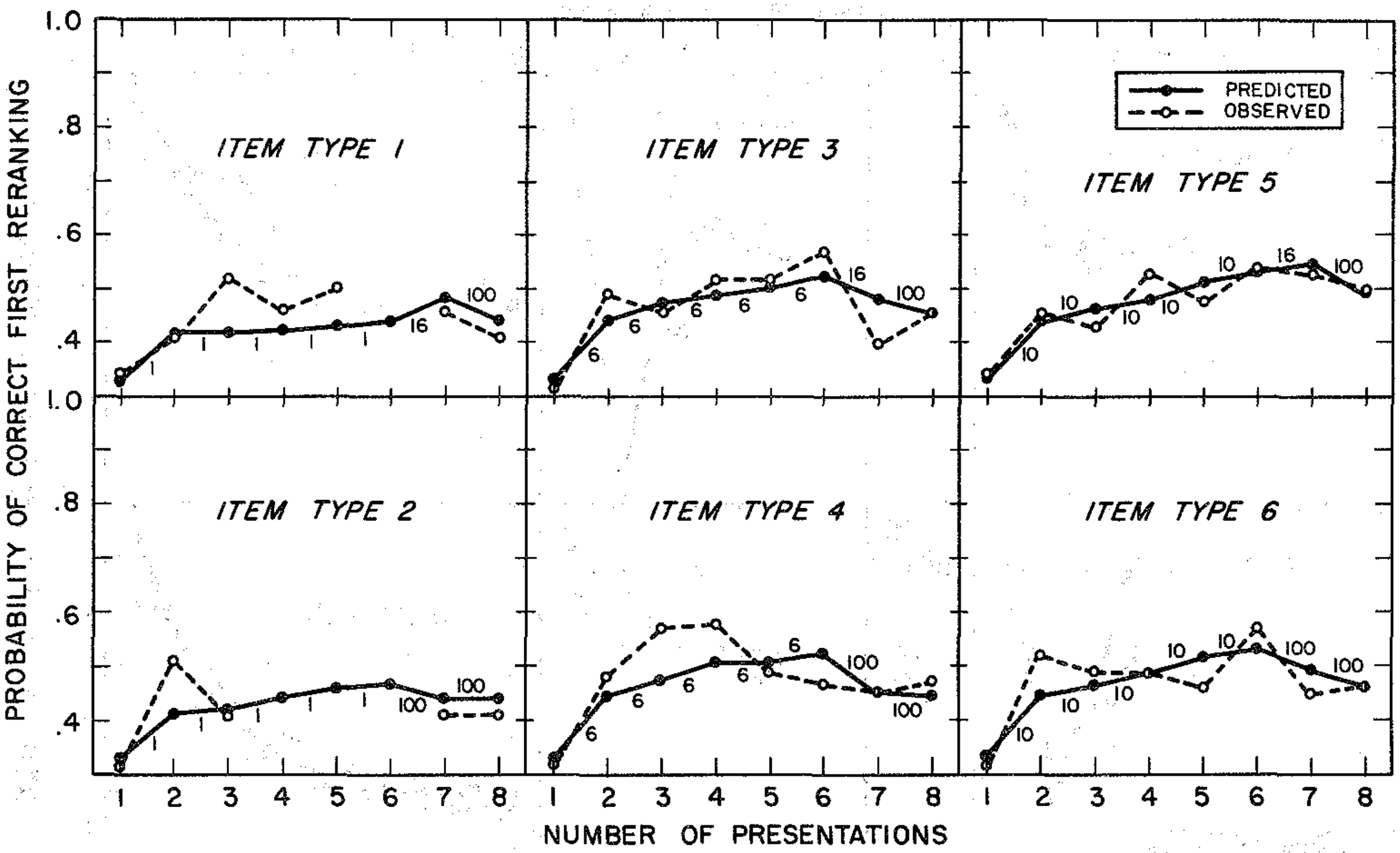


Figure II-5. Probability of Correct First-Reranking as a Function of Number of Presentations, for Item-Types 1-6.

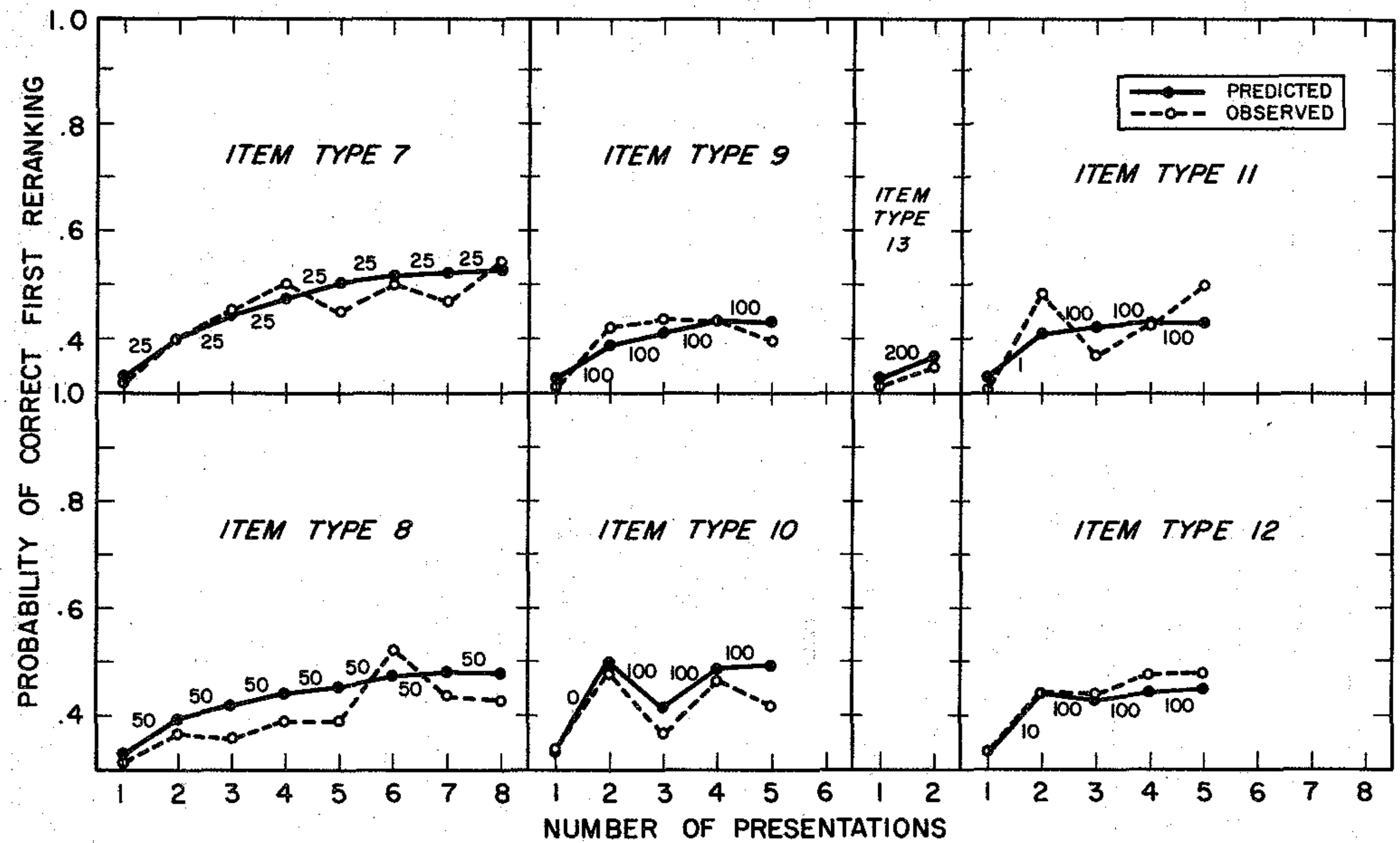


Figure II-6. Probability of Correct First-Reranking as a Function of Number of Presentations, for Item-Types 7-13.

the initial lag the better is performance after a long subsequent lag. A similar effect is seen in the graphs of item types 2, 4, and 6 in Figure II-3. Following five initial lags of either 1, 6, or 10, performance on two subsequent tests at lags of 100 rose from .52 to .62 to .65; i.e., performance is better at long lags the more spaced is the series of initial reinforcements.

It should be noted that item-types 9, 10, 11, and 12 seem to exhibit something like steady state characteristics; i.e., if reinforcements are given at lags of 100, performance seems to stabilize near the .50 level.* Item types 7 and 8 also seem to be approaching an asymptotic level of probability correct well below 1.0 (.75 and .63 respectively). These results further demonstrate that any model with a long term absorbing state which items enter an appreciable portion of the time will not provide an adequate description of the data. If the probability correct for an item in the absorbing state is p, then all curves at long lags should be asymptoting at p. This is not the case for these data even if p is allowed to be less than 1.0.

The Effects of Intervening Items. The lag curves above show that forgetting increases as the lag increases. It should be questioned whether it is the number of intervening items per se which determines the amount of forgetting. The theoretical position outlined in Chapter I implies that forgetting should, among other things, be a function of

---

*This result might lead to speculation that item-types 1-6, if given additional reinforcements at lags of 100, would exhibit a decrease in performance down toward the .50 level (which would be a strange sort of "learning," indeed).

the amount of new information stored during the intervening period. Therefore, the amount of forgetting should vary as a function of how well-known are the intervening items, if we accept the view that less new information is stored concerning well-known items. A similar expectation would hold if the degree of inter-stimulus interference were a determinant of forgetting; the greater the number of unknown stimuli that intervened, the greater the forgetting.* There are a number of experiments which bear on these points. Thompson (1967) demonstrated that a strong short-term effect exists in a situation where the subject adopts rehearsal as a predominant strategy; that is, a short series of extremely overlearned items following an item caused no forgetting, whereas an equal length series of unknown items caused dramatic decrements in performance. This short-term memory rehearsal effect should be differentiated, however, from the long-term memory retrieval effect proposed above; we shall return to this point shortly. Calfee and Atkinson (1965) proposed a trial-dependent-forgetting model for list-structured P-A learning. In this model, the amount forgotten from a short-term state of learning between successive reinforcements was proposed to decrease as the trial number increased, since the intervening items became better and better known as the experiment proceeded. While they found the trial-dependent-forgetting model to fit the data

---

*In principle, the various sources of forgetting should be separable. For example, an experiment could be run in which items are compared which are tested at equal lags and have equal numbers of intervening new stimuli; the items would differ in that the interreinforcement lags of the intervening items would be low in one case and high in the other.

more closely than the alternatives, one cannot directly conclude that the finding applies to individual items; since a list design was used, the changes in forgetting could be the result of some sort of reorganization or integration of the entire list over trials.

Although Experiment I was not expressly designed to systematically vary the makeup of the intervening items at a given lag, a fair amount of chance variation occurred and it is possible to capitalize upon this fact. Every trial in the trial sequence was assigned a number "K" representing how well "known" was its stimulus-response pair as follows:

$$K = (\text{reinforcement number}) \times (20)/(\text{lag}+1). \qquad \text{Eq. II-1}$$

In this formula the reinforcement number and the lag refer to the stimulus tested on that trial. K is very highly correlated with the probability correct on each trial and therefore provides a reasonably valid measure. Next we compute for each item presented the average value of K during the preceding lag, and call this average $\overline{K}$. We can now compare the probability correct for each item with how well "known" were the items making up the preceding lag. Table II-4 presents the resultant data (on page 36) for items tested following their first and second reinforcement, at each of several lags. At each lag, all items are divided into two roughly equal groups, those with high $\overline{K}$ and those with low $\overline{K}$. Thus the items with lag 1 and reinforcement 1 are split into a high-group and a low-group, all items in the high-group having values of $\overline{K}$ greater than any items in the low-group. The mean probability correct is then computed for items in the high-group and for items in the low-group, and these means are listed in columns 2 and 3 of the table. Hence column two of the table gives the mean probability correct for items whose intervening items are relatively well-known.

There are a number of points to be made regarding Table II-4. First, there is a definite, highly significant effect in the expected direction: intervening items which are less well-known cause more forgetting.* Almost certainly the magnitude of the differences would have been even larger than those observed if variations in $\overline{\overline{K}}$ had been larger; however, differences in $\overline{\overline{K}}$ arose by chance rather than by design. Of particular interest is the result for lag 1. In this case there is only a single intervening item and $\overline{\overline{K}}$ varies considerably from item to item; in fact, the mean probability correct for the intervening item was .31 for the low-group and .77 for the high-group. Nevertheless, only a difference of .05 was found in the measure tabled. If a rehearsal-type short-term process was causing the result, as in the Thompson study cited earlier, then this difference should have been far larger than was observed, and far larger than other differences in the table.** There is another feature of the data which makes this same point. The rehearsal model

*There is no question of significance. The results for reinforcements greater than 2 show essentially the same results as for those shown in the table. A sign test on the directions of the differences gives $p < .01$ and more rigorous tests would lower this probability considerably.

**The justification for this statement ultimately rests on a theoretical analysis in which the buffer model is applied to the data. It is beyond the scope of this report to go into the details of the analysis, but a buffer model was applied to the data of Experiment I. The best fit of the model was not adequate as a description of the data, and one of the major failings of the model was the extreme overprediction of the effects of known items at lag 1. Rather than the .05 difference at lag 1 which was presented in Table II-4, the buffer model predicted a difference of about .30.

explanation of the effect of known items holds that known items fail to cause decreases in performance because they do not enter rehearsal; if the intervening items do not enter rehearsal, then the target item will tend to stay in rehearsal in STS for a longer period of time, even until the moment of test. In this model, the first few items after the target item are crucial in determining the magnitude of the effect. In order to check this point, the analysis leading to the statistic in Table II-4 was repeated, except that $\overline{K}$ was calculated without including the K values of the first two intervening items. Nevertheless, the resultant pattern of results (excluding lag 1, of course) was virtually identical to that in Table II-4. A sign test on the direction of differences again gave a $p < .01$ as a level of significance. We therefore conclude that the $\overline{K}$ effect is not crucially dependent upon the K value of the first few intervening items. It seems reasonable, then, that the effect originates in the LTS retrieval process, rather than in a rehearsal mechanism. The explanation we propose, in terms of the theory of Chapter I, holds that the "age" of any code is dependent upon the number of new codes that are subsequently stored in LTS. Since the probability correct depends upon the "age" of a code, the effect found in Table II-4 follows directly.

Summary. There are several main results of Experiment I. First, the multiple-search nature of retrieval was established by a comparison of ranking and second-guessing effects on the same test trial. Second, performance was observed to tend toward chance as the lag increased; this and related findings demonstrated the inappropriateness of a model for this task which postulates a long-term memory absorbing state. Third, the forgetting of an item at a given lag, long or short, was observed to depend upon the degree to which the intervening items were known. Discussion of other results, and of the quantitative aspects of the data, will be reserved for Chapter III.

Experiment II

Experiment II was designed with the objective of providing a stringent test of the model used to predict the results of Experiment I. An integral feature of this model (to be discussed in detail in Chapter III) was the prediction of intrusion errors; i.e., incorrect retrievals from memory. In Experiment I responses were required on every trial, so that intrusions and pure guesses were not separable at the observable level. In Experiment II the response set size was increased and the subject was instructed to respond only when he felt he knew the answer. In this manner, intrusions may be observed directly. The ranking technique was not used - only a single first-guess was allowed - but second guesses were allowed following errors. A second objective of Experiment II was the collection of "interference" data which would allow for the natural expansion of the earlier model. Thus individual stimuli in the present experiment sometimes had their response assignment changed. Formally, a design was adopted which was the counterpart in a continuous paired-associate experiment of the standard proactive interference paradigm.

The design and procedure of Experiment II is in certain respects identical to that of Experiment I. Except where noted, the procedure was the same as in the previous experiment.

Design Justification. Each session involved an identical sequence of 400 trials; each trial consisting of a test phase followed by a study phase. The trial sequence, presented in Appendix 2, will be discussed shortly. As in Experiment I, the individual stimuli and responses were changed from one session to the next - only the sequence remained fixed. An individual stimulus could be presented on as many as 8 trials during the sequence, at varying lags. On some trials the response assignment of a stimulus was changed; on these trials the subject was notified following the test phase that the answer would be changing. The pair presented during the study phase would then contain the new response.

The item-types in the present experiment were constructed so as to provide a full test of proactive-interference phenomena with appropriate controls. Quite apart from considerations relating to the theory proposed in this paper, it is maintained that interference phenomena need reexamination in the context of continuous paradigms. Forgetting phenomena have been examined extensively for many years with the use of list-structured experiments: lists of paired-associates are successively learned, each list utilizing the same stimuli, but with response assignments shifted (i.e., the A-B, A-C design). The results of these experiments have been fairly successfully explained by some version of two-factor interference theory (Postman, 1961; Melton, 1963; Underwood, 1957; Keppel, 1968; etc.). The experimental effects are found to take place over whole lists, but it is often assumed that equivalent changes occur in individual stimulus-response assignments, the assumption based upon a seemingly natural inference. Thus, if, in an A-B, A-C design, it is found that increased training on the A-C list causes increased forgetting of the A-B list, it is then inferred that increased learning of a particular stimulus-response pair will result in increased forgetting of a previous pairing of that same stimulus with a different response. Recent research, however, has raised doubt about this inference (DaPolito, 1966; Greeno, 1967). Following A-B, A-C learning subjects were asked to give for each stimulus both responses previously paired with it; regardless of the presence of retroactive interference effects in the lists as a whole, it was found that the probability of a correct first-list response times the probability of a correct second-list response was equal to the combined probability of giving both responses correctly. This is a result to be expected if there were no individual item response interactions; i.e., if for a particular item the level of learning of the first list response does not affect the level of learning of the second list response, and vice versa. This implies that the usual inference from lists to items may not be valid, and theories of item interference should therefore be based on appropriate experiments which do not utilize a simple list structure.

Atkinson, Brelsford, and Shiffrin (1967) reported a continuous P-A experiment in which some indications of proactive interference were found for individual items. This finding was only incidental in that experiment, however, and could possibly have been caused by selection effects. Estes (1964) reported experiments in which proactive interference effects were sought for individual items buried in a list structure, but the results indicated no proactive effect. Peterson, Hillner, Saltzman, and Land (1963) reported a continuous task in which there were indications of retroactive interference. These experiments seem to delimit the current state of knowledge concerning individual item-interference: very little is currently established.

The present experiment was therefore designed to examine in depth the status of proactive item-interference. The item-types utilized for this purpose are listed in Table II-5. A stimulus is presented with its first response (R1) either 2 or 4 times for study. The response is then changed and 3 study trials are presented with the new response (R2), all at lag 10. The lags of the initial presentations are either (0-10) or (10-10) if there are two initial presentations, or (0-10-0-10) or (10-10-10-10) if there are four initial presentations. On the trial where the answer first changes, the test asks for the R1 response, the subject is then told the answer is changing, and the new pairing is presented. We denote these item-types by the initial sequence of lags. The column on the right margin of the table gives the number of instances of each item-type in the sequence of 400 trials.

A comparison of the first and second tests following the change of response, with the first and second tests before the change of response, should indicate any overall proactive effects. A comparison of the conditions in which the number of response 1 presentations varies (i.e., (10-10) vs. (10-10-10-10)) permits us to examine the probability of a correct R2 as a function of varying amounts of learning on R1. A comparison within the same number of initial presentations (i.e., (0-10) vs. (10-10)) should allow the same examination as above, but where the number of presentations is held constant (assuming that the 0 lags do not result in much learning). In this way it may be determined whether any proactive effect found is due to the amount learned about R1, or simply due to the number of presentations of R1.

The above item-types examine proactive interference only at lag 10. In order to study the effects of variations in lags, 16 other item-types were used. Each of these 16 item-types is given just three presentations; on the second presentation the response is changed. The lag between the first and second presentation is called lag 1; the lag between the second and third presentations is called lag 2. The item-types are listed in Table II-5a. Lag 1 takes on the values 0, 1, 4, 10; lag 2 takes on the values 1, 5, 10, 25. The entries in each cell of the 4 X 4 table are the number of occurrences of each item-type. These item-types will be denoted by their lag 1 and lag 2 separated by a comma; e.g. (4,25). Note that item-type (10-10) is different than item-type (10,10).

The subject is instructed to respond during each test with the response <u>most recently</u> paired with the stimulus presented. He is told to "forget" any old pairings once the response has changed. The subject does not have to respond if he does not know the answer. If he does respond and is wrong, he is told so and given an opportunity to respond again.

<u>Subjects</u>. The subjects were 14 students from Stanford University who received $2.00 per hour for their services. Each subject participated in a minimum of 8 and a maximum of 11 experimental sessions plus one initial practice session. The sessions were conducted on weekday evenings and took approximately 55 minutes each. The subjects were procured without regard for sex through the student employment service. The apparatus was identical to that for Experiment I.

<u>Stimuli and Responses</u>. The stimuli were 1600 common English words either 3, 4, or 5 letters in length selected in random fashion from

TABLE II - 5

ITEM-TYPES FOR EXPERIMENT II

Item-type	Response 1	Response 2	No. of Types
0-10	P1 -Lag- 0 P2 -Lag- 10	P3 -Lag- 10 P4 -Lag- 10 P5	7
10-10	P1 -Lag- 10 P2 -Lag- 10	P3 -Lag- 10 P4 -Lag- 10 P5	7
0-10-0-10	P1 -Lag- 0 P2 -Lag- 10 P3 -Lag- 0 P4 -Lag- 10	P5 -Lag- 10 P6 -Lag- 10 P7	8
10-10-10-10	P1 -Lag- 10 P2 -Lag- 10 P3 -Lag- 10 P4 -Lag- 10	P5 -Lag- 10 P6 -Lag- 10 P7	7

In the above table P followed by a number represents the presentation number of a stimulus of that item-type.

TABLE II - 5a

ITEM-TYPES FOR EXPERIMENT II

		Second Lag			
		1	5	10	25
First Lag	0	3	3	4	4
	1	3	3	3	3
	4	3	3	3	4
	10	3	3	4	3

In the above table the numbers in each cell are the numbers of instances of each item-type. Note that the first lag is previous to the changing of the response, and the second lag is subsequent to the changing of the response.

Thorndike (1921), with homonyms, personal pronouns, possessive adjectives, and the past tense of verbs eliminated. Ninety-five stimuli were randomly selected for use during each session, with the restriction that any stimulus used in a session could not be used in any succeeding session for that subject. Words were used as stimuli, rather than CCC's, in order to make the proactive interference comparisons meaningful. That is, the design does not use unique response pairings; hence the same response can be assigned to more than one stimulus. If two stimuli assigned the same response are not sufficiently different, it would be difficult to differentiate this case from the case where a single stimulus had a changed response assignment.

The responses were the 26 letters of the alphabet. At the start of each session all stimuli were assigned R1 and R2 responses randomly with the restriction that no word could be assigned its own initial letter as a response. Since no subject reported noticing this restriction, it may be assumed that the probability correct, if the subject decided to make a pure guess, would be 1/26.

Instructions. When a subject arrived for the first session he was given the following instructions to read:

> "This experiment will test your ability to remember responses to a series of common English words. The response will always be one of the letters of the alphabet. You must always try to remember the letter most recently paired with a particular word.
>
> The experiment will consist of a number of trials in succession and last about an hour (or less) each day. Each trial will begin when the word "test" will appear on the screen before you. Below the word "test" will appear an English word (which you may or may not have seen before on a previous trial.)

The task on this test trial is to give the response most recently paired with the word shown. If you have no idea what the answer is, then either type a "carriage return" (CR) or do not respond at all; if you have a guess, then type the letter you think is correct. Remember, the correct letter is the one most recently paired with a particular word.

If you type a letter and are wrong, the computer will tell you so and give you a second chance. Again, type a carriage return or do not respond if you have no idea as to the answer, and type the letter if you have a guess.

You must try to respond quickly, as there will be a time limit in which time you must give your response. If you exceed the time limit, the machine will go on to the study portion of the trial.

Following the "test" portion of the trial will be a pause. Then the word "study" will appear on the screen. Below the word "study" will appear the English word you were just tested on paired with the currently correct answer. This is always the correct response which you must try to remember. Feel free to use any coding mnemonics which help you to remember the response.

Sometimes the response presented for study will be different than the previously correct response associated with the given word. In this case, forget the previously correct response and learn the new response (the old one is now wrong). You will be warned just before the study trial if the response is being changed, so that you will never fail to notice that a change has occurred. This warning will be: "answer changes."

You will be given several seconds to study the current word-letter pair, and then, after a brief pause, the next trial will begin (i.e., a new test trial will occur). Each session will consist of a continuous sequence of these trials.

The experimenter will give you instructions regarding which booth to use, how to start each session, and what to sign each day."

The experimenter reviewed the instructions with the subject and then introduced him to the computer and its operation. The entire first session was used to familiarize the subject with the apparatus and instructions, and to give him practice at the task.

Procedure. As noted earlier, each session consisted of a sequence of 400 trials. Each trial involved a standard series of events. (1) The word Test appeared on the upper face of the CRT. Beneath the word Test appeared the member of the stimulus set indicated by the presentation schedule of Appendix 2. The subject then typed a letter if he felt he knew the response. If he was sure he did not know the response, then he could terminate the test trial by typing a carriage return. If an incorrect response was typed, then the words WRONG. TRY AGAIN appeared on the CRT below the previous response, which remained displayed. The subject could then respond, not respond, or type a carriage return, as for the first guess. If the subject had not typed a response within 3 sec. for the first-guess, or within 2.7 sec. for the second-guess, then the test phase was terminated. (2) The computer next determined whether the response to the current stimulus was to be changed; if so, the CRT was blanked momentarily, and then the following words appeared: ANSWER CHANGES. After 1/2 sec. the study phase began. If the response was not to be changed, then the CRT was simply left blank for 1/2 sec. until the study phase began. (3) The screen was blanked and then the word STUDY appeared at the top of the CRT. Beneath the word STUDY appeared the stimulus just tested along with the correct response to be remembered (changed or not as was appropriate). This display remained for 3.0 seconds. (4) The CRT was blanked for 1/2 sec. and then the next trial began. Using this procedure, the session of 400 trials took about 55 minutes.

At the start of each session, the computer randomly assigned each subject 95 stimuli he had not seen in previous sessions. Each stimulus was then randomly assigned two different letters as responses, with the restriction that the first letter of a stimulus could not be used as its response. The first 14 trials consisted of 10 filler items, items which appeared only seldom thereafter.

Altogether 147 subject-sessions of data were collected (not counting the practice sessions). Due to computer shutdown and other extraneous factors, only 122 of these sessions were entirely completed, the remainder being close to completion. The data collected consisted of the entire sequence of events within each session, including the latencies of the responses. At the conclusion of the experiment each subject filled out a written questionnaire.

Results of Experiment II

A large amount of data will be presented in the present section. As it is rather difficult to grasp without a theoretical basis, detailed discussion will be put off until the next chapter. An attempt will be made here to limit discussion to certain highlights. In the following the first response given by the subject is termed a "first-guess," and the second response when given by the subject is termed a "second-guess." Table II-6 presents the probability of a correct first-response for each subject, lumped over all trials and sessions. The results are listed in order of increasing probability correct. It is evident that there is a wide range in subject ability at this task. Despite this, the remaining data is presented in a form averaged over all subjects in order to gain precision of estimates. This should not overly distort the observed effects, since a subject by subject breakdown of the data seemed to show the same qualitative effects holding for individual subjects as for the group average.

TABLE II - 6

MEAN PROBABILITY CORRECT
FOR SUBJECTS OF EXPERIMENT II

Subject Number	7	6	2	14	3	13	11	8	9	12	5	1	4	10
Probability Correct First-guess	.29	.30	.34	.36	.41	.49	.51	.51	.51	.51	.53	.56	.68	.69

TABLE II - 7

MEAN PROBABILITY CORRECT
FOR SUCCESSIVE DAYS OF EXPERIMENT II

Day Number	1	2	3	4	5	6	7	8	9	10
Probability Correct First-guess	.52	.48	.44	.48	.45	.50	.47	.42	.49	.52

Table II-7 gives the probability of a correct first-guess on successive days of the experiment (practice day not included). There is no evidence for a trend over days. Apparently, as in Experiment I, proactive interference from session to session was not an important factor. The data to follow will be averaged over all sessions. In the following discussion an error will be taken to mean the absence of a correct response; the term intrusion will be reserved for overt errors.

First-Response Data. Figure II-7 presents, in the top panel, the probability of a correct first-guess for each of the item-types listed, at each of their presentations. Figure II-8 presents the same probability for the remaining item-types. Consider first the top panel of Figure II-7. The observed data is represented by open circles; ignore the predictions for the present. The vertical line in each graph delineates the point at which the R1 response is changed. Following the change of response all lags are 10. The successive lags previous to the change are presented in the item-type name at the top of each graph. There are slightly more than 1000 observations at each point shown. The most important features of these data relate to the question of proactive interference. In conditions (10,10), (10-10), and (10-10-10-10), the probability correct after one reinforcement is about .55. The first test after the response changes, however, has a probability correct of about .41. Hence an overall proactive effect is present. A comparison of all five conditions reveals that the proactive effect is not dependent upon the number of reinforcements prior to the change of response, nor upon the terminal probability correct just prior to the change. This is true despite a reasonable range in both variables:

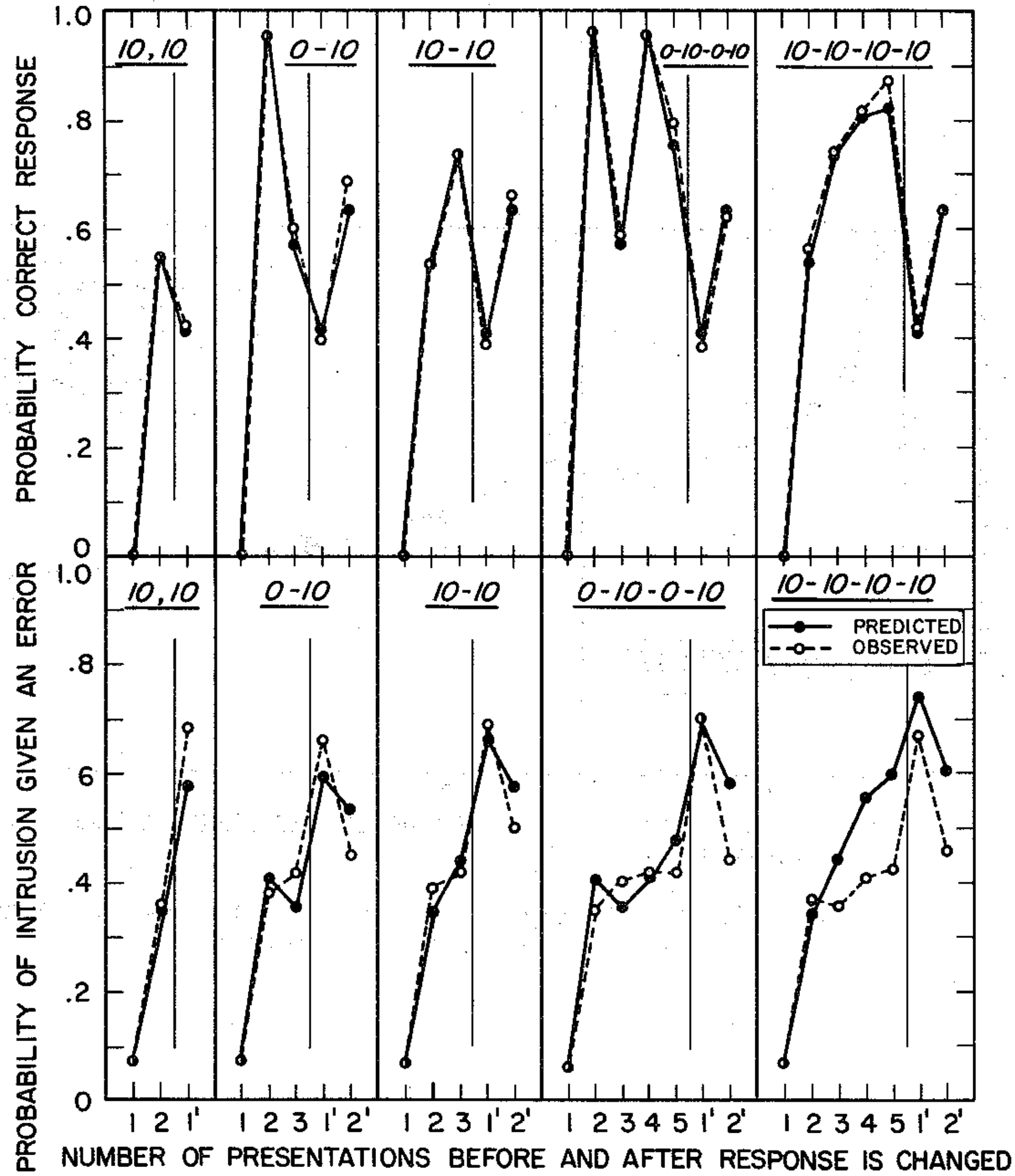


Figure II-7. Probability of First-Guess Correct Responses and First-Guess Intrusions, for the Major Item-Types.

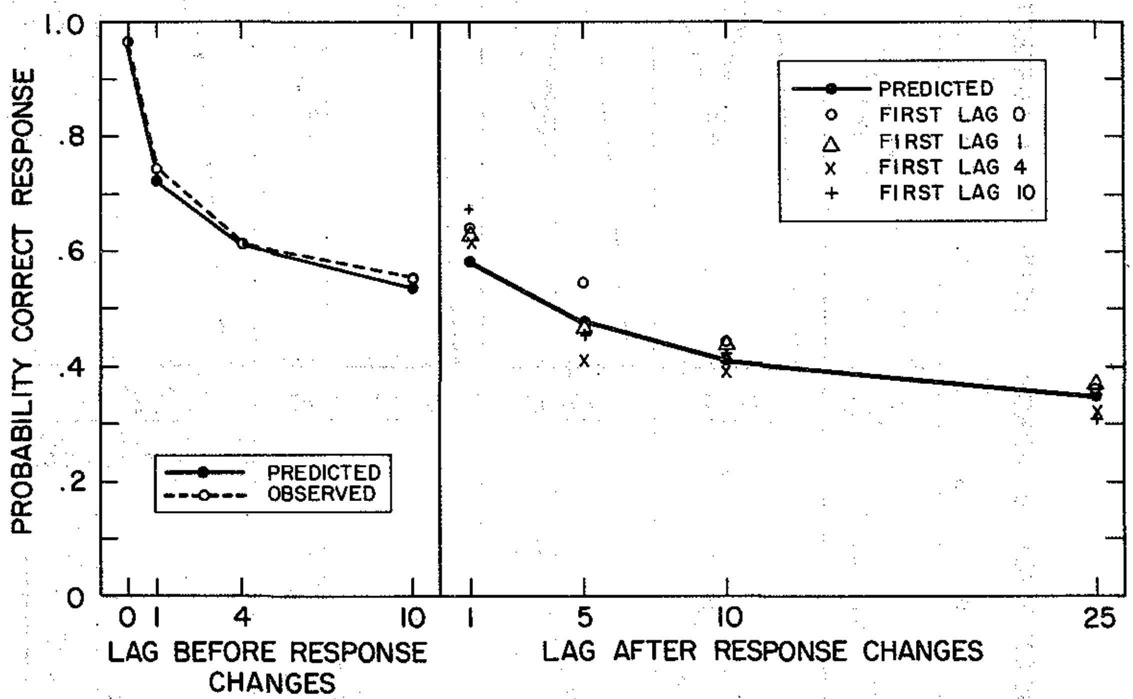


Figure II-8. Probability of Correct First-Guesses as a Function of Lag, for the Matrix Item-Types.

the number of initial reinforcements takes on the values 1, 2, and 4; the terminal probability correct takes on the values .55, .61, .74, .80, and .87; the probability correct after the change of response takes on the values .42, .40, .39, .39, .42. A similar result appears to hold for the second test following the change of response. This lack of dependence upon the degree to which the first response is learned raises some questions about the source of the overall proactive effect. In particular, one must consider the hypothesis that the subjects, having been informed that the response is changing, attempt to code the new pairing with a probability smaller than for an R1 reinforcement. This hypothesis, and a number of models which can account for the observations, will be dealt with in the following chapter.

Figure II-8 presents much the same pattern of results as those just discussed. This figure gives the probability of a correct first-guess for the test before and after the response is changed, where the lag previous to, and following, the change of response is varied. The left-hand panel presents the first-reinforcement lag curve for lags 0, 1, 4, and 10. The observations are the open circles. Following each of these lags the response is changed and a second lag of 1, 5, 10, or 25 ensues. The right-hand panel in the figure presents the results for the 16 resultant conditions, henceforth termed the "matrix" item-types. If variations in the first lag did not have a differential proactive effect, then the four observations at each lag in the second panel should not differ from each other, which seems to be the case. The data are somewhat more unstable than in the previous figure because each point in the right-hand panel is based on approximately 400 to

500 observations. Points in the left-hand panel are based on about 1800 observations.

Figure II-7 presents, in the bottom panel, the probability that a false intrusion response was given, conditionalized upon the fact that a correct response was not given (the unconditional probability of an intrusion was divided by 1.0 minus the probability correct). In the following we refer to a response given in error which had previously been associated with the tested stimulus as an old-intrusion. Other intrusions are called new-intrusions. In Figure II-7 both types are lumped. The observed points are represented by open circles. Several points should be noted concerning these graphs. The intrusion rate for newly presented items is above zero (about .07), but well below that observed on succeeding trials. If the subject searched his memory for an answer on every new trial, it might be expected that an intrusion rate higher than those on succeeding trials would result. The relatively low rates observed would be expected if the subject was often recognizing quickly that the stimulus presented was new, and thereby ceasing further memory search. Note also that there is a considerable increase in intrusions following the change of response - in fact, the increase in number of intrusions is considerably larger than the decrease in probability correct at those points. Most of the increase in intrusions following change of response is of course in old-intrusions. Table II-8a gives the probability of an old-intrusion for the major item-types, conditional upon the fact that a correct response was not made. The numbers in parentheses are predictions which may be ignored for the moment. Before the change of response the probability of an old-intrusion

TABLE II - 8

FIRST-GUESS INTRUSIONS
(Predicted Values in Parentheses)

Table II - 8a: Probability of Old-Intrusion Given an Error

		Item Type			
		0-10	10-10	0-10-0-10	10-10-10-10
Number of presentation after change of response	1	.461 (.345)	.517 (.443)	.552 (.471)	.514 (.470)
	2	.171 (.216)	.225 (.153)	.171 (.238)	.236 (.238)

Table II - 8b: Probability of Intrusion Given an Error

		First Test		Second Test			
				Second Lag			
First Lag	0	.38 (.41)	0	.84 (.78)	.60 (.73)	.64 (.68)	.60 (.65)
	1	.40 (.31)	1	.69 (.65)	.74 (.62)	.66 (.59)	.61 (.57)
	4	.40 (.34)	4	.63 (.62)	.71 (.61)	.68 (.58)	.64 (.56)
	10	.37 (.35)	10	.59 (.60)	.62 (.59)	.69 (.58)	.65 (.55)

Table II - 8c: Probability of Old-Intrusion Given an Error

		Second Lag			
		1	5	10	25
First Lag	0	.65 (.66)	.34 (.53)	.39 (.47)	.33 (.39)
	1	.42 (.49)	.57 (.40)	.47 (.35)	.33 (.30)
	4	.40 (.46)	.45 (.37)	.47 (.33)	.44 (.29)
	10	.36 (.41)	.40 (.35)	.44 (.32)	.41 (.28)

is zero, so these trials are not tabled. Note that in the table the old-intrusion rate shows a tremendous decrease from the first to the second test of R2. This might be explained if the subject was learning on the first trial that the old-intrusion he had given was wrong - this intrusion would then be repressed on the next trial. The intrusion results for the item-types where the lag was varied are presented in Table II-8b and II-8c. Table II-8b gives the lumped results, and Table II-8c the old-intrusion results. Discussion of these tables are reserved until the next chapter.

For a number of reasons it might be felt that intrusion rates should increase as the duration of the session lengthened. This possibility may be examined by considering intrusions on items presented for the first time at different locations in the trial sequence. Figure II-9 presents these results. Intrusion rates are averaged for successive groups of eight new items during the trial sequence. The graph demonstrates that a fairly orderly increase in intrusion rates occurs, though not of large magnitude.

Second-Guess Data. Figure II-10 presents data for second-guesses following new-intrusions on the first guess. The top panel presents the probability of a correct second-guess for the major item-types. Table II-9a presents the same probabilities for the item-types on which the lag was varied. It may be observed that the second-guess curves follow the first-guess curves in general form; there is a rise before the change in response and then a sharp drop after the change. Furthermore, across conditions, variations in presentation schedules prior to the change do not seem to affect the second-guessing rate following the change; this fact conforms to the first-guess finding.

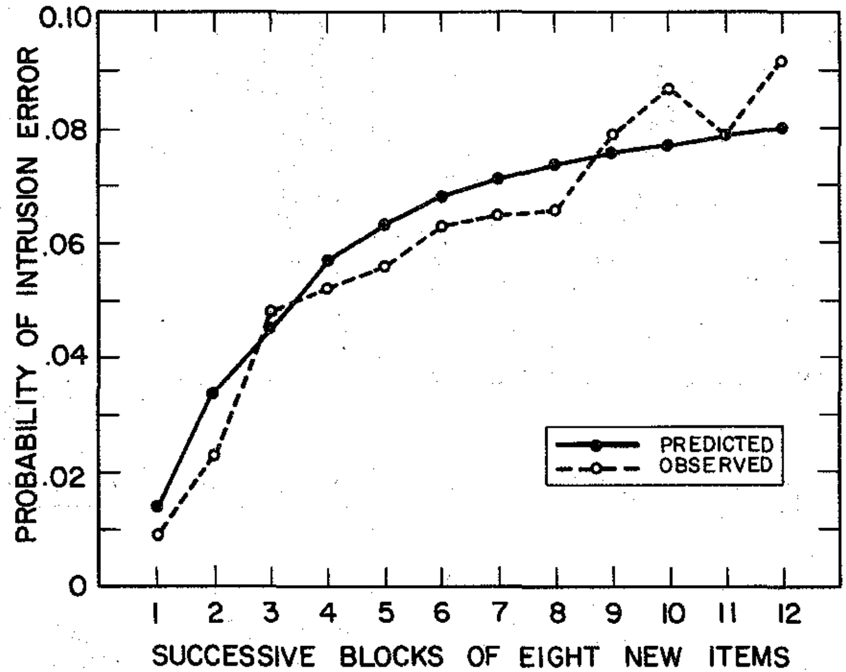


Figure II-9. Probability of Intrusions for New Items, as a Function of Duration of Session.

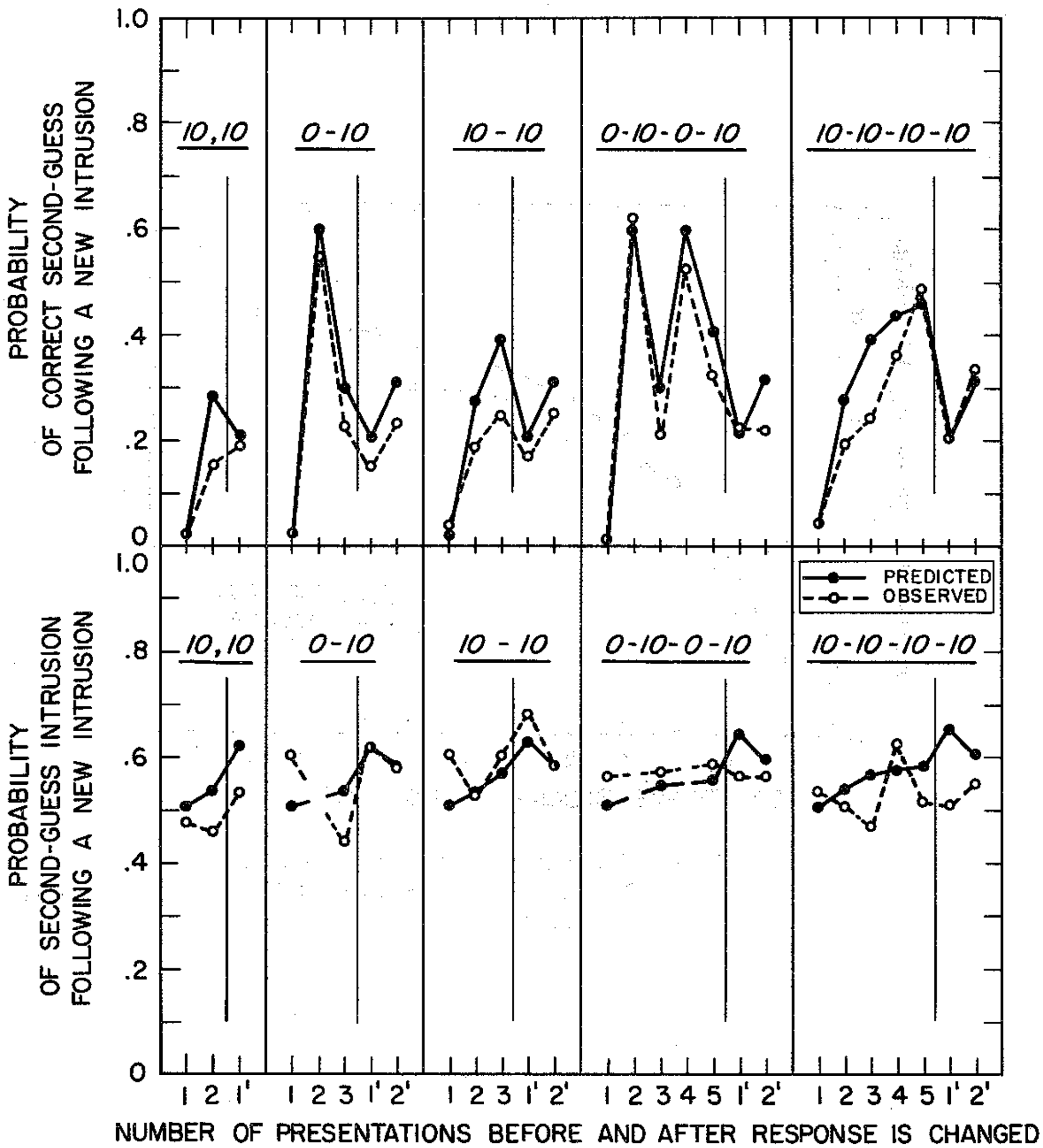


Figure II-10. Probability of Second-Guess Correct Responses and Second-Guess Intrusions, for the Major Item-Types.

TABLE II - 9
SECOND-GUESS INTRUSIONS
(Predicted Values in Parentheses)

Table II - 9a: Probability of Correct Second-Guess Following a New Intrusion

First Test

First Lag	
0	.65 (.60)
1	.25 (.32)
4	.15 (.33)
10	.20 (.28)

Second Test

First Lag \ Second Lag	1	5	10	25
0	.15 (.24)	.19 (.24)	.18 (.23)	.13 (.17)
1	.35 (.21)	.17 (.23)	.10 (.21)	.11 (.17)
4	.18 (.21)	.18 (.23)	.16 (.22)	.07 (.18)
10	.15 (.20)	.28 (.23)	.20 (.21)	.11 (.18

Table II - 9b: Probability of Second-Guess Intrusion Following a New Intrusion, Conditional Upon a Second-Guess Error. Top Matrix for Second-Guess New Intrusions. Lower Matrix for Second-Guess Old-Intrusions.

First Test

First Lag	
0	---
1	.45 (.46)
4	.45 (.51)
10	.44 (.53)

Second Test

First Lag \ Second Lag	1	5	10	25
0	.53 (.30)	.37 (.40)	.36 (.45)	.44 (.49)
1	.50 (.35)	.37 (.43)	.42 (.46)	.39 (.51)
4	.28 (.39)	.39 (.44)	.52 (.47)	.51 (.50)
10	.36 (.42)	.32 (.45)	.48 (.50)	.41 (.51)
0	.06 (.32)	.12 (.25)	.15 (.21)	.19 (.16)
1	.18 (.19)	.18 (.18)	.12 (.15)	.17 (.12)
4	.16 (.17)	.19 (.16)	.11 (.14)	.15 (.12)
10	.11 (.14)	.13 (.14)	.06 (.13)	.09 (.11)

Table II - 9c: Second-Guess Old Intrusions

	0-10	10-10	0-10-0-10	10-10-10-10
1	.12 (.13)	.17 (.17)	.23 (.18)	.22 (.22)
2	.07 (.06)	.10 (.07)	.10 (.07)	.07 (.08)

The lower panel in Figure II-10 presents the probability of any intrusion on the second-guess following a new-intrusion on the first-guess. The probability plotted is conditional upon a second-guess error. Table II-9b presents the same data for the item-types on which lag was varied. Table II-9c presents the second-guess old-intrusion rate for the item-types in Figure II-10. The first point to notice about the observations is the rather high rate of intrusions as compared with the rates observed on the first guess. Whereas the intrusion rates on the first-guess lie at about the .40 level, the second-guess intrusions are between probabilities of .5 and .6.* One possible interpretation of this finding would hold that the subject's decision criterion for output of responses found during memory search has been lowered on the second-guess. Particularly interesting is the intrusion rate for new items: Having made a wrong first-guess on a new item, subjects will then make a wrong second-guess with a probability of almost .60 (which can be compared with the first-guess new-intrusion rate of .07). An implication of this result is that once a decision has been made to search LTS on the first-guess, a search will always be made on the second-guess.

Table II-10 presents the data dealing with second-guesses following old-intrusions given on the first-guess. The results should be noted carefully because they are rather crucial to the model used in Chapter III. Table II-10a gives the probability correct following an old-intrusion. This probability is quite high -- higher even than that

---

*A part of this rise might have been due to subject selection, but a subject-by-subject breakdown showed 13 out of 14 subjects to have higher overall second-guess than first-guess intrusion rates.

following a new-intrusion. Table II-10b gives the probability of second-guess new-intrusions following first-guess old-intrusions. We shall merely note for the present that this new-intrusion rate is lower than the new-intrusion rate following first-guess new-intrusions.

Latencies. It is beyond the scope of this report to make a thorough analysis of the latency results. Tables II-11 through II-15 present the mean latencies for all item-types for the following conditions: a) correct first-guess responses, b) first-guess old-intrusions, c) first-guess new-intrusions, d) correct second-guesses following old-intrusions, and e) correct second-guesses following new-intrusions. We mention here the following results. (1) The latencies of a correct response decrease as the number of reinforcements increase; i.e., for the (10-10-10-10) condition the mean latencies are successively 1.52, 1.42, 1.36, 1.33. (2) The longer the lag, the longer the latency of a correct response. For initial lags of 0, 1, 4, and 10, the mean latencies of a correct response are 1.03, 1.37, 1.50, and 1.56. This result would have a natural interpretation if memory search were temporally ordered to some degree, but could also be handled if there were a significant amount of correct retrieval from a fast access short-term store at the shorter lags. (3) The latencies of a correct response following the change of response are slower than the corresponding latency for the first response. Nevertheless, these latencies after the change of response do not vary as a function of the type of sequence prior to the change. This result is in good accord with the response data; i.e., the change of response has an effect, but an effect independent of the history preceding it.

TABLE II - 10

SECOND-GUESSES FOLLOWING OLD-INTRUSIONS AS FIRST GUESSES

Table II - 10a: Probability Correct

	Number of Presentations 1	2
0-10	.31	.54
10-10	.27	.50
0-10-0-10	.23	.51
10-10-10-10	.27	.39

First Lag	Second Lag 1	5	10	25
0	.35	.28	.28	.29
1	.42	.33	.41	.24
4	.43	.34	.29	.22
10	.42	.30	.24	.29

Table II - 10b: Probability New Intrusions Conditional Upon a Second Guess Error

	Number of Presentations 1	2
0-10	.36	.52
10-10	.36	.51
0-10-0-10	.30	.44
10-10-10-10	.34	.44

First Lag	Second Lag 1	5	10	25
0	.31	.33	.38	.36
1	.29	.30	.25	.37
4	.26	.36	.27	.32
10	.35	.33	.35	.32

TABLE II - 11

MEAN LATENCIES FOR CORRECT FIRST-GUESSES

First Response Test

	P1	P2	P3	P4
0-10	1.04	1.51		
10-10	1.55	1.45		
0-10-10-10	1.04	1.53	1.14	1.42
10-10-10-10	1.52	1.42	1.36	1.33

Second Response Test

	P1	P2
0-10	1.66	1.54
10-10	1.63	1.57
0-10-0-10	1.63	1.54
10-10-10-10	1.67	1.59

P= number of previous presentations of the stimulus-response pair

---

First Response Test

First Lag	
0	1.03
1	1.37
4	1.50
10	1.56

Second Response Test

Second Lag

First Lag	1	5	10	25
0	1.46	1.56	1.52	1.61
1	1.42	1.57	1.56	1.73
4	1.48	1.72	1.64	1.67
10	1.37	1.60	1.64	1.63

TABLE II - 12

MEAN LATENCY OF FIRST-GUESS
OLD INTRUSIONS

First Response Test

	P1	P2
0-10	1.60	1.83
10-10	1.63	1.83
0-10-0-10	1.67	1.77
10-10-10-10	1.62	1.94

P= number of previous presentations of the stimulus-response pair

---

Second Response Test

Second Lag

First Lag	1	5	10	25
0	1.52	1.63	1.68	1.77
1	1.60	1.56	1.59	1.65
4	1.57	1.55	1.60	1.57
10	1.43	1.57	1.60	1.65

TABLE II - 13

MEAN LATENCIES OF FIRST-GUESS
NEW-INTRUSIONS

First Response Test

	P1	P2	P3	P4
0-10	1.40	2.00		
10-10	2.03	2.06		
0-10-10-10	1.85	2.03	1.56	2.03
10-10-10-10	2.04	1.93	1.98	1.94

Second Response Test

	P1	P2
0-10	2.05	2.11
10-10	2.03	2.05
0-10-0-10	2.05	2.07
10-10-10-10	2.07	1.92

P= number of previous presentations of the stimulus-response pair

---

First Response Test

First Lag	
0	1.44
1	1.99
4	1.98
10	2.07

Second Response Test

Second Lag

First Lag	1	5	10	25
0	1.79	1.87	2.02	2.12
1	1.85	2.08	2.01	2.06
4	1.91	1.97	2.10	2.06
10	1.93	2.10	1.92	2.17

TABLE II - 14

MEAN LATENCY FOR CORRECT SECOND-GUESSES
FOLLOWING OLD-INTRUSIONS

First Response Test

	P1	P2
0-10	1.54	1.29
10-10	1.61	1.00
0-10-0-10	1.49	1.27
10-10-10-10	1.74	1.26

P= number of previous presentations of the stimulus-response pair

---

Second Response Test

Second Lag

		1	5	10	25
	0	1.50	1.73	1.66	1.46
First	1	1.52	1.83	1.52	1.57
Lag	4	1.59	1.52	1.55	1.50
	10	1.59	1.53	1.50	1.48

TABLE II - 15

MEAN LATENCIES OF CORRECT SECOND-GUESSES FOLLOWING NEW-INTRUSIONS

First Response Test

	P1	P2	P3	P4
0-10	0.70	1.36		
10-10	1.37	1.23		
0-10-10-10	1.08	1.28	0.81	1.53
10-10-10-10	1.44	1.35	1.16	1.21

Second Response Test

	P1	P2
0-10	1.33	1.55
10-10	1.40	1.20
0-10-0-10	1.35	1.12
10-10-10-10	1.33	1.35

P= number of previous presentations of the stimulus-response pair

---

First Response Test

First Lag	
0	0.63
1	1.34
4	1.37
10	1.33

Second Response Test

First Lag \ Second Lag	1	5	10	25
0	0.93	1.43	1.19	1.32
1	1.50	1.06	1.30	0.82
4	1.60	1.65	1.15	1.42
10	1.42	1.64	1.20	1.27

We next turn to the intrusion latencies. The mean latencies of intrusions, both old and new, are slower than the corresponding correct latencies in all cases; however, the latencies of new-intrusions are markedly longer than those of old-intrusions. This result, as will be seen in the next chapter, has important implications regarding the temporal ordering of the memory search. The latency of new-intrusions, as opposed to the correct latencies, does not vary as the number of reinforcements of R1 increases. The latency of a new-intrusion seems to be slower the longer the lag since the correct response, but the effect is essentially eliminated if lag = 0 is not considered. Finally, turning to the second-guess results, we will mention here only the following fact: after the change of response, the mean latency for a correct second-guess is shorter following new-intrusions than following old-intrusions. This would be surprising if the source of first-guess old-intrusions arose in confusion of the old and new responses. That is, if the old and new responses were confused and the subject chose one to output, then it might be expected that it would not take long to output the other after a wrong first-choice.

Conclusions

A rather large amount of diverse data has been collected in the two experiments. The variables examined include lag between study and test, number of reinforcements, second-guessing, rankings, negative transfer, intrusion rates for both first- and second-guessing, and latencies of response. A storage and retrieval model of long-term memory was described in Chapter I which, at least theoretically, had the capacity to deal with these variables simultaneously. In the next chapter it will be seen whether an explicit model based on the general theory can deal quantitatively with the data.

CHAPTER III

THEORETICAL ANALYSIS: A STORAGE AND RETRIEVAL MODEL

The derivation of a quantitative model from the theory presented in Chapter I involves a large number of individual decisions. The number of possible models that could be derived is extremely large, and this report cannot compare and contrast them all. Rather, an attempt will be made to construct the simplest possible model consistent with both the overall theory and the data. A few variations of the resultant model will also be discussed.

A model will first be presented for the data of Experiment I. This model will then be extended, but not altered, in an attempt to predict the data of Experiment II, data involving a number of additional variables.

Experiment I

The Short-Term System. The subject is assumed to pay some attention to each item presented for study, and thereby enter it into STS, at least momentarily. Therefore a test at lag 0 should result in nearly perfect performance (since the study phase and the test phase of the next trial are separated by only 3/4 sec.). We do not wish to involve ourselves in predicting just how good performance on such a zero-lag test should be (we would have to consider typing mistakes, and so forth) and therefore will treat the few zero-lag trials that occur as special cases. The first-guess and second-guess predictions for performance at zero-lag are simply set equal to the mean probability which was observed in all such instances, .97 and .50 respectively.

The present task was designed so that the short-term control processes utilized would tend to be single-trial coding mechanisms, rather than multi-trial rehearsal operations. That the design was successful in this regard is indicated both by subject reports and by the relative lack of an effect due to the type of intervening item at a lag of 1. Nonetheless, some items are undoubtedly maintained in STS beyond the trial of presentation -- this could occur if the subject takes more than one trial to encode certain items, or if some items previously encoded are given a small amount of additional rehearsal. It is therefore proposed that any item for which a storage attempt is not made decays rapidly from STS and is lost by the termination of the following trial. On the other hand, items which are coded decay from STS at a rate independent of the type of intervening items. Specifically, let $P(A)$ represent the probability that a storage attempt is made for a particular item; note that $P(A)$ includes the probability that the item is already in STS when presented on a trial. Let $P(R_i)$ represent the probability that the item will be present in STS at a lag of $\underline{i}$. Then we have the following:

$$P(R_0) = .97 \qquad \text{Eq. III-1}$$

$$P(R_i) = P(A) \times (1-\alpha_1)^i$$

where $\underline{\alpha_1}$ is a parameter governing decay from STS. It might be asked whether there is a reason other than intuitive for including a decaying short-term process in the model. As it will be seen later, it is through the action of this process that a distributed learning effect is predicted by the model.

There is one important exception to the stated results concerning lack of organized rehearsal. The design of the experiment was such that a test of an item at lag 1 was almost always followed by a sequence of further tests of that item at lag 1. All subjects reported noting this fact, and a majority of them reported specifically rehearsing these items when they were noticed. As a result, performance on Type 1 and Type 2 items was abnormally high for presentation numbers 3, 4, 5, and 6. Rather than add to the model a specific rehearsal process to account for these observations, we will merely comment that it would be easy to do so.

Storage. When a currently unretrievable item is presented for study, an attempt may be made to store it. Let α be the probability of attempting to store such an item. The information stored will involve three components: stimulus, response, and associative information ($F(I_s)$, $F(I_r)$, and $F(I_a)$). As the present experiment is not designed to emphasize the differences between these information measures, we will characterize the amount of information transmitted to LTS by a single measure, $F(I)$, where the components of $F(I)$ include the three measures above. The exact form of $F(I)$ is not crucial to the model, but a reasonable spread in its distribution is necessary (a spread in the distribution is needed to predict both the first-guess lag curve and the rather low, and invariant, second-guessing performance over lags). For the purpose of simplifying calculations $F(I)$ will be approximated by a two-point distribution as follows. $F(I)$ is divided at its median; codes with strengths above the median will be called hi-codes and defined to have strength σ_H; codes with strengths below the median will be called lo-codes and defined to have strength σ_L ($\sigma_H > \sigma_L$). Thus an attempt to store information will result in a lo-code with probability .5 and will result in a hi-code with probability .5. The information stored will be placed in a location determined by stimulus characteristics, but because the present experiment uses a continuous task with homogenous items, the placement will not be ordered from the point of view of the model. Hence the model will treat placement as an essentially random process.

There are a number of decision rules which determine whether a storage attempt will be made for a particular item. Basically, a storage attempt will be made with probability α only when a correct response has not been retrieved from STS or LTS on the test phase of the trial. The only exception to this rule occurs at zero-lag. Term the state in which an item enters STS only momentarily, and is not coded, as the null-state. Items in the null-state at test, even though in STS, are treated as if a successful retrieval had not occurred. Thus an attempt may be made to store these items with probability α. These decision rules imply that a code which has just resulted in a successful retrieval will not be disturbed by further storage attempts, a reasonable strategy for the subject to adopt. On the other hand, the act of successful retrieval itself could reasonably be expected to make future retrieval easier. For this reason, lo-codes which have been successfully retrieved from LTS are treated thereafter as hi-codes (the alternative model, in which retrieved lo-codes are not altered, will be discussed later). One final informational change occurs in a code that has been successfully retrieved from LTS: the code is updated temporally to the present.

There are two processes which may occur when an item is given a reinforcement beyond the first. In one, a code which has not been retrieved from LTS will be left untouched, and a new and different code will be introduced during the study phase of the trial. In the other, the unretrieved code will be retrieved while a new storage attempt is made during the study phase, since the correct response is supplied at that time. If the code is retrieved during study, then it may be assumed that the ongoing storage attempt will consist of amending or changing the retrieved code; thus only a single code will result. Most likely, a mixture of these processes will take place during an experiment of the present type. However, because it greatly simplifies matters computationally, we shall assume that only the second hypothesis occurs; thus only a single code can exist for an item at any one time in LTS.*

The proportion of times a coding attempt is made, based on α, should be closely related to the decay rate from STS, α_1; that is, the more coding effort expended on intervening trials, the more likely is an item's loss from STS. For simplicity, we shall assume $\alpha_1 = \alpha$ in the remainder of this chapter.

Retrieval. At zero lag the subject is correct with probability .97 and second-guesses correctly with probability .50. The following

*The extended model, in which a mixture of the two possibilities occurs, will necessarily predict the data more closely than the restricted model actually used. However, the type of data collected in the present experiments is such that the extended model will not be better to an appreciable degree. As it will be seen, the restricted model fits quite well.

discussion does not deal with the zero-lag case. At test, a search is first made of STS; if the item is found, then it is reported correctly with probability 1.0. If the item is not found in STS, a search is made of LTS. We continue to use the terminology of Chapter I: if the stimulus currently being tested has a code stored in LTS, this code is termed the c-code; the other codes stored in LTS are termed i-codes.

For any stimulus tested, only a small subset of the codes stored in LTS will be examined during the search. This subset (termed the examination-subset) will be defined by the characteristics of the stimulus presented, characteristics that lead the subject to examine certain memory regions rather than others. Of course, once the search begins, the successive members of the examination subset will be determined to a large degree by associative factors. For the current experiment, however, the associative factors must be treated as essentially random, and the probability that a c-code will be in the examination subset depends only upon the "age" of the code, and the strength of the code.

Although the search through memory proceeds one code at a time, the clearest exposition results if we consider the search process in two stages. First we define a potential examination-subset, containing all those codes that will eventually be examined if the search continues long enough. In the second stage we define the order of search through the subset, and the probability of terminating the search and emitting a response at some point. Let $P(Z_i)$ be the probability that a c-code will be in the examination-subset, if the current test is at lag i. Then

$$P(Z_i) = \frac{\sigma}{\sigma + \beta \text{ (age)}} \qquad \text{Eq. III-2}$$

where σ is the strength of the c-code (either σ_H or σ_L), age is some function of i, and β is a parameter ($0 \leq \beta < \infty$) governing the dependence of $P(Z_i)$ upon age. Since evidence was presented in the previous chapter that the probability correct depended upon the degree to which the intervening items were "known," the age of an item is defined to equal the mean number of new codes that were stored during the lag since the item's last presentation. The mean is taken over all possible realizations of the experiment; it is used rather than the actual number of new codes stored as an approximation to make the mathematics of the model tractable. The particular function presented in Eq. III-2 was utilized because it conforms to the criteria mentioned in Chapter I, and because of its simplicity. At large i, the value of $P(Z_i)$ decreases quite slowly as i increases, but at small i an appreciable decrease occurs.

If a c-code is examined during the search two processes can occur; first, a response may be recovered; second, the subject engages in a decision process to decide whether to emit any response recovered. In the following, the possibility that a response other than the one encoded will be recovered from the c-code will not be considered; this possibility will be taken up instead in the intrusion rate from i-codes. The probability of recovery and output should then be a straightforward function of the strength of the code: designate ρ_1 as the probability of recovery and output on the first-guess search, given a code was examined. Then,

$$\rho_1 = 1 - \exp(-\sigma) \qquad \text{Eq. III-3}$$

where exp is the exponential function ($\exp(k) = e^{-k}$) and σ is the strength of the code examined.

Next we turn to a consideration of intrusions, where an intrusion refers to the recovery and output of a response, as the result of the examination of an i-code during the search. The probability that an i-code will be in the examination-subset will depend in part upon the similarity of its stimulus to the stimulus being tested, but on the average this probability will be considerably smaller than for a c-code. Similarly, the probability that examination of an i-code results in the recovery and output of a response is considerably less than for a c-code. Each of these possibilities may be incorporated into the model by introducing the concept of effective-strength of an i-code, σ_I, where σ_I is less than either σ_H or σ_L. The degree to which σ_I is less than σ_H or σ_L should depend upon the similarity, or amount of generalization, between the stimuli used in the experiment. Note that it does not matter whether an i-code is a hi-code or a lo-code; its strength is σ_I in both cases. (While on the one hand a hi-code will be in the examination subset and lead to response recovery more often than a lo-code, on the other hand a hi-code is more likely to contain information which will inhibit intrusions during response-production.) Equations III-2 and III-3 can now be generalized to include i-codes: depending on the code being examined, σ in these equations will take on the value σ_I, σ_H, or

σ_L. Note that the age in Equation III-2 applies to the code under examination, and not necessarily to the item being tested.

The final component of the search process to be specified is the order of search through the examination-subset. To begin with, note that the experimental design utilized does not induce an order in the search (as might be the case if the stimuli were grouped in some obvious manner). In Chapter I it was suggested that an item would tend to be examined earlier in the search, the greater its strength and the lesser its age. We choose here to assume a strictly temporal search, independent of the strength of the codes. While this assumption cannot be entirely accurate, it should prove instructive to see how far it can be carried. Furthermore, it has the advantage of making the mathematics of the model tractable.

The memory search is assumed to be terminated when the first response is recovered and output; this seems reasonable if responding is required to be fairly rapid. As noted in Chapter I, this assumption leads to predictions that rankings and rerankings beyond the first choice will be at the chance level, which is close to the effect observed. If every code in the examination-subset is examined without a response being recovered and output, then the subject guesses randomly.

Following an error (an incorrect first-ranking) the subject engages in a second search of LTS. The second search is identical to the first, except that the decision criterion for output of recovered responses is lowered. This assumption is based on the results of Experiment II, where it was observed that the intrusion rates were considerably higher for second-guesses than for first-guesses. The change in decision criterion is assumed to apply to all codes, and is governed by a parameter γ as follows: let ρ_2 be the probability of recovery and output on the second-guess search, given that a code was examined. Then,

$$\rho_2 = 1 - \exp(-\gamma\sigma), \quad \gamma > 1. \qquad \text{Eq. III-4}$$

Equation III-4 is of course the counterpart of Equation III-3 for the first-guess search. The second-guess search is assumed to proceed independently of the first-guess search, but a c-code examined and rejected on the first-guess cannot give rise to a response on the second-guess.*

Review of the Model. The model utilizes six parameters:

α: governs the probability of a coding attempt, and decay from STS;

β: adjusts the degree to which an item's probability of being examined during the search depends upon age;

σ_H: the strength (amount of information stored) for a hi-code;

σ_L: the strength for a lo-code;

σ_I: the strength for an i-code (a code for an item other than the item currently being tested)--governs intrusions;

γ: adjusts the decision criterion for output of a recovered response during the second-guess search.

When an item is presented for test, a memory search commences. At zero-lag the probability correct is .97 and the probability of a correct second-guess is .50. Otherwise, if the item is currently present in

*In fact, this assumption makes almost no difference in the predictions for the data of Experiments I and II, compared with the complete independence assumption. It was used here because it seemed reasonable that the same c-code examined twice within a second or two would seldom give rise to differing results. The same does not apply to i-codes because σ_I is low enough that the change in decision criterion on the second-guess will make a significant difference.

STS, then a correct response is output. If the item is not in STS, then a search of LTS begins. The search takes place through a subset of the codes stored in memory, termed the examination-subset. The probability that a particular code will be in the examination-subset is given by Eq. III-2. The subject considers each code in the examination-subset in temporal order, the most recent first. The probability of recovering and outputting a response while considering a particular code is given by Eq. III-3. If all the codes in the subset are examined, but no response is emitted, then the subject guesses randomly. Whenever a response is recovered and emitted, the search is terminated and the subject ranks the remaining alternatives randomly. If the first-ranking proves to be incorrect, then a second search is initiated. This search is identical to the first, except that the decision criterion for output of a recovered response is lowered. In addition, a c-code examined and rejected during the first search cannot give rise to a response on the second search.

During the study phase of a trial the following events take place. If a successful retrieval had been made from LTS, then the code utilized is temporally updated to the present; in addition, a lo-code retrieved successfully becomes a hi-code. If a retrieval had been made from STS, then no new code is stored. Following any incorrect retrieval, or a pure guess, or a retrieval at zero-lag from the null-state, an attempt is made to store with probability α. If a storage attempt is made, then a hi-code will result with probability .5, and a lo-code will result with probability .5. Following a storage attempt, an item will leave STS with probability α on each succeeding trial.

In the following sections of the paper the model will be used to predict second-guessing data, among other phenomena. It should be noted that these data are conditional upon first-guess errors, and therefore are subject to considerable selection effects due to subject-item differences. The model predicts such selection effects since codes are assumed to be stored which have differing strengths. Thus selection due to subject-items should present no difficulties. This is not true, however, if items are selected on the basis of their performance on previous trials. Large subject differences are observed in both experiments; these differences will result in a considerable distortion of sequential phenomena which will not be predicted by the model. For this reason, this paper will not deal with sequential phenomena (such as two-tuples of errors on successive reinforcements, etc.).

Mathematical Analysis. The following discussion will be facilitated by a number of definitions. Let $c_{i,j}$ represent a correct response on the ith trial and the jth guess (i gives the trial number in the sequence of 439; j = 1 implies the rankings; j = 2 implies the rerankings). Let $e_{i,j}$ represent the corresponding error function. Let $\Omega_{i,k}$ represent the state of the memory system at trial i, for some realization of the experiment, k. The state of the system is described by three lists: the stimuli which are currently in STS, the stimuli which have lo-codes stored in LTS, and the stimuli which have hi-codes stored in LTS.

We shall deal in the following only with $P(c_{i,j})$, and not with the rankings and rerankings beyond the first choice -- the model predicts these to be at the chance level. We therefore have:

$$P(c_{i,j}) = \sum_{k=1}^{\infty} P(c_{i,j,k} \mid \Omega_{i,k}) \, P(\Omega_{i,k}) \, , \qquad \text{Eq. III-5}$$

where summation is taken over all realizations of the experiment, denoted by k. For certain models this sum would be unwieldy to work with, but for the present model in which search is strictly temporally ordered and in which age is approximated by the mean number of intervening new codes, it is possible to bypass the summation and deal with the average state of the system at each trial, called $\overline{\Omega}_i$. $\overline{\Omega}_i$ may be iteratively calculated trial by trial, and $P(c_{i,j})$ is a relatively simple function of $\overline{\Omega}_{i-1}$. The details of the calculations, which are straightforward but require a cumbersome amount of notation, are reserved for Appendix 3. We note here only the following observation, which has not been stressed previously. When generating the predictions for the second-guess data, one must take into account the selection effect on the proportions of hi- and lo-codes introduced by the first-guess error. For example, many more errors occur if the item being tested has no code stored, or a lo-code stored, than if a hi-code is currently stored. As a result, the second-guess rates conditional on an error can be surprisingly stable over reinforcements and lags.

Using the computational methods described in Appendix 3, predictions can be generated from the model for any given set of parameter values. These predictions consist of the following vector for each of the 439 trials of the experiment: $[P(c_{i,1}); P(c_{i,2}); 1-P(c_{i,1})-P(c_{i,2})]$. Note that $P(c_{i,2})$ is not conditional upon a first-guess error; the numbers graphed in Figures II-5 and II-6 are conditional and equal $P(c_{i,2})/P(e_{i,1})$.

Given predictions for any given set of parameter values, we next define a goodness-of-fit measure. Corresponding to the predicted probabilities above, we define three observational quantities. $O_{i,1}$ is defined to be the observed number of correct first-guesses on the *i*th trial; $O_{i,2}$ is defined to be the observed number of correct second-guesses on the *i*th trial; $E_{i,2}$ is defined to be $N_i - O_{i,1} - O_{i,2}$, where N_i is the total frequency of all responses on the *i*th trial. The goodness-of-fit measure to be used is termed π^2 (Holland, 1967), and is calculated identically to χ^2 as follows:

$$\pi^2(\alpha,\beta,\sigma_H,\sigma_L,\sigma_I,\gamma) =$$

$$\sum_{i=1}^{439} \left\{ \frac{[N_i P(c_{i,1}) - O_{i,1}]^2}{N_i P(c_{i,1})} + \frac{[N_i P(c_{i,2}) - O_{i,2}]^2}{N_i P(c_{i,2})} \right. \qquad \text{Eq. III-6}$$

$$\left. + \frac{[N_i P(e_{i,2}) - E_{i,2}]^2}{N_i P(e_{i,2})} \right\} .$$

N_i in the above equations decreases from 83 when i=1, to 58 when i=439. Although the π^2 distribution is not identical to that of χ^2 because certain independence assumptions are not satisfied in the above sum, a crude approximation to the levels of significance of π^2 can be made by use of the χ^2 tables. In using the tables, the degrees of freedom (d.f.) is equal to twice the number of trials, i, over which the π^2 is summed, minus the number of parameters being estimated (6 in the present case).

The next step is to estimate parameters by minimizing the π^2 function over all possible sets of parameter values. A grid search procedure was used to accomplish the minimization; i.e., a reasonably exhaustive search was made through the possible sets of parameter values, the computer generating predictions and computing π^2 for each set. The set of parameters giving rise to the lowest value of π^2 is assumed to generate the best fit of the model to the data. We will first state that the minimization carried out over all 439 trials resulted in predictions that consistently underestimated presentations 3 through 6 for item-types 1 and 2. As pointed out earlier, however, this was expected since the subjects reported rehearsal schemes for these trials. Therefore, in order not to bias the predictions for the remaining data, the 32 trials of the above type were deleted from the π^2 sum. Thus the π^2 function in what follows is summed over only 407 trials.

Predictions of the Model. The values of parameters which minimized the π^2 function for Experiment I were $\alpha = .68$, $\beta = .286$, $\sigma_H = 10.5$, $\sigma_L = 1.16$, $\sigma_I = .17$, $\gamma = 2.3$. The minimum π^2 value was 871.4, and the number of d.f. = (407)(2)-6 = 808. Since for large d.f. $\sqrt{2\chi^2} - \sqrt{(2)(\text{d.f.})-1}$ is approximately normally distributed with a one-tailed test appropriate, a χ^2 value of 871.4 would be just above the .05 significance level. This is a strong indication that the model and the data were in close agreement on a trial-by-trial basis (if we ignore the abnormal points for item-types 1 and 2). The predictions of the model for the lag curves and the various item-types are shown in Figures II-2 through II-6 (pages 43 through 47) as the solid black points connected by unbroken lines. Except for the central portions of the Type 1 and Type 2 curves, the predictions are quite accurate. Even for the Type 1 and 2 curves the predictions are quite accurate for presentations 1 and 2, before rehearsal has begun, and for presentations 7 and 8, after rehearsal has ceased. Particularly noteworthy are the second-guess predictions, since only a single parameter, γ, has been utilized for adjustment of the second-guessing probability. It is instructive to note how the model predicts the maximum in the second-guess lag curve in Figure II-2 (page 43). At very small lags, all stored c-codes are likely to be retrieved correctly, so that most of the errors will occur when no c-code is stored in LTS; hence second-guesses will not be accurate. At longer lags, more and more intrusions occur before the c-code is reached in the first-guess search, hence more and more c-codes are available in LTS during second-guessing. At very long lags, even though many intrusions occur before the c-code is reached in the first-guess search, and therefore many c-codes are available during second-guessing, the lag is so long that the probability correct drops again. Note also that the distributed practice effect is predicted by the model. Such an effect arises from a short-term decaying store from which little learning takes place (Greeno, 1964). In the present model recovery from STS maintains lo-codes which would otherwise probably be transformed to hi-codes.

We may ask how the model performs under various restrictions and alterations. If $\gamma = 1.0$, which implies that the same bias applies during second-guessing as first-guessing, the predictions of the second-guessing probability are consistently above the observations, and the minimum π^2 almost doubles in value. Hence the altered output criterion implied by $\gamma = 2.3$ is necessary in the model. No restrictions among the three strength parameters, σ_H, σ_L, and σ_I can come close to fitting the data; that is, no two of the strength parameters may be set equal without losing accuracy of the model. An interesting alternative model results if we eliminate the assumption that successfully retrieved lo-codes become hi-codes. The minimum π^2 for the resultant model is 1020.4; the primary reason this model mispredicts is that very little learning is predicted to take place over the first few reinforcements of an item. Reference to Figure II-3 (page 44) shows a large rise in probability correct over the first few reinforcements. The transforming of retrieved lo-codes to hi-codes should not be misconstrued as antithetical to the finding from 3-state Markov models (Greeno, 1967a) that learning from the intermediate state is minimal. There is no simple correspondence between the three states of the Markov models, and the various states of the present model; rather they overlap each other. In any event, the present model does have a state from which little learning occurs: STS. To the extent that one is willing to equate this state and the intermediate Markov state, there is no conflict.

Finally, we may ask how the model predicts if "age" is based upon the number of intervening trials, rather than the number of intervening new codes. The minimum π^2 for this model is 920.0, perhaps not a dramatic increase, but one which confirms the empirical finding in Chapter II that "unknown" intervening items cause more forgetting.

The fit of the model to the data of Experiment I is quite good. The model is able to deal quantitatively, and simultaneously, with variations in number of reinforcements and in lag, with first-guesses and second-guesses, and with rankings and rerankings (in a sense). Nevertheless, the model as it stands has the power to deal with a considerably richer set of data. To be precise, an integral feature of the model is the prediction of intrusions, but intrusions were not observable in Experiment I. Experiment II, therefore, should provide a considerably more stringent test of the model. In addition, the model is extended to predict phenomena relating to the changing of response assignments for individual stimuli.

Experiment II

Before discussing Experiment II we wish to reiterate some important terminology. The term "intrusion" denotes the emission of an incorrect response. Two types of intrusions are possible: "new-intrusion" is used to denote the emission of a response which has never been paired with the stimulus being tested; "old-intrusion" is used to denote the emission of a response which is incorrect but has been paired at some earlier point in the session with the stimulus being tested. That is, an old-intrusion denotes the emission of the R1 response, if the R2 response is currently correct. The term "first-guess" denotes the subject's response during the initial portion of the test trial. If a first-guess intrusion is given, then the subject is given another chance to respond called the "second-guess." Thus, for example, the results of a hypothetical test trial might be described as a "second-guess old-intrusion following a first-guess new-intrusion." This terminology should be noted carefully, since it will be used throughout the remainder of this chapter.

There is one extension of the model that is not related to the change of response. As seen in Figure II-7 in the lower panel (page 65) there is a considerable rise in the intrusion rate following the first presentation of an item. The most likely interpretation of this finding is the one outlined in Chapter I. When the stimulus is presented for test, it is presumably scanned for salient characteristics. If a very salient characteristic is found, a search is then made in the memory location indicated by that characteristic, and if appropriate information is not found there, then the stimulus is identified as new and the search ceases. We therefore introduce a parameter δ to govern this process. Let δ be the probability that a normal search is made for a new item. Thus with probability $1 - \delta$ the stimulus is recognized as new and no search is made. We assume that no previously presented item is recognized as new (presumably old stimuli with high-salient characteristics always have enough information stored in the appropriate location that a recognition occurs and the search continues).

The model must now be extended to account for change-of-response phenomena. In order to make the following discussion clear, we define an o-code to be the code which encodes the R1 response for the item being tested, if the R2 response is currently correct. Thus the image encoding the previously correct response is called an o-code. It will be assumed that when a change of response occurs the o-code, if it is present in LTS, will not be updated temporally, it will simply remain in LTS and may be found during a later search. During a later search of LTS the probability that an o-code will be in the examination subset, and the probability that the R1 response will be recovered, will be the same as for a c-code at that same age. That is, since the stimuli are the same for the two codes, the same strengths apply in Equations III-2 and III-3: σ_H if a hi-code is stored, and σ_L if a lo-code is stored. However, the probability of output of the recovered response must depend upon whether information has been added to the o-code that it is "old" and hence wrong. We shall assume that whenever an R1 response has been retrieved, output, and is incorrect, that this information will be added to the o-code, so that the o-code cannot give rise to an old-intrusion on following trials. During the trial on which the answer is changed, however, the R1 response is correct when given. We therefore introduce a parameter κ defined as the probability that an o-code is tagged as wrong. The tagging is a result of the message ANSWER CHANGES which appears on the CRT, and a result of the changed pairing which is then presented for study. Note that κ applies only on the trial on which the answer changes, and applies only to o-codes which were correctly retrieved during the test phase of the trial.

The model as it now stands, due to the strictly temporal search characteristic, predicts no proactive effect. This is true because the c-code will always be encountered in the search before the o-code, if both are in the examination-subset. It was seen in Figure II-7 (page 63), however, that an overall proactive effect existed: the probability correct following the change of response was less than the probability correct following the first presentation of the R1 response. A parameter α_0 is therefore defined as the probability of attempting to encode the R2 response during the trial on which the change of response occurred, where $\alpha_0 < \alpha$. It is assumed that α_0 applies because the message ANSWER CHANGES appears on the screen. On trials where this message does not appear, α is assumed to apply in the usual way. Presumably the message sometimes induces the subject to pass by the new pairing, perhaps as a result of fear of confusion.

The extended model to be applied to Experiment II has three parameters not used in the model for Experiment I: δ, the probability of searching LTS when a new stimulus is tested; κ, the probability of tagging an o-code with the information that the response has been changed; and α_0, the probability of attempting to store on the trial when the response changes. Note that κ and α_0 apply only on the trial on which the response changes. When a search is made of LTS and no response is recovered and output, then the subject refrains from responding -- he does not guess.

<u>Mathematical Analysis</u>. For a given set of parameter values, the predictions of the model are generated in a manner quite similar to the method used for Experiment I. Appendix 4 presents the alterations in the iterative procedures used that enable us to predict the data for Experiment II. A natural next step would be the definition of an appropriate π^2 function, followed by a minimization routine. Unfortunately there is too much observed data for an attempt to minimize π^2 to succeed in a reasonable length of time, if all the data is considered simultaneously. Therefore, as a first step, we will fit the first-guess data only. The resultant parameter values, except for γ, will then be fixed. As a second step, the model will be applied to the second-guess data, but only γ will be estimated freely; the other parameters will retain the values giving the best fit to the first-guess data. The reason for estimating γ from the second-guess data is that γ is most sensitive to this data.

Let N_i be the total number of observations at the $\underline{i}$th trial; let O_i be the observed number of correct first-guesses at the $\underline{i}$th trial;

let Z_i be the observed number of intrusions (both old- and new-) at the <u>i</u>th trial. Let $P(c_i)$ be the predicted probability of a correct response at the <u>i</u>th trial; let $P(z_i)$ be the predicted intrusion probability at the <u>i</u>th trial (unconditional, and including both old- and new-intrusions). Then the following π^2 function is defined as a goodness-of-fit measure.

$$\pi^2(\alpha,\alpha_0,\beta,\sigma_H,\sigma_L,\sigma_I,\delta,\kappa,\gamma) =$$

$$\sum_{i=1}^{400}\left\{\frac{[N_iP(c_i) - O_i]^2}{N_iP(c_i)} + \frac{[N_iP(z_i) - Z_i]^2}{N_iP(z_i)} + \frac{[N_i(1-P(c_i)-P(z_i)) - (N_i-O_i-Z_i)]^2}{N_i(1-P(c_i) - P(z_i))}\right\} . \qquad \text{Eq. III-7}$$

The general comments made regarding Equation III-6 apply here also. N_i in the above π^2 function varies from 147 when i=1 to 122 when i=400. The number of degrees of freedom of π^2 in this instance is $(2 \times 400)-9 = 791$.

A grid search procedure was used to minimize π^2 over the possible sets of parameter values. When the parameters giving rise to the minimum value of π^2 were found, the second step of the estimation procedure was carried out. First a new π^2 function called π_1^2 was defined; π_1^2 was identical to π^2 except that all quantities were redefined to apply to the <u>second</u>-guess (thus N_i became the total number of intrusions, both new and old; etc.). All of the parameter values giving rise to the minimum value of π^2 were fixed except for the value of γ. Then $\pi_1^2(\gamma)$ was minimized. The minimum value of π_1^2 was 937.4 which occurred when $\gamma = 4.9$. This value of γ, along with the fixed values of the other parameters, was then used to recalculate π^2. The resultant value of π^2

was not appreciably higher than the minimum value based only on the first-guess data. As a result, we shall accept as "best" the predictions as generated by the parameter set with $\gamma = 4.9$. The values of the other parameters giving rise to the minimum π^2 are as follows: $\alpha = .94$, $\alpha_0 = .74$, $\beta = .25$, $\sigma_H = 45.1$, $\sigma_L = 1.25$, $\sigma_I = .117$, $\delta = .33$, and $\kappa = .30$. The minimum π^2 value was 872.6 (treated as a χ^2 this value would correspond to a level of significance between .05 and .01).

Predictions of the Model. The predictions of the model for the first-guess data are presented in Figures II-7, II-8, and Table II-8 (pages 65, 66, 69). The predictions, overall, are quite accurate; intrusion rates and correct guesses are predicted accurately both before and after the response changes, as a function of the number of reinforcements, and as a function of lag. The model predicts the overall proactive effect (due to the parameter α_0), and also the lack of a proactive effect as a function of the sequential history before the change of response (due to the strictly temporal search). There are several discrepancies that should be examined, however. First, note that the probability correct is considerably underpredicted after four reinforcements in the (10-10-10-10) condition (the discrepancy is .05 which is equivalent to a z-score of about 4.2). The model in general will underpredict after a large number of reinforcements for the following reason. Because the search is strictly temporally ordered, there is always a minimum average number of intrusions which occur before the c-code is ever examined, no matter how well the c-code is stored. Thus there is a ceiling for the probability correct at a given lag, as long as new items are continually introduced. In Experiment I some items were given up to 7 reinforcements, but the lags in these cases were large, and the probability correct never got near enough to the arbitrary ceiling for discrepancies to occur. In the present experiment, there are only four consecutive reinforcements before the response changes; as a result only a single discrepant point occurs. Thus it is not safe to conclude without further experimentation with greater numbers of reinforcements that the model definitely fails in predicting such a ceiling effect. (However, we will shortly examine evidence of a rather different character which will definitely show that the strictly ordered search hypothesis is in error.) A second discrepancy of the predictions occurs in the intrusion rates following the change of response, especially old-intrusion rates. Even though a proactive effect is not predicted for the probability correct, old-intrusions are predicted to rise as the amount of learning concerning R1 increases. The data, however, show a quite stable old-intrusion rate over conditions.

The above points notwithstanding, the predictions for the first-guess data are quite accurate. There is another statistic which bears this out. The model predicts that the new-intrusion rate will increase during the session, since more and more items are available to give rise to new-intrusions. This is easiest to check for new items. The observations and predictions are given in Figure II-9 (page 71). The overall level of the predictions in the Figure is governed by the parameter δ, and its accuracy is not surprising; however, the form of the predicted increase is quite close to that observed. The meaningfulness of this statistic is difficult to determine. The overall reduction in intrusion rates (reflected by δ) is assumed to occur because new items are recognized as such; it might seem logical that this recognition process would be a function of the duration of the session. It is possible to argue, however, that recognition via extremely salient stimulus characteristics is not appreciably affected by the number of stimuli input. This question should prove susceptible to further experimental research; for the present, it is not unreasonable to accept the second hypothesis above, an hypothesis in accord with the model.

Before turning to the second-guess results it would be instructive to consider the values attained by several of the parameters. It has been suggested earlier that the value of σ_I should be reflective of the amount of inter-stimulus generalization in the experiment. Since Experiment I utilized highly confusable consonant trigrams, and Experiment II utilized words, the value of σ_I should be smaller in the second experiment. The values attained were in the expected direction (.18 and .117 respectively). At first glance, the value attained by σ_H, 45.1, seems far too high; for example, this value would lead to predictions that the probability correct at a lag of near 300 would be as high as .30 (depending upon the condition). Fortunately this prediction can be roughly checked in the data since there were a few instances of very long lags. For example, stimulus number 10 (in the trial sequence of Appendix 2) was given successive reinforcements on trials 13, 39, and 389. The predicted probability correct for trials 39 and 389 was 44.6 and 28.5 respectively. The observed values on these trials were 42.1 and 42.4 respectively. Thus, the observed values were even higher than those predicted. Similarly, stimulus number 47 was given its final two reinforcements on trials 77 and 380. The predicted values for these trials were 35.4 and 26.3; the observed values were 35.3 and 42.3.

These results indicate that the high value of σ_H estimated in the present case was quite appropriate.

The second-guess predictions are presented in Table II-9 and in Figure II-10 (pages 73, 72). Figure II-10 gives the probability correct in the top panel and the overall intrusion rate in the lower panel, both following first-guess <u>new</u>-intrusions. In addition, the predictions in the lower panel are conditional upon a second-guess error. In both panels the fit is fairly accurate. The high intrusion rates predicted occur because $\gamma = 4.9$, considerably lowering the decision criterion for output of second-guess responses. A very high intrusion rate is predicted even for new items, items not previously presented. The model predicts this effect because the rates shown are conditional upon a first-guess error; an error implies that during the first-guess the subject did not recognize that the item was new, and made a decision to search LTS. Under these circumstances, a second-guess search will also be made, and since the stimulus being tested is new, this search will quite often result in intrusions (there is no c-code in LTS to lower the intrusion probability). Table II-9c gives the breakdown of the predictions in the lower panel of the figure, i.e., it gives the second-guess <u>old</u>-intrusion probability for the major item-types, following new-intrusions on the first-guess (the combined old- and new-intrusion rates were given in the figure). The predictions for these cases seem quite accurate, lending support to the hypothesis that o-codes and c-codes are quite similar, even with respect to their probability of being given following an extraneous intrusion.

Tables II-9a and II-9b give the second-guess predictions for the matrix item-types, following a new-intrusion on the first-guess. The first comment to be made is that the predictions in these tables are consistently high; this results from a failing of the model to be discussed shortly (under-predictions following old-intrusions on the first guess); if the second-guess data following first-guess old-intrusions were not part of the π^2 minimization, then these data would be fit more closely. Qualitatively, the effects predicted are observed with several minor exceptions. For example, in Table II-9a, a maximum probability correct is predicted at a second lag of 5: this prediction is observed if one ignores the observation at (1,1). In fairness to the model it should be pointed out there are very few observations in the (0,1) and (1,1) conditions. Similarly, in Table II-9b, the predicted increase in second-guess new-intrusions as a function of the second lag is observed if one eliminates the (0,1) and (1,1) points. More serious are the deviant predictions for second-guess old-intrusions after the second lag. The old-intrusion rate is predicted to rise as the second lag increases; this is observed for first lags of 1, 4, and 10, but just the opposite is seen for a first lag of 0. This misprediction could be rectified by assuming that the zero-lag is a special case that results in a very high probability of coding the old-response as being wrong. In the previous model, this coding only occurs after a non-null-state retrieval.

As a whole the predictions discussed so far are quite accurate. We turn now to a prediction which conclusively demonstrates that the assumption of a strictly temporally ordered LTS search is not adequate. These predictions are the counterpart to the observations presented in Table II-10 (page 76). The predictions were not given there because they are so extremely discrepant from the observations. The observed probability of a second-guess correct response following a first-guess old-intrusion is quite high -- about .30. Without giving the predictions cell by cell, we can state that the predicted probability correct varies between .02 and .05, depending upon the condition. The model predicts such low probabilities following first-guess old-intrusions because a c-code will always be examined before an o-code, if both are in the examination subset. This occurs because the LTS search is strictly temporal, and the c-code is always more recent than the o-code. If an old-intrusion first-guess is given, then it is certain that the c-code is either not present or has been bypassed in the search. A c-code present in LTS is not bypassed often, but when it is, it is almost always a lo-code; thus the probability of recovering it correctly during second-guessing is very low. The predicted second-guess intrusion probabilities following first-guess old-intrusions are also fairly deviant. Because the probability correct is predicted to be quite low, the intrusion predictions are quite high, about .45.

These failures of the predictions of the model make it clear that the assumption of a strictly temporal LTS search must be altered. The precise manner of alteration, which will still allow prediction of the previous observations, is not trivial and will be discussed later.

The failure of the temporal search assumption would make it presumptious to extend the present model to the latency results. Nevertheless, there are a number of theoretical remarks that may be made concerning the observed latencies. A simple model which can be used as a base for speculations holds that items retrieved from STS have a relatively short mean latency; items retrieved from LTS have a latency proportional to the number of codes examined before the response is output. The observed increase of correct response latency with lag can be explained either by considerations of recovery from STS (which decreases with lag) or by a partially temporal LTS search. The decrease in correct response latency with the number of reinforcements cannot be explained by a strictly temporal search; however, a search that examines codes in an order partially dependent upon the code's strength can predict this effect nicely. As the number of reinforcements increase, more and more of the c-codes stored will be hi-codes; hi-codes will tend to be examined earlier in the search than lo-codes because of their greater strength and hence will result in lower latencies. Previous studies have reported latency decreases with increases in reinforcements (i.e., Rumelhart, 1967), but responding in these studies was required on every trial. The results could therefore be explained as the result of averaging guesses and retrievals. Rumelhart also found that the latency of correct responses decreased after an item's terminal error, a result not explicable by guessing considerations. The effect is predicted quite easily by the present model, however. The same assumption regarding order of search can help explain why correct response latencies after the change of response are higher than before the change: The o-code will be examined occasionally before the c-code; even when the o-code response is inhibited, the latency of giving the c-code response will be lengthened by the prior consideration of the o-code. At first glance, it might appear that an occasional prior consideration of an o-code will not significantly alter the latency predictions, but this is not so. The predicted mean number of i-codes in the potential examination-subset is only 5.0 for the present model, even on the last trials of the session. The mean number actually examined prior to a correct response is considerably less than this figure, perhaps less than 1.0. In these circumstances, only a small proportion of o-codes additionally examined prior to emission of the correct response will greatly affect the predicted latency of such a correct response.

That intrusion latencies would be larger than correct response latencies would not be unexpected even in the strictly temporal search model. The model in which the search order depends upon the strength of the codes, however, does not only explain this result, but also why the latencies of old-intrusions are markedly smaller than those of new-intrusions (since the strength of i-codes is much less than that of o-codes, the o-codes will be examined earlier in the search). The fact that latencies of old-intrusions are greater than those of correct responses, even though in most cases there is a higher proportion of high strength codes for o-codes than c-codes, indicates that there is at least some temporal component to the search.

In the absence of a specific model, we will not discuss the latency results further. The major import of these results is that the order of the LTS search through the examination-subset must be only partially temporally ordered, and partially dependent upon the strength of the codes in the subset. This is the same conclusion arrived at through a consideration of the probability of a correct second-guess following an old-intrusion on the first guess. We might turn then to a discussion of the necessary features of such a model.*

Extensions of the Model. The most reasonable extension of the model lets the order of search through the examination-subset depend upon both the strength and temporal position of the codes. However, as soon as the strictly temporal search is altered, a proactive effect will be predicted which depends upon the amount of learning of the R1 response. That is, in the extended model the proportion of times the o-code is encountered prior to the c-code will be greater the more often the o-code is stored, and will be greater the larger the strength of the o-code. Similarly, the number of old-intrusions should be markedly affected by the level of learning of the R1 response, but neither of these predictions is observed. Apparently what is needed in the model is a mechanism by which well-known o-codes are marked as being wrong (old), but in which the number and strength of the unmarked o-codes remain very nearly constant over a wide range of reinforcement histories. The formulation of such a process would undoubtedly entail the use of several new parameters, but several parameters of the current model could very probably be eliminated, namely α_0 and κ. The precise formulation of an appropriate model to deal with the change-of-response data is beyond the scope of the present report; it must await further research to verify the results found, and to extend the range of variables studied. The major change

*The entire question of order of search can probably be settled unconditionally by engaging in further research in which each stimulus has a unique response assignment. Then all intrusions could be precisely placed temporally.

of response result, that proactive item interference does not depend upon the degree of learning of the R1 response, is certainly surprising in the light of the list structure results, and from the point of view of two-factor interference theory. This alone is sufficient reason for engaging in further research dealing with individual item-interference.

Concluding Discussion

We may summarize the major results of Experiment I as follows. First, it was found that the second-guessing probability could be considerably above chance even when responses ranked after the first choice were correct at the chance level. This result was interpreted as implying that the subjects used a retrieval strategy which output the first acceptable response recovered in the memory search. If this strategy is adopted, then the subject will give the recovered response as his first-ranking and guess for the remaining three rankings. Thus only the first-ranking will be above chance. Second-guessing, on the other hand, is based upon the result of an additional search of memory and may therefore be above chance. Second, it was found that performance in a continuous task decreased toward the chance level as the study-test interval became very large; in addition, when the lag between reinforcements was large, learning curves did not asymptote at a probability correct of 1.0, but rather seemed to stabilize at some intermediate value related to the size of the lag between reinforcements. These results demonstrated that any model which assumes a long-term absorbing state is not an appropriate representation of the memory process for tasks of the present type. In order to predict the above results, it was proposed that codes of varying strength are stored in LTS, and that the probability of retrieval at test is dependent upon the age and strength of the stored codes. This model was able to predict the learning, forgetting, and second-guessing data quite accurately. Third, it was found that the amount of forgetting at a given lag was dependent upon how well-known were the intervening items. The model predicted this result because the "age" of an item was made dependent upon the number of new codes that were stored during the intervening period.

The primary empirical results of Experiment II were concerned with proactive interference. It was found for both the probability of a correct response and the probability of an intrusion that an overall proactive effect was present. The magnitude of the effect, however, was not dependent upon the reinforcement and lag history prior to the change of response. The model predicted this proactive effect for probability correct because it assumed a strictly temporally ordered memory search. However, it was found that the probability of correctly second-guessing following an old-intrusion was about .30, markedly higher than the predictions of about .05. This latter finding demonstrated that the memory search could not be strictly temporally ordered; it was argued that search order is dependent upon the strength of codes as well as their age. This hypothesis was given further support by the analysis of response latencies. First, the latency of a correct response decreased with the number of reinforcements; second, the latency of a correct response was greater following the change of response than prior to the change. These latency results would be expected if codes of greater strength tended to be examined earlier in the memory search. Although this extension of the model seems quite natural, it results in the prediction that proactive effects will depend upon the reinforcement and lag history prior to the change of response. Since this prediction was not confirmed, further extensions of the model were suggested which would handle the observations.

Because an important feature of the storage and retrieval model was the prediction of intrusions, Experiment II was designed to examine intrusion probabilities over a wide range of conditions. In general, the model predicted the intrusion probabilities quite accurately. Two findings are especially noteworthy. First, the intrusion probabilities during second-guessing were found to be considerably higher than those during first-guessing; this result was taken to imply that the criterion for output of recovered responses was considerably lowered during second-guessing. Second, the intrusion probability when a new stimulus was presented for test was very much lower than that observed for previously presented items. This result reflects a recognition process in which certain new stimuli are recognized as being new; when presented stimuli with very salient characteristics do not trigger a recognition response in the expected location, it is assumed that a decision is made to cease further memory search. However, if a decision is made to search LTS, then a second-guess following an error should result in a very high intrusion probability, and this was also observed.

Taken as a whole, the predictions of the model were quite accurate. The model proved capable of dealing quantitatively and simultaneously with a wide variety of data, including lag, number of reinforcements, second-guessing performance, intrusion rates on first- and second-guessing, and change of response phenomena. The primary way in which this model differed from its predecessors was its emphasis upon an ordered search through a small subset of the codes stored in LTS. The value of such a process was confirmed by the analysis of the data; in fact, the analysis gives considerable support to the theory outlined in the first chapter of this report.

APPENDIX 1

Column a = trial number
Column b = stimulus number
Column c = number of reinforcements of current stimulus

a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	1	0	45	24	0	90	30	0	135	36	1	180	58	0
2	2	0	46	23	1	91	26	3	136	36	5	181	41	0
3	3	0	47	19	2	92	21	5	137	20	4	182	37	6
4	4	0	48	21	1	93	56	0	138	37	1	183	39	5
5	3	1	49	13	3	94	27	2	139	11	5	184	42	0
6	5	0	50	16	4	95	56	1	140	32	1	185	19	7
7	6	0	51	25	0	96	24	1	141	12	7	186	20	6
8	7	0	52	25	1	97	30	1	142	22	1	187	38	4
9	2	1	53	51	0	98	31	0	143	38	0	188	13	7
10	8	0	54	19	3	99	31	3	144	27	6	189	32	3
11	9	0	55	52	0	100	57	0	145	37	2	190	58	1
12	10	0	56	26	0	101	57	1	146	35	2	191	57	2
13	5	1	57	16	5	102	26	4	147	23	1	192	41	1
14	11	0	58	20	1	103	29	6	148	24	2	193	21	6
15	12	0	59	21	2	104	30	2	149	39	0	194	56	2
16	13	0	60	13	4	105	27	3	150	40	0	195	40	2
17	12	1	61	19	4	106	60	0	151	25	2	196	31	2
18	14	0	62	53	0	107	61	0	152	37	3	197	38	5
19	12	2	63	53	1	108	62	0	153	51	1	198	24	3
20	15	0	64	11	2	109	59	0	154	38	1	199	39	6
21	12	3	65	52	1	110	20	3	155	63	0	200	43	0
22	16	0	66	54	0	111	30	3	156	39	1	201	29	7
23	12	4	67	14	1	112	26	5	157	35	3	202	43	1
24	17	0	68	19	5	113	32	0	158	16	6	203	41	2
25	12	5	69	26	1	114	11	4	159	37	4	204	43	2
26	17	1	70	21	3	115	33	0	160	53	2	205	44	0
27	13	1	71	13	5	116	27	4	161	20	5	206	43	3
28	17	2	72	27	0	117	14	2	162	32	2	207	42	1
29	16	1	73	28	0	118	30	4	163	39	2	208	43	4
30	17	3	74	28	1	119	59	1	164	11	6	209	20	7
31	18	0	75	28	2	120	15	1	165	38	2	210	43	5
32	17	4	76	29	0	121	34	0	166	37	5	211	26	6
33	19	0	77	55	0	122	55	1	167	14	3	212	44	1
34	17	5	78	29	1	123	34	1	168	35	4	213	32	4
35	20	0	79	26	2	124	35	0	169	28	2	214	41	3
36	16	2	80	29	2	125	30	5	170	39	3	215	33	1
37	21	0	81	21	4	126	36	0	171	52	2	216	40	3
38	13	2	82	29	3	127	27	5	172	40	1	217	45	0
39	11	1	83	27	1	128	36	1	173	54	1	218	14	4
40	19	1	84	29	4	129	18	1	174	28	3	219	44	2
41	12	6	85	19	6	130	36	2	175	64	0	220	15	2
42	22	0	86	29	5	131	37	0	176	38	3	221	46	0
43	16	3	87	20	2	132	36	3	177	39	4	222	34	2
44	23	0	88	13	6	133	17	6	178	55	2	223	47	0
			89	11	3	134	36	4	179	35	5	224	30	6

APPENDIX 1 (CONT.)

a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
225	41	3	270	48	5	315	33	2	360	63	1	405	64	1
226	44	3	271	40	5	316	75	0	361	86	0	406	69	2
227	43	6	272	45	5	317	14	6	362	87	0	407	80	0
228	45	1	273	28	4	318	70	1	363	53	4	408	71	6
229	55	3	274	49	1	319	49	6	364	75	2	409	62	1
230	47	1	275	47	6	320	15	3	365	52	4	410	74	6
231	18	2	276	46	5	321	40	7	366	74	2	411	81	0
232	46	1	277	52	3	322	34	3	367	14	7	412	72	6
233	44	4	278	54	2	323	59	3	368	54	3	413	68	2
234	17	7	279	35	6	324	76	0	369	71	3	414	1	2
235	42	2	280	53	3	325	30	7	370	72	3	415	33	3
236	41	5	281	49	2	326	1	1	371	61	1	416	75	4
237	47	2	282	55	4	327	43	7	372	76	2	417	2	2
238	36	6	283	37	7	328	4	1	373	74	3	418	70	2
239	42	2	284	42	4	329	55	5	374	88	0	419	49	7
240	44	5	285	66	0	330	50	6	375	47	7	420	15	4
241	32	5	286	67	0	331	18	3	376	7	1	421	59	4
242	22	2	287	48	6	332	42	6	377	8	1	422	34	4
243	46	2	288	49	3	333	71	0	378	55	6	423	76	4
244	47	3	289	45	6	334	60	1	379	35	7	424	3	2
245	27	7	290	32	7	335	72	0	380	71	4	425	77	0
246	40	4	291	58	2	336	36	7	381	67	1	426	4	2
247	28	3	292	21	7	337	73	0	382	72	4	427	77	1
248	24	4	293	46	6	338	6	1	383	74	4	428	55	7
249	59	2	294	56	3	339	73	1	384	66	1	429	77	2
250	45	3	295	49	4	340	44	6	385	89	0	430	50	7
251	47	4	296	40	6	341	75	1	386	9	1	431	77	3
252	25	3	297	38	6	342	22	3	387	48	7	432	18	4
253	41	6	298	24	5	343	76	1	388	75	3	433	77	4
254	46	3	299	31	3	344	30	5	389	45	7	434	5	2
255	51	2	300	39	7	345	71	3	390	58	3	435	77	5
256	65	0	301	57	3	346	72	1	391	71	5	436	6	2
257	16	7	302	49	5	347	23	4	392	74	5	437	82	0
258	47	5	303	68	0	348	24	6	393	46	7	438	83	0
259	42	3	304	50	0	349	30	6	394	72	5	439	73	2
260	48	0	305	69	0	350	28	7	395	76	3	440	44	7
261	45	4	306	50	1	351	25	4	396	56	4			
262	48	1	307	42	5	352	74	0	397	38	7			
263	32	6	308	50	2	353	41	7	398	24	7			
264	48	2	309	70	0	354	51	3	399	31	4			
265	46	4	310	50	3	355	69	1	400	90	0			
266	48	3	311	26	7	356	71	2	401	57	4			
267	49	0	312	50	4	357	42	7	402	10	1			
268	48	4	313	68	1	358	72	2	403	78	0			
269	14	5	314	50	5	359	74	1	404	79	0			

APPENDIX 2

Column a = trial number Column b = stimulus number
Column c = 0 for study of first response, 1 for second
Column d = number of reinforcements of latest response

a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
1	1	0	0	45	45	1	1	90	51	1	0	135	21	1	2	180	28	0	1
2	2	0	0	46	14	1	0	91	48	1	0	136	22	0	3	181	22	1	3
3	3	0	0	47	43	1	1	92	15	1	2	137	60	1	0	182	64	1	0
4	4	0	0	48	15	0	2	93	48	1	1	138	54	0	0	183	26	0	2
5	2	0	1	49	12	0	4	94	52	1	0	139	18	1	2	184	26	0	3
6	5	0	0	50	47	0	0	95	18	0	2	140	54	1	0	185	25	1	2
7	6	0	0	51	47	1	0	96	46	1	1	141	23	0	2	186	63	0	0
8	4	0	1	52	13	1	1	97	19	0	2	142	23	0	2	187	27	0	1
9	7	0	0	53	42	0	0	98	19	0	3	143	60	1	1	188	63	1	0
10	8	0	0	54	16	1	0	99	20	0	1	144	58	1	1	189	24	1	2
11	9	0	0	55	17	0	0	100	49	1	1	145	24	0	0	190	63	1	1
12	1	0	1	56	17	0	1	101	21	0	0	146	54	1	1	191	28	1	0
13	10	0	0	57	14	1	1	102	21	0	1	147	22	1	0	192	65	0	0
14	3	0	1	58	42	1	0	103	22	0	0	148	55	0	0	193	64	1	1
15	12	0	0	59	15	0	3	104	53	0	0	149	56	0	0	194	65	1	0
16	12	0	1	60	44	0	0	105	53	1	0	150	55	1	0	195	26	0	4
17	5	0	1	61	44	1	0	106	18	0	3	151	25	0	0	196	67	0	0
18	45	0	0	62	41	1	2	107	58	0	0	152	25	0	1	197	67	1	0
19	13	0	0	63	13	1	2	108	15	1	3	153	23	1	0	198	27	0	2
20	45	1	0	64	42	1	1	109	19	1	0	154	56	1	0	199	66	0	0
21	8	0	1	65	16	1	1	110	20	0	1	155	24	0	1	200	65	1	1
22	14	0	0	66	46	0	0	111	53	1	1	156	56	1	1	201	66	1	0
23	14	0	1	67	17	1	0	112	50	1	1	157	62	0	0	202	28	1	1
24	43	0	0	68	14	1	2	113	21	1	0	158	22	1	1	203	76	0	0
25	3	0	2	69	49	0	0	114	22	0	1	159	61	0	0	204	29	0	0
26	15	0	0	70	15	1	0	115	51	1	1	160	61	1	0	205	64	1	2
27	12	0	2	71	46	1	0	116	57	0	0	161	55	1	1	206	26	0	5
28	40	0	0	72	44	1	1	117	18	1	0	162	61	1	1	207	70	0	0
29	40	1	0	73	18	0	0	118	58	1	0	163	25	1	0	208	76	1	0
30	13	0	1	74	49	1	0	119	52	1	1	164	23	1	1	209	27	0	3
31	40	2	0	75	13	1	3	120	19	1	1	165	56	1	2	210	30	0	0
32	16	0	0	76	16	1	2	121	20	0	3	166	54	1	2	211	30	0	1
33	6	0	1	77	47	1	1	122	59	0	0	167	24	1	0	212	66	1	1
34	14	0	2	78	17	1	1	123	59	1	0	168	62	1	0	213	29	1	2
35	14	0	3	79	51	0	0	124	21	1	1	169	22	1	2	214	29	0	1
36	43	1	0	80	48	0	0	125	22	0	2	170	62	1	1	215	24	1	3
37	15	0	1	81	15	1	1	126	60	0	0	171	26	0	0	216	69	0	0
38	12	0	3	82	50	0	0	127	57	1	0	172	26	0	1	217	26	0	6
39	10	0	1	83	52	0	0	128	18	1	1	173	61	1	2	218	70	1	0
40	41	0	0	84	18	0	1	129	23	0	0	174	25	1	1	219	76	1	1
41	13	1	0	85	19	0	0	130	23	0	1	175	23	1	2	220	27	1	0
42	41	1	0	86	10	0	1	131	19	1	2	176	27	0	0	221	69	1	0
43	16	0	1	87	50	1	0	132	20	0	4	177	64	0	0	222	30	0	2
44	41	1	1	88	20	0	0	133	57	1	1	178	24	1	1	223	30	0	3
				89	17	1	2	134	59	1	1	179	28	0	0	224	67	1	1

APPENDIX 2 (CONT.)

a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
225	29	1	0	270	31	1	0	315	88	1	0	360	83	0	3
226	31	0	0	271	75	0	0	316	87	0	0	361	94	1	0
227	69	1	1	272	75	1	0	317	38	0	2	362	93	1	1
228	70	1	1	273	33	1	0	318	38	0	3	363	84	0	0
229	32	0	0	274	75	1	1	319	86	1	1	364	89	1	2
230	32	0	1	275	78	0	0	320	36	1	2	365	11	1	2
231	27	1	1	276	77	1	1	321	87	1	0	366	39	1	0
232	68	0	0	277	78	1	0	322	39	0	0	367	92	0	0
233	68	1	0	278	34	1	0	323	87	1	1	368	74	0	0
234	30	1	0	279	78	1	1	324	35	0	5	369	92	1	0
235	72	0	0	280	35	0	1	325	37	1	0	370	85	1	0
236	29	1	1	281	31	1	1	326	88	1	1	371	83	1	0
237	31	0	1	282	80	0	0	327	83	0	0	372	94	1	1
238	71	0	0	283	80	1	0	328	89	1	0	373	74	1	0
239	68	1	1	284	33	1	1	329	38	1	0	374	84	0	1
240	72	1	0	285	73	1	1	330	95	0	0	375	71	1	2
241	32	1	0	286	36	0	0	331	95	1	0	376	11	1	1
242	27	1	2	287	36	0	1	332	93	0	0	377	39	1	1
243	71	1	0	288	70	1	2	333	39	0	1	378	91	0	0
244	30	1	1	289	80	1	1	334	34	1	4	379	91	0	1
245	71	1	1	290	34	1	1	335	35	0	6	380	47	1	2
246	72	1	1	291	35	0	2	336	37	1	1	381	85	1	1
247	29	1	2	292	31	1	2	337	93	1	0	382	83	1	1
248	31	0	2	293	81	0	0	338	83	0	1	383	90	0	1
249	77	0	0	294	33	1	2	339	89	1	0	384	74	1	1
250	77	1	0	295	81	1	0	340	38	1	1	385	84	1	0
251	33	0	0	296	73	1	2	341	11	0	0	386	89	1	3
252	32	1	1	297	82	0	0	342	11	0	1	387	11	1	2
253	79	0	0	298	36	1	0	343	95	1	1	388	39	1	2
254	34	0	0	299	82	1	0	344	39	0	2	389	10	1	0
255	34	0	1	300	34	1	2	345	89	1	1	390	91	0	2
256	30	1	2	301	81	1	1	346	36	1	3	391	91	0	3
257	23	1	3	302	35	0	3	347	37	1	2	392	85	1	2
258	73	0	0	303	37	0	0	348	82	1	2	393	83	1	2
259	31	0	3	304	88	0	0	349	83	0	2	394	71	1	3
260	73	1	0	305	38	0	0	350	94	0	0	395	92	1	1
261	27	1	3	306	38	0	1	351	38	1	2	396	84	1	1
262	33	0	1	307	86	0	0	352	81	1	3	397	94	1	2
263	32	1	0	308	86	1	0	353	11	0	2	398	47	1	2
264	79	1	0	309	36	1	1	354	11	0	3	399	10	1	1
265	34	0	2	310	82	1	1	355	39	0	3	400	10	1	2
266	79	1	1	311	34	1	3	356	90	0	0				
267	34	0	3	312	81	1	2	357	90	1	0				
268	30	1	3	313	35	0	4	358	85	0	0				
269	37	0	0	314	37	0	1	359	85	0	1				

APPENDIX 3

ITERATIVE PROCEDURES FOR CALCULATING PREDICTIONS FOR EXPERIMENT I

Let $b_{n,j}$ be the probability that the item being tested is in STS, at lag j.
Let $c_{n,k}$ be the probability correct on trial n, guess k.
Let $e_{n,k}$ be $1.0 - c_{n,k}$.

Let $\overline{\Omega}_n$ be the average state of memory at trial n. $\overline{\Omega}_n$ is equivalent to the status of the following five vectors, each of length n:

1) $code_i$ is the probability that a new code was stored on trial i.
2) buf_i is the probability that the item presented on trial i entered STS (but not the null-state).
3) hic_i is the probability that a hi-code for the item presented on trial i is temporally placed in memory at trial i.
4) loc_i is the probability that a lo-code for the item presented on trial i is temporally placed in memory at trial i.
5) q_i is a dummy variable; equals zero only if the stimulus tested on trial i is later tested on a trial previous to n, else equals one.

We now show how to derive $\overline{\Omega}_n$ as a function of $\overline{\Omega}_{n-1}$. Assume we have $\overline{\Omega}_{n-1}$.

We need the following definitions.

$CR1_n$ is the probability of a correct response recovery given a first-guess LTS search, on trial n.
$CR2_n$ is the same for a second-guess search.
$IN1_n$ is the probability of an incorrect response recovery given a first-guess LTS search, on trial n.
$IN2_n$ is the same for a second-guess search.
$CE1_n = 1-CR1_n - IN1_n$.
$CE2_n = 1-CR2_n - IN2_n$.
SC_j is the probability of a correct recovery in an LTS search given that the search has proceeded as far as the jth trial. (Note: the search proceeds backwards, from trial n to trial 1.)
SI_j is the same for incorrect recoveries from LTS during the search.
Let j* be the trial number of the c-code.
Let fpi be the probability of an incorrect intrusion between trials n and j*.

Let $P(Z_k)$ be the probability that a code of type k is in the examination subset, where k= H,L, or I, depending upon the code type.
Let $P(P_k)$ be the probability that an examined code of type k gives rise to the response encoded, where k= H,L, or I, depending upon the code.

The status of a search of memory is defined by (SC_j, SI_j). This vector may be calculated recursively. If $j-1 \neq j^*$ then

$$SI_{j-1} = SI_j + q_j(1-SC_j-SI_j)(hic_j + loc_j)P(Z_I)P(P_I)(3/4).$$

$$SC_{j-1} = SC_j + q_j(1-SC_j-SI_j)(hic_j + loc_j)P(Z_I)P(P_I)(3/4).$$

But if $j-1 = j^*$ then,

$$SI_{j^*} = SI_j$$

$$SC_{j^*} = SC_j + (1-SC_j-SI_j)(hic_j P(Z_H)P(P_H) + loc_j P(Z_L)P(P_L)).$$

In the above recursions, the age of an item at trial j is required (in $P(Z_k)$). The age is calculated as follows:

$$age_j = \sum_{i=j}^{i=n} code_i.$$

As the result of the recursion, we have (SC_1, SI_1). Then $CR1_n = SC_1$; $IN1_n = SI_1$.

We now have,

$$c_{n,1} = b_{n,j} + (1-b_{n,j})(CR1_n + IN1_n/4 + CE1_n/4), \text{ where } b_{n,j} = (buf_{n-j+1})\alpha^{n-j}.$$

Before the second-guess search predictions may be calculated, adjustment must be made for the selection effect due to the first-guess error. Hence, we must temporarily alter the proportions of hi- and lo-c-codes stored.

$$HIC_{j^*} = \left\{(1-b_{n,n-j^*+1})(hic_{j^*})(fpi + [1-(4/3)(fpi)][1-P(Z_H)][3/4]\,)\right\} / e_{n,1}.$$

$$LOC_{j^*} = \left\{(1-b_{n,n-j^*+1})(loc_{j^*})(fpi + [1-(4/3)(fpi)][1-P(Z_L)][3/4]\,)\right\} / e_{n,1}.$$

The second-guess recursion now proceeds identically to the first-guess recursion, except that the quantities above are substituted for hic_{j^*}, loc_{j^*}. The result is $CR2_n$, $IN2_n$, and $CE2_n$. Then we have,

$$c_{n,2} = (1-c_{n,1})(CR2 + IN2_n/3 + CE2_n/3).$$

This concludes the predictions on the <u>n</u>th trial; to calculate $\bar{\Omega}_n$, however, we must complete the <u>n</u>th trial of the five vectors making up the state of memory.

Let $Y = (1-b_{n,n-j^*+1})$; Let $W = CR1_n + IN1_n/4 + e_{n,1}(CR2_n + IN2_n/3)$.

Then,

$$code_n = Y(1-W)\alpha.$$

$$hic_n = Y(W+[1-W][\alpha/2]).$$

$$loc_n = Y(1-W)\alpha/2.$$

$$q_{j*} = 0.$$

$$\mathrm{buf}_n = 1 - Y + Y(1-W)\alpha.$$

The above five equations transform $\overline{\Omega}_{n-1}$ into $\overline{\Omega}_n$. The iterative process then continues until the 439 trials are predicted. The boundary conditions on the above process, and special cases such as zero-lag, are not given here: they are straightforward, and their presentation merely increases the terminology needed.

APPENDIX 4

ITERATIVE PROCEDURES FOR CALCULATING PREDICTIONS FOR EXPERIMENT II

The iterations used for Experiment II are very close in character to those for Experiment I, and little purpose is served by repeating them here in full detail. Instead, we present only the equations which normalize the proportions of hi- and lo-codes for selection effects prior to second-guessing.

Before the answer changes, all intrusions are new, hence, there are just two conditions: HIC_{j*} represents adjusted hi-codes; LOC_{j*} represents adjusted lo-codes.

$$HIC_{j*} = \left\{Y(hic_{j*})(fpi + [1-(26/25)(fpi)][1-P(Z_H)][IN1_n(25/26)-fpi])\right\} / e_{n,1}.$$

$$LOC_{j*} = \left\{Y(loc_{j*})(fpi + [1-(26/25)(fpi)][1-P(Z_L)][IN1_n(25/26)-fpi]\right\} / e_{n,1}.$$

After the answer changes we must consider two possibilities: the intrusion could have been old- or new-. We denote the adjusted probabilities with primes (') if there was a new-intrusion; we denote the adjusted probabilities with quotes (") if there was an old-intrusion. Then,

$$'HIC_{j*} = \left\{Y(hic_{j*})(fpi + [1-(26/25)(fpi)][1-P(Z_H)][(1-f1)+f1(1-c2)(1-f2)]\right\} / n_{n,1}.$$

$$'LOC_{j*} = \left\{Y(loc_{j*})(fpi + [1-(26/25)(fpi)][1-P(Z_L)][(1-f1)+f1(1-c2)(1-f2)]\right\} / n_{n,1}.$$

$$"HIC_{j*} = \left\{Y(hic_{j*})(1-[26/25]fpi)(1-P[Z_H])(f1)(c2)\right\} / e_{n,1}.$$

$$"LOC_{j*} = \left\{Y(loc_{j*})(1-[26/25]fpi)(1-P[Z_L])(f1)(c2)\right\} / e_{n,1}.$$

The above equations use several definitions not used in Appendix 3.

Set $Y = (1 - b_{n,n-j*+1})$.

Let $n_{n,1}$ represent the probability of a new intrusion on the first-guess on trial n.

Let $e_{n,1}$ represent the probability of an old intrusion on the first-guess on trial n.

Let c2 represent the probability of giving the R1 response after examining the ȯ-code.

Let 1-f1 be the probability of emitting a new intrusion as a result of examining a i-code temporally between the c-code and the ȯ-code.

Let 1-f2 be the probability of emitting a new intrusion as a result of examining a i-code temporally older than the ȯ-code.

Then the above equations give the correction for selection effects. The remaining calculations are straightforward, similar to those given in Appendix 3, and are therefore not presented.

REFERENCES

Atkinson, R. C., Brelsford, J. W., and Shiffrin, R. M. Multi-process models for memory with applications to a continuous presentation task. Journal of Mathematical Psychology, 1967, 4, 277-300.

Atkinson, R. C., and Estes, W. K. Stimulus sampling theory. In R. D. Luce, E. Galanter, and R. R. Bush (Eds.), Handbook of mathematical psychology, Vol. 2. New York: Wiley, 1963, 121-268.

Atkinson, R. C., and Shiffrin, R. M. Mathematical models for memory and learning. Tech. Rep. No. 79, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965. (To be published in D. P. Kimble (Ed.), The proceedings of the third conference on learning, remembering and forgetting. New York: New York Academy of Science).

Atkinson, R. C., and Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory, Vol. 2. New York: Academic Press, 1968, in press.

Binford, J. R., and Gettys, C. Nonstationarity in paired-associate learning as indicated by a second guess procedure. Journal of Mathematical Psychology, 1965, 2, 190-195.

Bjork, A. Learning and short-term retention of paired-associates in relation to specific sequences of interpresentation intervals. Tech. Rep. No. 106, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1966.

Bower, G. H. Notes on a descriptive theory of memory. In D. P. Kimble (Ed.), _The organization of recall_. New York: New York Academy of Science, 1967, 112-185.

Calfee, R. C., and Atkinson, R. C. Paired-associate models and the effects of list length. _Journal of Mathematical Psychology_, 1965, _2_, 254-265.

DaPolito, F. J. Proactive effects with independent retrieval of competing responses. Unpublished doctoral dissertation, Indiana University, 1966.

Estes, W. K. Statistical theory of distributional phenomena in learning. _Psychological Review_, 1955, _62_, 369-377. (a)

Estes, W. K. Statistical theory of spontaneous recovery and regression. _Psychological Review_, 1955, _62_, 145-154. (b)

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), _Psychology: a study of a science_, Vol. 2. New York: McGraw-Hill, 1959, 380-491.

Estes, W. K. All-or-none processes in learning and retention. _American Psychologist_, 1964, _19_, 16-25.

Greeno, J. G. Paired-associate learning with massed and distributed repetitions of items. _Journal of Experimental Psychology_, 1964, _67_, 286-295.

Greeno, J. G. Paired-associate learning with short-term retention: mathematical analysis and data regarding identification of parameters. _Journal of Mathematical Psychology_, 1967, _4_, 430-472.

Greeno, J. G. Some attempts to locate associative information. Prepublication, 1967.

Holland, P. W. A variation on the minimum chi-square test. *Journal of Mathematical Psychology*, 1967, *4*, 377-413.

Keppel, G. Retroactive and proactive inhibition. In T. R. Dixon and D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968, 172-213.

Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 1-21.

Peterson, L. R., Saltzman, D., Hillner, K., and Land, V. Recency and frequency in paired-associate learning. *Journal of Experimental Psychology*, 1962, *63*, 396-403.

Peterson, L. R., Hillner, K., and Saltzman, D. Supplementary report: time between pairings and short-term retention. *Journal of Experimental Psychology*, 1962, *64*, 550-551.

Postman, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal Learning and Verbal Behavior*. New York: McGraw-Hill, 1961, 152-179.

Rumelhart, D. E. The effects of interpresentation intervals in a continuous paired-associate task. Tech. Rep. No. 116, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1967.

Shiffrin, R. M., and Atkinson, R. C. Some speculations on storage and retrieval processes in long-term memory. Tech. Rep. No. 127, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1968.

Thompson, W. J. Recall of paired-associate items as a function of interpolated pairs of different types. *Psychonomic Science*, 1967, *9*, 629-630.

Thorndike, E. L. *The teacher's word book*. New York: Teachers College, Columbia University, 1921.

Underwood, B. J. Interference and forgetting. *Psychology Review*, 1957, *64*, 49-60.