LEARNING AND SHORT-TERM RETENTION OF PAIRED ASSOCIATES IN RELATION TO SPECIFIC SEQUENCES OF INTERPRESENTATION INTERVALS

by

Robert A. Bjork

TECHNICAL REPORT NO. 106

August 11, 1966

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

ACKNOWLEDGMENTS

I am indebted to Professor William K. Estes, chairman of my dissertation committee, for his constant good counsel and his generosity with time and suggestions. I also wish to thank the other members of my committee, Professors Richard C. Atkinson and James G. Greeno, for their encouragement and instruction. Professor Harley A. Bernbach gave freely of his time to read and criticize the dissertation in draft form. I am indebted to David E. Rumelhart for many profitable discussions and for his help with the data analysis, and to Donald P. Horst for his assistance in the programming necessary for the experiment. My thanks to Sydney Burkhart, Helen Butler, and Madlyn Lees, who guided me through a number of procedural matters, and to Maria Jedd for her painstaking preparation of the figures in this dissertation. Finally, I am especially grateful to my wife, Tele, who not only relieved me of concerns extraneous to this project, but also gave me efficient and cheerful assistance.

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 Grant MH6154 from the United States Public Health Service and, in part, by Contract NONR-225(73) between the Office of Naval Research and Stanford University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Inferences concerning the nature of simple verbal association have been based primarily on measures of average performance during the learning of lists of paired associates. There is some evidence, however, that the learning of any single paired associate in a list may exhibit effects idiosyncratic to the spacing of its presentations which are lost in the averaging of performance measures for all items; the particular sequence of intervals separating the presentations of any given paired associate may result in significant short- and long-term effects peculiar to the sequence.

This dissertation focuses on the learning of the individual members of a list of paired associates rather than on the overall learning of the list. The conceptual and experimental stress of this paper is on the acquisition of the individual members as a function of the spacing sequence between their presentations.

The experimental behavior with which this dissertation is most directly concerned is the learning of paired associates by means of a series of anticipation trials. An anticipation trial begins with the presentation of a stimulus to which the subject attempts to anticipate the correct response and ends with a presentation of the correct response (reinforcement).

It is customary in paired-associate experiments which utilize the anticipation method to present one randomization of the list to be learned, then another, and so on until the end of the experimental session. Consider, in such an experiment, the presentations of any particular paired associate. The interval, filled with interpolated presentations of other items, between any two successive presentations of the item varies from a few seconds to a number of minutes in the usual experiment; i.e., the number of other items which intervene between two occurrences of the given item in consecutive randomizations of the list varies from zero to twice the number of other items in the list. Thus any one paired associate, during the course of an experimental session, is characterized by a specific sequence of intervals between its presentations.

In general, there are two main ways in which the interval between two presentations of an item could affect the performance on the item. (1) The subject's immediate memory for the item may vary inversely with the length of the interval; if so, performance on a trial following a short interval will be better than performance on a trial following a long interval. (2) The effectiveness of the reinforcement occurring on a trial, as reflected by performance on later trials, might vary as a function of the interval since the preceding presentation.

Chapter 1 of this paper presents a preliminary survey of some progress and problems in the study of simple association that are especially relevant to the present study. Chapters 2, 3, and 4 present the method, results, and theoretical analysis of an experiment designed to allow direct analysis of the idiosyncratic effects on paired-associate learning of specific presentation sequences.

CHAPTER 1

SOME PROGRESS AND PROBLEMS: A PRELIMINARY SURVEY

Since the middle 1950's, research in the areas of short-term memory and paired-associate learning has resulted in considerable progress relevant to the present study. This progress derives from the joint development of some experimentally fruitful conceptions of the learning-memory process and some conceptually fruitful experimental procedures for studying short-term memory-learning.

The One-Element Model

One very important theoretical advance was the development of the simple all-or-none (one-element) model for verbal association (Bower, 1961; Estes, 1961). The one-element model assumes that there is no partial learning; a given paired associate is either in a perfectly learned state or an unlearned guessing state. Any opportunity to learn an unlearned association (the pairing of stimulus and response) has an all-or-none effect: the association is either learned completely or not learned at all with fixed probabilities c and $1-c$, respectively.

There are two reasons why the all-or-none model is so important. First, in certain ideally simple experimental situations, the model predicts numerous statistics of performance during learning with striking accuracy. Second, when in more complex situations, the learning does not conform to the predictions of the all-or-none model, the specific way in which it differs is often instructive.

The learning of a list of paired associates by the anticipation method usually corresponds well with the predictions of the one-element model when (1) the stimuli are simple and easily discriminated, (2) and the responses come from a well-learned set, are few (two or three), and are each paired with the same number of stimuli. A primary example of such an experiment and the accuracy with which the all-or-none model accounts for the learning is reported by Bower (1961).

If the experimental situation just described is changed along one or more of several dimensions, the (observed) learning tends to diverge from the predictions of the all-or-none model. Even in the "ideal" cases, however, it appears to the present author that there are aspects of the data which cannot be accounted for by the one-element model. In spite of the accuracy with which the models fit a variety of statistics averaged over all items, this paper contends that performance on individual items is subject to immediate memory effects not predicted by the model. The most apparent instance of such effects should occur in the standard anticipation procedure when a paired associate presented last in one randomization of the list is presented first in the next randomization. Since no other items intervene, it seems reasonable to expect that performance on the second of the two presentations should be essentially perfect, due to the subject's short-term memory, independent of the performance level prior to the presentation.

One could augment the one-element model, however, to include a short-term memory state. Assume that, on any one trial a paired associate can be in any one of three states: unlearned (U), learned (L), or short-term memory (S). The subject responds correctly with probability one if an item is in the learned or short-term memory states and with only chance (guessing) probability if the item is in the unlearned state. The effect of a reinforcement and the effect of an intervening trial on the state of an item are shown below.

THE EFFECT OF A REINFORCEMENT:

$$
\begin{array}{c} \\ L_n \\ S_n \\ U_n \end{array}
\begin{array}{c} \begin{array}{ccc} L_{n+1} & S_{n+1} & U_{n+1} \end{array} \\ \left[\begin{array}{ccc} 1 & 0 & 0 \\ c & 1-c & 0 \\ c & 1-c & 0 \end{array}\right] \end{array}
\qquad
\begin{array}{c} P(\text{correct}\,|\,\text{row state}) \\ \left[\begin{array}{c} 1 \\ 1 \\ g \end{array}\right] \end{array}
$$

THE EFFECT OF AN INTERVENING TRIAL:

$$
\begin{array}{c} \\ L_n \\ S_n \\ U_n \end{array}
\begin{array}{c} \begin{array}{ccc} L_{n+1} & S_{n+1} & U_{n+1} \end{array} \\ \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1-f & f \\ 0 & 0 & 1 \end{array}\right] \end{array}
$$

These two transition matrices summarize the assumed associative and forgetting effects of a reinforcement and an interpolated trial, respectively. Upon the presentation of an item in the unlearned or short-term memory states, the item is permanently learned with probability c or is stored in short-term memory with probability $1-c$.

With each intervening trial an item in short-term memory is forgotten into the unlearned state with probability f.

Applied to the average learning of a list of paired associates, this three-state (one-element forgetting) model differs only slightly from the one-element model, especially if the forgetting rate, f, is large. In either model there is a fixed probability c that an unlearned item will be learned as a result of a reinforcement. The only differentiating predictions of the models derive from those instances in which an unlearned item remains in state S from the time of one reinforcement of the item until the next test on the item. One statistic which should reflect such instances is the mean proportion of correct responses prior to the last error. The one-element model predicts only chance performance before the last error and the memory model predicts somewhat above chance performance before the last error. This difference arises because, although in both models an item cannot be in the learned state prior to the last error, the memory model allows some items to remain in short-term memory from which they are retrieved correctly.

The size of the difference between the predictions of the models will depend on the forgetting rate, f, and the average number of interpolated trials (which increases with the length of the list to be learned) between presentations of any given item. To illustrate that this difference is usually negligible, consider the following typical situation and parameter values: there are ten items in the list, $g = .50$ and $f = .30$. The one-element model predicts chance performance, $g = .50$, prior to the last error. In this situation the average number of other items intervening between successive presentations of an item is nine.

Hence, for the one-element forgetting model, the approximate predicted proportion correct before the trial of last error is

$$(1-f)^9 + [1-(1-f)^9]g = (1-f)^9(1-g) + g$$

$$= (0.70)^9(0.50) + 0.50$$

$$= 0.52 \ .$$

The one-element forgetting model seems a promising model for the learning of individual paired associates in those situations where the one-element model accounts well for the average performance. The model predicts in detail the effects of any specific sequence of intervals between the presentations of any individual paired associate. It contains a reasonable geometric forgetting assumption and when applied to average performance measures it reduces, essentially, to the one-element model. But, in spite of its promise, there is a collection of experimental results which suggest that, to account for all the effects of spacing schedules, some more elaborate extension of the one-element model is required. The next section reviews some of these results with respect to the general conception of paired-associate learning as a three-state process.

Three-State Models: Some Problematic Experimental Results

The one-element forgetting model is only one of many possible three-state models. The general notion that on any trial an item may be in either long-term (permanent) memory, in short-term (transient) memory, or not in memory is common to models with very different assumptions concerning learning and memory. It seems worthwhile to investigate whether any model within this general framework is adequate to account for the effects of spacing on the learning of simple paired associates.

The following experimental results from several different procedures raise some important considerations about the assumptions of an adequate three-state model.

(1) The effects of a given interval between two successive presentations of a paired associate appear to be more complicated than the short-term memory loss postulated in the one-element forgetting model. An experiment by Greeno (1964) illustrates that there may be significant long-term as well as short-term effects of the interval length. In one condition of Greeno's experiment subjects learned a list of paired associates by the anticipation method except that some items were presented twice in each randomized block. The repetitions of the repeated items were massed with only zero or one of the other items intervening. Figure 1 shows the mean learning curve for the repeated items and the mean learning curve for the nonrepeated items across the randomized blocks. These learning curves suggest that when presentations are spaced very closely their long-term efficacy, as measured by later performance, is no better than a single presentation. Even though performance on the second of each pair of massed repetitions is very high owing to the subjects' immediate memory, the second presentation appears to add nothing to later performance.

On the face of it, Greeno's result implies that the probability of learning out of a short-term memory state is less than the probability

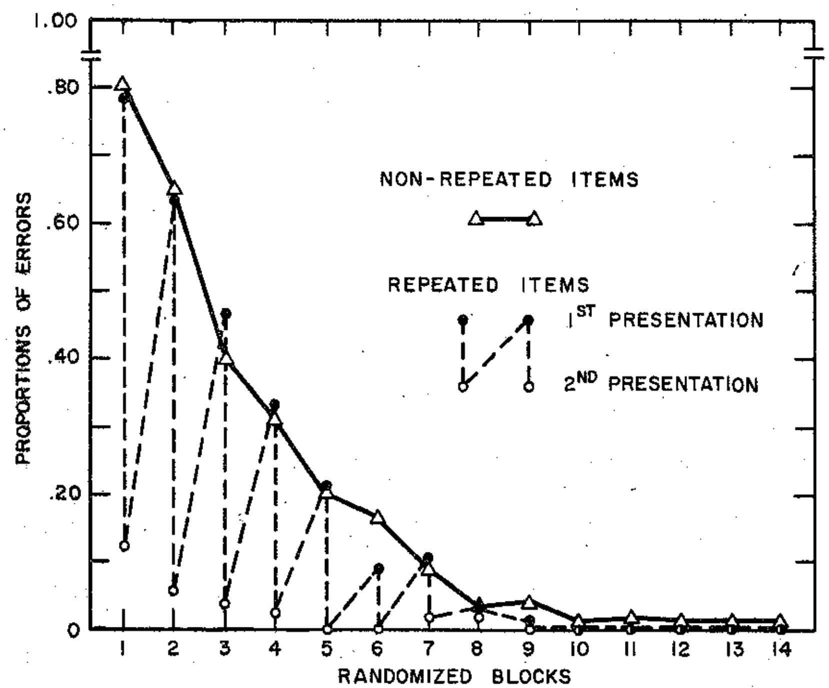


Figure 1. Mean Learning Curves for Repeated and Non-Repeated Items (Greeno, 1964).

of learning out of a guessing state and may be negligible. This result is inconsistent with the assumption of the one-element forgetting model that the probability of transition to learned state is independent of spacing.

(2) The results of a "miniature" experiment by Peterson, Hillner, and Saltzman (1962) suggest an even more complex interaction of the long- and short-term effects of the interval between two presentations. Subjects were presented with a running series of study trials and test trials. On a study trial (R) both the stimulus and correct response were presented; on a test trial only the stimulus was presented. Each of a number of paired associates had two study trials followed by a single test trial. The study trials were separated by zero or four other trials, and one, two, four, or eight trials intervened between the second study trial and the test trial. Performance on the test trial is shown in Figure 2 for each of the two spacings of the study trials as a function of the interval between the second study trial and the test. The result that performance on a test following four or eight trials is better the more the study trials are spaced replicates, in general, Greeno's finding. But when the test follows only one or two trials, performance seems better with the study trials closely spaced. Thus, the effect of the spacing between two presentations as measured by later performance may depend on how much later the performance is measured.

(3) In similar miniature experiments, one by Peterson, Wampler, Kirkpatrick, and Saltzman (1963) and the other by Young (1966), the results imply that there may be a limit to the improvement of later

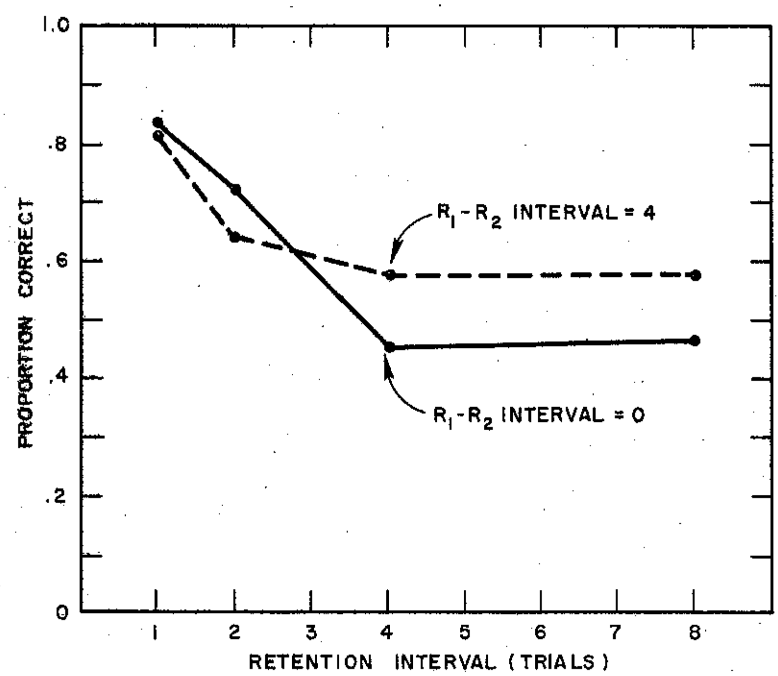


Figure 2. Mean Retention Curves for Different Spacings Between a First and Second Study Trial (Peterson, et al., 1962).

performance with an increase in the spacing of two study trials. In both experiments a single test trial follows the second of two study trials after a fixed number of intervening trials. Eight trials intervened in Peterson's experiment and ten trials intervened in Young's experiment. The $R_1 - R_2$ interval between the two study trials, however, was varied in both experiments. In Figure 3 the proportion of correct responses on the test trial appears in both experiments to rise initially and then drop slightly as the $R_1 - R_2$ interval increases.

The results of the two preceding miniature experiments are for experimental tasks somewhat different from the task of primary interest to this paper, the learning of a list of paired associates. They do illustrate, however, the likelihood that any three-state model adequate to account for all the effects of spacing will require a fairly complicated short-term memory structure.

(4) There are also some indications that the long-term and short-term effects of a given interval between presentations may change during the course of an experimental session. This possibility is inferred from the results of standard anticipation method experiments when the list of paired associates to be learned is long (twenty or longer). In such experiments the learning rate is often slower during the early trials than during the later trials and the conditional probability of an error given an error on the preceding trial tends to decrease across trials. Both of these effects could reflect an increase in the probability of transition to the learned state across trials. They could also reflect instead, or in addition, an improvement across trials in the short-term memory of items not in the learned state.

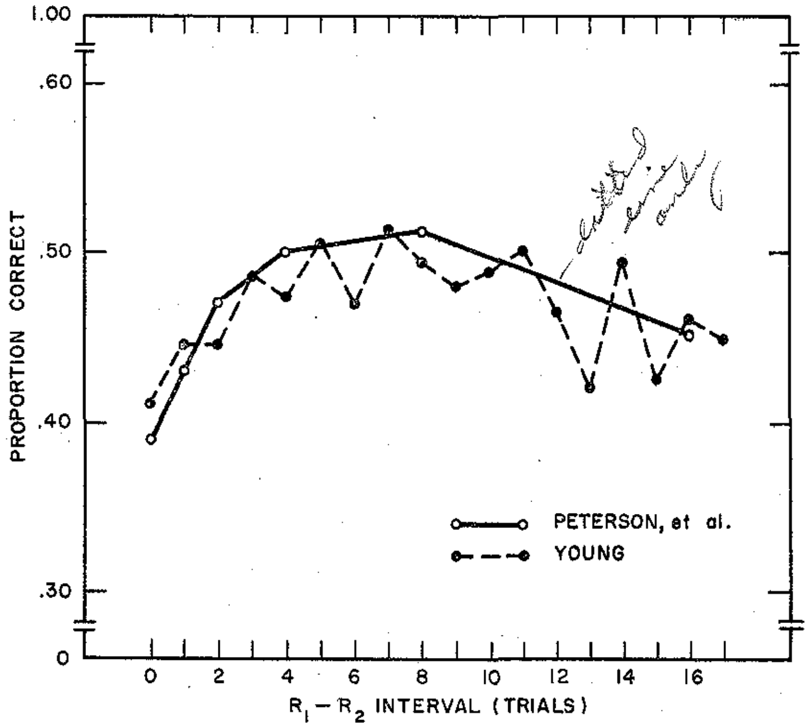


Figure 3. Proportions of Correct Responses as a Function of the Spacing of Two Study Trials (Peterson, et al., 1963; Young, 1966).

Calfee and Atkinson (1965) postulated a three-state "trial-dependent-forgetting model" for such experiments. The model assumes that only intervening items not in the learned state tend to interfere with the short-term retention of other items. Hence, during the course of an experimental session as more and more members of the list are learned, short-term memory for items yet to be learned improves. The model also allows for different rates of learning out of the guessing and short-term memory states. Thus, if learning is faster out of the short-term memory state, the trial-dependent-forgetting model predicts both an increase in the learning rate and an increase in the conditional probability of an error given a preceding error $(P(e_n|e_{n-1}))$ across trials. Figure 4 compares the predictions of the trial-dependent-forgetting model with the observed learning curve and $P(e_n|e_{n-1})$ curve from an experiment reported by Calfee and Atkinson.

The predicted curves in Figure 4 are obtained from parameter estimates in which the probability of learning out of the short-term memory state (0.42) is almost four times the probability of learning out of the guessing state (0.11). Thus, there seems to be an inferred but fundamental conflict between the results in Figure 4 and Greeno's results in Figure 1. In the former, the probability of transition to the learned state is estimated to be higher from the short-term memory state than from the guessing state. In the latter, it appears that the probability of learning out of short-term memory is negligible.

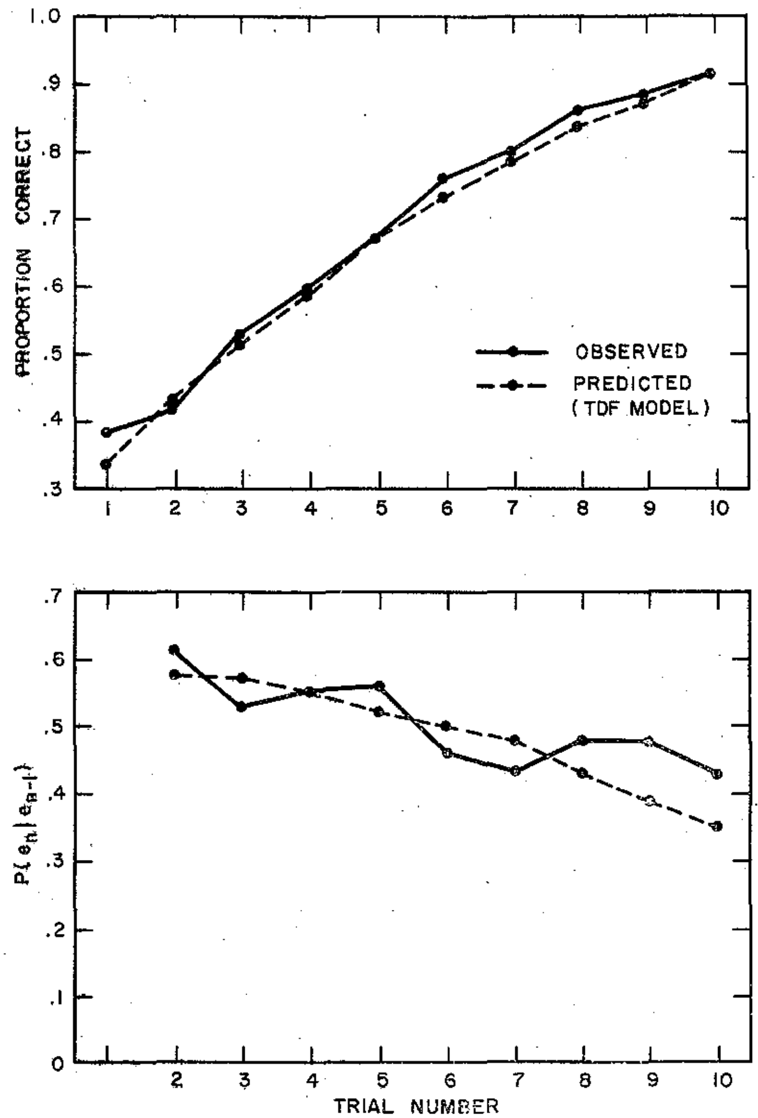


Figure 4. Predicted vs. Observed Learning Curves; Predicted vs. Observed Probabilities of an Error on Trial n+1 Given an Error on Trial n (Calfee and Atkinson, 1965).

It is also possible that the guessing state should not be thought of as a single state (Atkinson and Crothers, 1964). There may be a difference between the effect of a trial on which the stimulus is familiar and the response forgotten and the effect of a trial on which the stimulus is unfamiliar. If the probabilities of transition to the learned and short-term memory states are lower from an unfamiliarized guessing state than from a familiarized-forgotten guessing state, learning should be slower on the early trials.

Part of the difficulty in clarifying the effects of spacing on the learning and short-term retention of individual paired associates is experimental. Most of the evidence about these effects in standard list-learning experiments is inferential. When, in the typical experiment, performance measures on all items are averaged together the effects of spacing are largely averaged away. The next section points out some of the shortcomings of the standard anticipation procedure and suggests some changes to allow a more direct study of spacing effects.

Experimental Analysis: Some Unfortunate Properties of the Standard Anticipation Method

There are two main reasons in standard experiments that the average learning of a list of paired associates does not significantly reflect the effects of spacing operative on the individual members of the list.

(1) The typical procedure has a structure which, statistically, avoids the interpresentation intervals most likely to have significant effects, the very long and very short intervals. Consider any two successive randomizations of a list of L paired associates. The number

of other items interpolated between the presentation of any particular item in the first randomization and its presentation in the second randomization can vary from zero to $2L - 2$. In order for the number of interpolations to be small the item must occur late in the first randomization and early in the next, a statistically unlikely event.[1] In order for the number to be large the item must occur early in the first randomization and late in the next, another unlikely event. The actual probability distribution over the possible numbers of interpolations is triangular. If I_n is the number of other items interpolated between the n^{th} and $n+1^{st}$ presentations of a given item,

$$P(I_n = k) = \begin{cases} \dfrac{k+1}{L^2} & k = 0, 1, \ldots, L-1 \\[2ex] \dfrac{2L-(k+1)}{L^2} & k = L, L+1, \ldots, 2L-2. \end{cases}$$

Figure 5 presents, as an example, the distribution which occurs with a ten-item list $(L = 10)$.

Given the high probability of interpresentation intervals close to the length of the list in the standard procedure, the specific sequence of intervals characterizing the presentations of an individual item does not tend to exhibit much variation. In order for any effects of spacing to be more apparent, the standard procedure needs to be changed so that short and long intervals are more likely to occur.

---

[1] This event is made even less likely by some experimenters who constrain the randomization to prevent the occurrences of very short intervals.

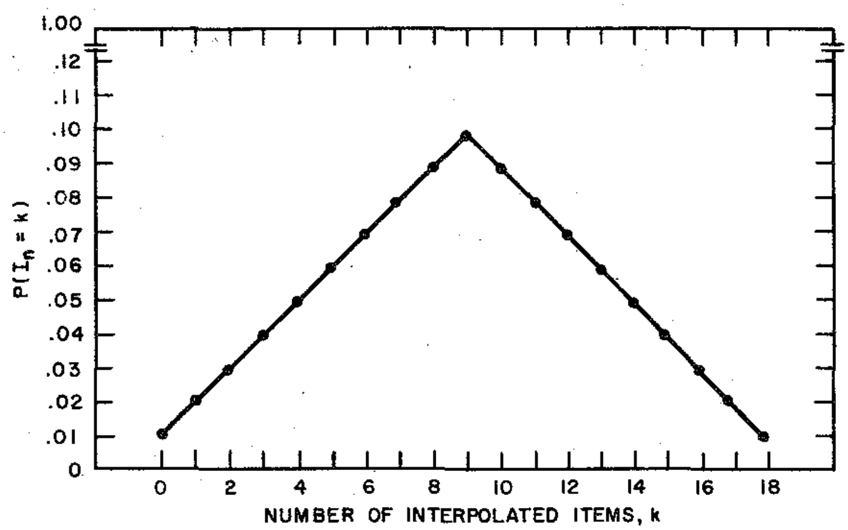


Figure 5. Distribution of Interpresentation Intervals for a Ten-Item List.

(2) The second reason that the effects of spacing on the acquisition of individual items are not reflected in the learning of the list is due to the averaging of performance measures for all items. Those effects which do occur in the standard procedure are spread over all items and trials by the averaging and subject-to-subject randomization processes. If one's main interest is in list learning, it is obvious why presenting a unique series of randomized cycles of the list to each subject is desirable. The procedure reduces the chances that effects peculiar to single items or presentation schedules, and not characteristic of items or schedules in general, will color the performance on the list.

In the study of spacing effects, it is still important to randomize across subjects the assignment of paired associates to presentation sequences to avoid confounding item differences with the spacing variable of interest. But it is self-defeating to change presentation schedules from subject to subject. For example, if it is of interest to know the effects of having only one interpolation between the second and third presentations of an item all subjects should have an item with that specific property.

The next chapter describes an experiment which modifies the standard anticipation method in the ways suggested above. It is designed to reveal more directly the learning and short-term retention of paired associates in relation to specific sequences of interpresentation intervals.

CHAPTER 2

METHOD

Design of the Experiment

In order to emphasize the differences between the experimental method and the standard anticipation method, it is necessary to introduce some conventions and notation. One possible source of confusion is in the use of the word "trial." In what follows, the word "trial" is used to designate a single anticipation trial of the experiment and is not used, as is often the case, to designate the set of trials on which the various members of the list all have their n^{th} presentation. The word "presentation," unless followed by "of the experiment" refers to the presentation of any particular item.

The following notation is used to describe the experiment:

$1, 2, \ldots, k, \ldots, L$:	The members of a list of L paired associates.
$1, 2, \ldots, N, \ldots$:	The trials of the experiment.
$1, 2, \ldots, n, \ldots$:	The presentations of any one item.
N_n :	The trial number of the n^{th} presentation of a particular item.
$t_n = N_{n+1} - N_n$:	The number of trials from the n^{th} presentation of an item until the $n+1^{st}$ presentation of the item. (Note that t_n equals the number of interpolated items plus one.)
$(t_1, t_2, \ldots, t_n)$:	The sequence of interpresentation intervals (presentation sequence) characterizing the presentations of any particular item.

The standard anticipation method was altered in two significant ways in this experiment.

(1) The randomized cycle structure of the standard anticipation method was replaced by a procedural algorithm designed to generate a series of trials in which any interpresentation interval in the range, $t_n = 1, 2, \ldots, 2L - 1$, is equally likely to occur. That is, for any item k and presentations $n, n+1$,

$$P(t_n = j) = \frac{1}{2L - 1}, \quad j = 1, 2, \ldots, 2L - 1.$$

Compared to the standard procedure, the algorithm was designed to increase the frequency of short and long intervals without changing either the range or the average of the intervals generated. The Appendix contains a description of the algorithm along with some comments on its uses and its distributional properties.

(2) During an experimental session consisting of a series of anticipation trials on the members of a list of paired associates, the presentations of any individual item are characterized by a sequence of interpresentation intervals, $(t_1, t_2, \ldots, t_n)$. The second distinguishing feature of this experiment was that every subject had the same series of trials in the sense that each had the same set of presentation sequences. That is, every subject had exactly one item in the list assigned to each of the specific $(t_1, t_2, \ldots, t_n)$ sequences. The confounding of item differences with the effects of the presentation sequences was avoided,

however, by counterbalancing across subjects the assignment of items to presentation sequences.

There were 21 paired associates in the experimental list, and each experimental session consisted of 409 trials. The 21 specific presentation sequences used for all subjects are shown in Table 1. The trial of the first presentation of each presentation sequence, N_1, is shown in the left column of the table. The trial of any later presentation, N_n, is just the sum of N_1 and the first $n-1$ interpresentation intervals. That is,

$$N_n = N_1 + \sum_{i=1}^{n-1} t_i \; .$$

Subjects

The subjects were 50 freshmen at Stanford University, 25 men and 25 women. They were obtained from the freshman dormitories, and they were paid $1.50 to participate in the experiment. The experiment ran about 50 minutes, including instructions.

Experimental Materials

A list of 21 paired associates was constructed, with nonsense syllables as stimuli and the digits 3, 5, and 7 as responses. Each of the responses was paired with seven of the stimuli.

There were several restrictions on the selection of the nonsense syllables. (1) No two of them overlapped in more than one letter. (2) No consonant was used more than twice as the first letter or more than twice as the third letter. (3) No consonant was used only once or more than three times in the twenty-one syllables. (4) The vowels

Table 1

Experimental Sequences of Interpresentation Intervals

Present. Sequence	N_1	Interpresentation Intervals																												
		t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}	t_{29}
1	1	25	23	5	18	19	37	16	33	26	28	31	3	15	24	2	29	4	40	6	12									
2	2	19	2	35	31	31	15	22	33	21	19	30	18	15	22	16	25	27												
3	3	14	16	1	29	19	28	6	21	37	8	14	28	14	14	7	19	30	41	11	3	24								
4	4	32	19	29	28	24	12	19	2	14	11	7	7	7	34	37	25	27	12	3	31									
5	5	34	4	28	9	35	35	35	34	20	4	25	39	10	26	28	28													
6	6	12	28	35	18	41	22	37	19	14	18	35	41	1	15	19	39													
7	7	24	31	30	14	11	14	39	23	33	28	12	16	11	8	1	12	33	22	4	3									
8	8	29	32	26	38	30	34	19	17	9	9	30	35	16	12	22	14													
9	9	6	37	1	3	37	16	33	19	27	39	14	22	8	12	36	23	20	1	32	4									
10	10	30	4	29	2	13	23	23	34	7	9	20	13	30	29	33	13	35	18											
11	11	11	26	11	7	19	36	8	27	25	17	27	3	18	2	31	12	6	15	20	21	39								
12	12	18	2	19	36	15	2	37	10	9	6	25	16	30	40	28	23	20	34											
13	13	3	34	26	38	16	17	5	2	24	9	8	15	13	22	8	7	1	33	4	17	16	5	22	30					
14	14	24	41	39	4	17	4	29	40	8	24	40	36	35	36															
15	19	5	40	10	26	8	11	7	29	37	14	29	32	29	23	37	11	1	6											
16	20	9	38	31	7	20	28	23	33	4	17	27	30	12	35	12	35	8												
17	25	32	33	17	17	22	34	34	41	37	32	13	33																	
18	27	38	36	37	41	10	33	34	39	17	17	35	32																	
19	28	14	28	16	41	31	28	14	29	35	5	3	28	21	2	29	26	22												
20	35	12	13	37	6	20	26	15	38	32	41	35	35	41																
21	41	4	16	7	9	1	5	11	2	17	19	13	14	6	6	2	32	6	25	22	15	1	15	1	40	29	13	5	6	12

A, I, O, and U were each used in four of the pairs and the vowel E was used in five pairs. (5) The nonsense syllables were of medium-low meaningfulness (Archer, 1960).

The stimuli used were:

JUC	TAW	NAL
BIJ	PAF	RUW
GEB	VUR	MOH
VOS	ZEG	ZAN
FIP	HIF	CEH
BEM	SEJ	TUL
LIR	WOV	JOM

For each subject, one of the responses 3, 5, or 7 was paired with each stimulus. The assignment of responses to particular stimuli was randomized across subjects.

Experimental Apparatus

The experiment was run on an automated verbal associative learning apparatus located at the Institue for Mathematical Studies in the Social Sciences.

Each subject sat by himself in a soundproofed, air-conditioned room. On a table in front of the subject there was a response panel with three 7/8 in. by 1 in. response keys, 2 inches apart. The keys were set in a column on the panel and were marked with the digits 3, 5, and 7 in ascending order. The subject sat with his right hand on the response panel, his thumb, index finger, and middle finger resting on the response keys marked 3, 5, and 7, respectively. On a second table

5.5 feet away from the subject there were two visual display boxes, one on top of the other. The front of each box was a 2 in. by 12 in. display panel. The stimulus members of the paired associates were presented on the top panel and the response member was presented on the bottom panel.

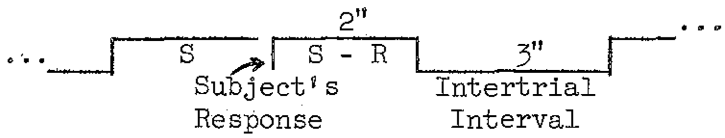
In an adjoining room, a key punch and additional timing and storage equipment allowed the experimental session to be preprogrammed. A deck of 409 IBM cards was prepared for each subject. Each card in the deck contained the information necessary to determine the stimulus and response events of a single trial. The apparatus presented the stimulus, recorded the subject's response and response latency, presented the stimulus and response together for two seconds, and after a three-second interval started the next trial. A more complete description of the verbal associative apparatus can be found in Yellott (1965) or Izawa and Estes (1965).

Experimental Procedure

After being seated, each subject was instructed on the nature of the experimental task. The anticipation trial procedure was explained in detail and two example trials were run through to familiarize the subject with the procedure. When the subject indicated that he understood the procedure, the experimenter left the room and the subject started the series of trials by pushing one of the response buttons.

Every subject was presented with a total of 409 trials. The first nine trials were pretraining trials. They utilized seven dummy items, two of which were repeated during the nine trials. There was no break between the pretraining trials and the 400 trials on the twenty-one items to be learned.

Each anticipation trial began with the presentation of a nonsense syllable. The subject then attempted to anticipate the correct response by pushing one of the buttons on the response panel. Immediately following the subject's response, the correct response appeared together with the stimulus on the display panel. The timing sequence of these events is shown below.



Although the subject was free to take an indefinite time to respond, he was encouraged to respond as quickly as he could without making mistakes.

CHAPTER 3

EMPIRICAL ANALYSIS OF THE RESULTS

The experiment permits several empirical analyses of the course of learning as a function of the individual presentation sequences. It also permits the standard analyses of learning data averaged over all presentation sequences. These two types of analyses are presented in the first and second sections of this chapter, respectively.

Performance in Relation to the Individual Presentation Sequences

In Figure 6 the learning curve and the average success latency curve are shown for each of the twenty-one presentation sequences. Note that the points on any curve are spaced along the abscissa in direct correspondence to the sequence of interpresentation intervals which characterizes the particular presentation sequence. This correspondence is accomplished as follows: all 400 trials of the experiment are laid out on the abscissa of each graph and the presentation sequences (see Table 1) are indicated with hash marks. That is, for any presentation sequence, the trials of the experiment on which the presentations occurred are designated by short vertical marks below the abscissa. This representation permits the performance level at any point to be related visually to the spacing of successive presentations.

There are several significant effects of spacing on the performance measures shown in Figure 6. The most striking are the perturbations of the learning curves attributable to short-term retention. When the interval between two presentations is very short there is a marked

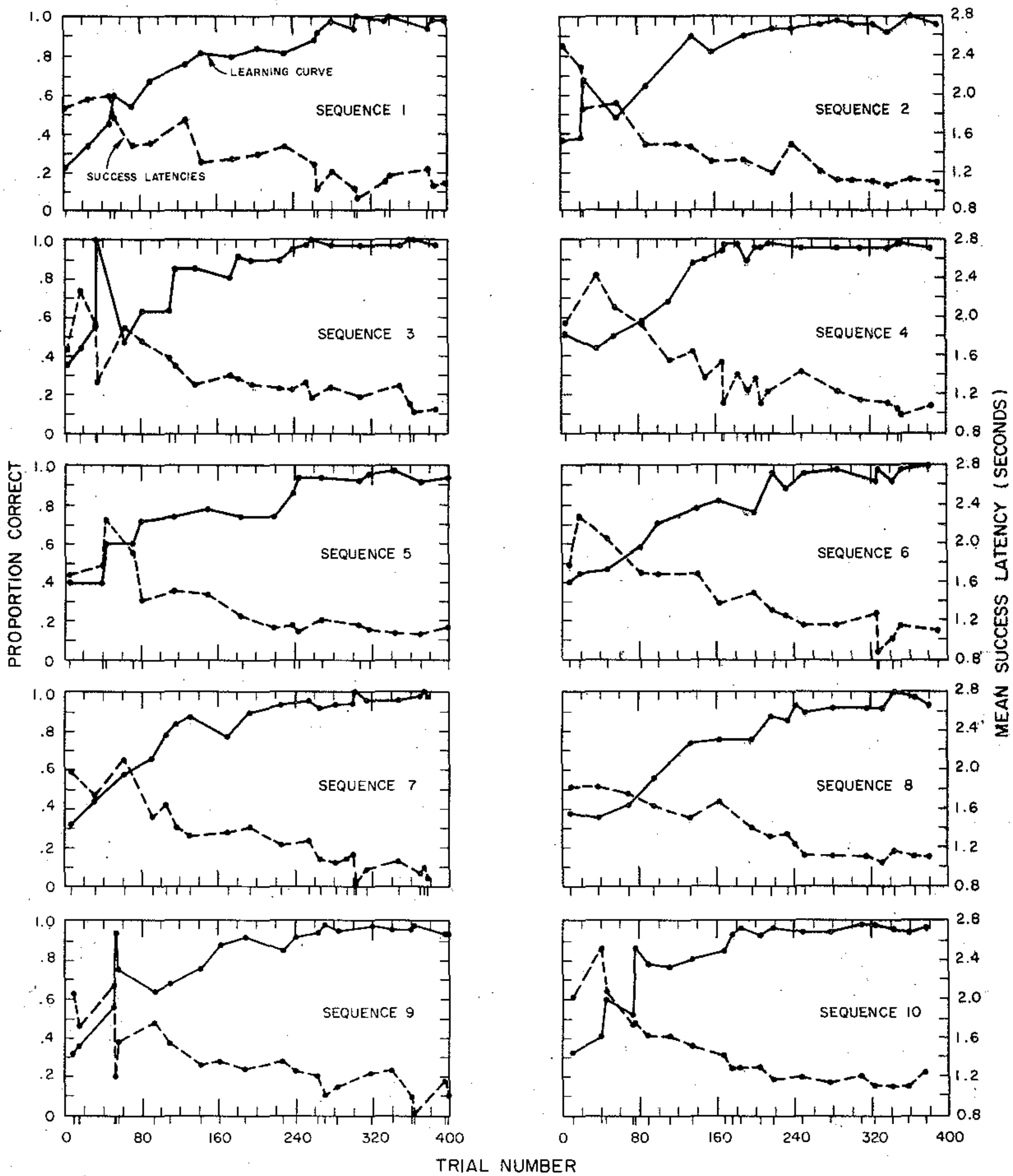


Figure 6. Mean Learning Curves and Mean Success Latency Curves for the Individual Presentation Sequences. (The hash marks below each abscissa designate the trials of the experiment on which the presentations occurred.)

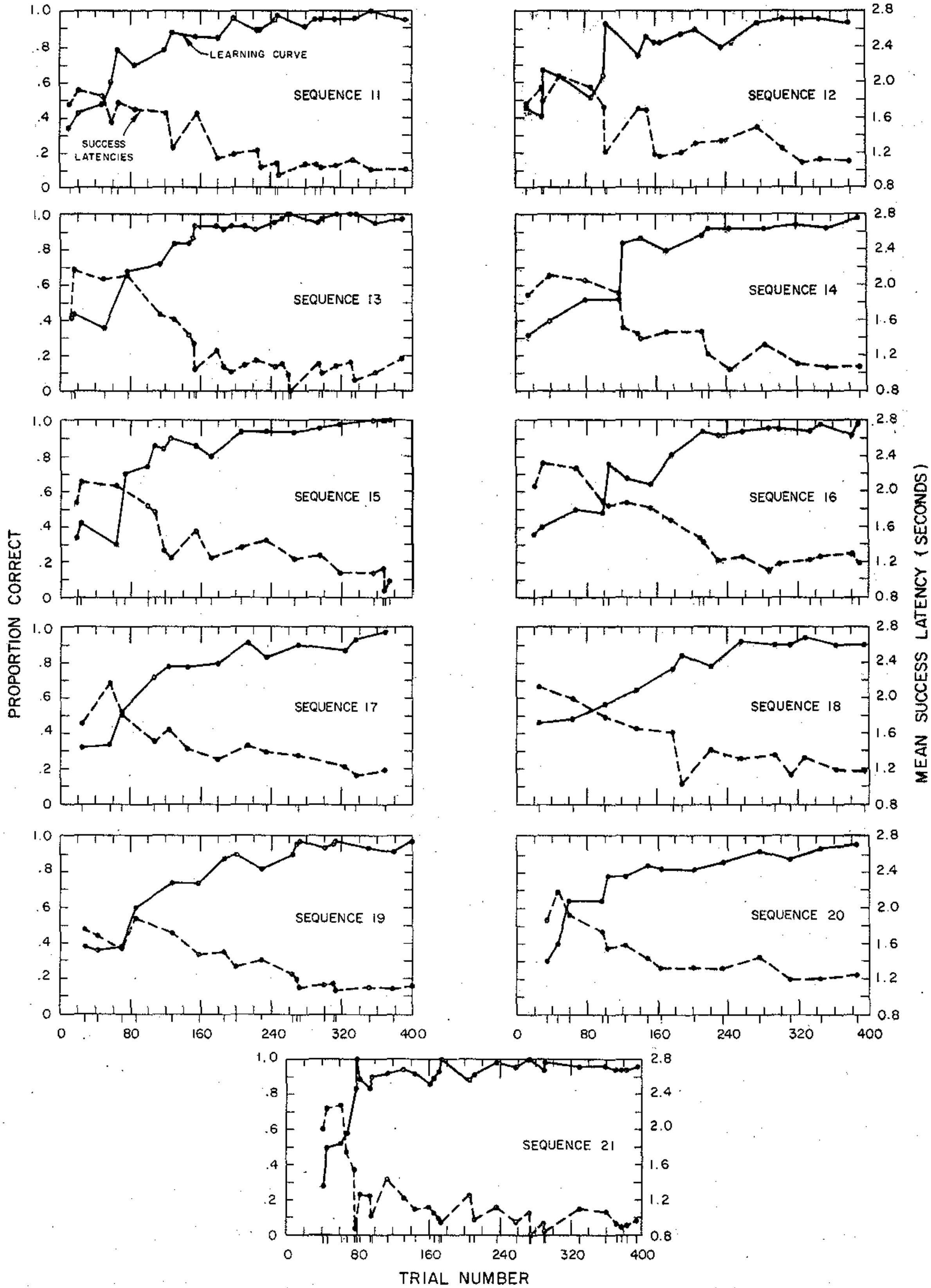


Figure 6 (continued).

temporary increase in the proportion of correct responses. Some obvious examples of this occur on the fourth presentation of sequence 3, the fourth presentation of sequence 9, the third and fifth presentation of sequence 10, and the third and seventh presentations of sequence 12. The temporary nature of these short-term memory effects is illustrated by their spike-like shape; if an interval of even moderate length follows a very short interval, the sharp rise in the proportion of correct responses is followed by a distinct drop.

The mean latency of correct responses seems to be similarly sensitive to the interval between presentations. In fact, the mean success latency curve in many cases looks like a reflection of the learning curve. There is often a negative spike in the mean latency of a correct response following a short interpresentation interval which matches the positive spike in the frequency curve. Two examples of this are the fourth presentation of sequence 3 and the fourth presentation of sequence 9. In both the learning curves and the mean success latency curves the short-interval spikes seem to occur late in the experiment as well as early in the experiment. Even between trials 300 and 400, when the proportion of correct responses is typically between 0.90 and 1.00 and the mean success latency tends to level off between 1.00 and 1.20 seconds, the sharp spikes occur following very short intervals.

In contrast to the learning curves and success latency curves, the mean error latency across presentations does not seem clearly sensitive to spacing. Table 2 gives the average error latencies on

Table 2

Mean Error Latencies on Presentations 1 through 13 of the Individual Presentation Sequences

Present. Sequence	Presentation Number												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1.64 (39)	2.23 (33)	1.95 (28)	2.40 (20)	2.16 (23)	2.08 (16)	1.72 (12)	3.18 (9)	1.62 (10)	1.60 (8)	2.07 (9)	1.81 (6)	1.34 (4)
2	1.81 (32)	2.07 (31)	1.91 (16)	1.95 (26)	1.98 (18)	1.67 (9)	3.38 (5)	2.20 (9)	1.70 (5)	1.28 (3)	1.77 (3)	1.01 (2)	1.51 (1)
3	2.20 (32)	2.07 (28)	2.39 (22)	(0)	2.46 (26)	1.73 (18)	2.12 (18)	2.01 (7)	2.54 (7)	1.67 (9)	1.78 (4)	3.15 (5)	2.15 (5)
4	1.86 (24)	2.28 (28)	1.94 (25)	2.26 (21)	2.38 (16)	1.86 (6)	1.66 (5)	1.15 (3)	1.93 (1)	1.04 (1)	1.90 (5)	1.71 (2)	2.36 (2)
5	1.80 (30)	2.17 (30)	2.32 (20)	2.12 (20)	2.48 (14)	2.24 (13)	2.77 (11)	2.33 (13)	2.81 (13)	2.01 (7)	2.00 (3)	1.21 (3)	2.06 (4)
6	1.86 (30)	2.69 (28)	1.88 (27)	2.39 (21)	2.39 (15)	2.37 (11)	2.19 (9)	2.34 (12)	1.78 (2)	2.18 (6)	2.69 (2)	2.13 (1)	1.58 (4)
7	2.37 (34)	2.53 (28)	2.25 (21)	2.46 (17)	2.25 (11)	2.42 (8)	2.70 (6)	1.98 (11)	1.73 (5)	2.12 (3)	1.70 (2)	2.85 (4)	1.49 (3)
8	1.85 (31)	2.02 (32)	2.29 (29)	1.98 (22)	1.63 (13)	1.87 (12)	2.25 (12)	1.48 (6)	1.82 (7)	1.19 (3)	1.38 (5)	1.75 (4)	1.80 (4)
9	1.90 (34)	2.09 (32)	1.87 (22)	1.47 (3)	1.88 (12)	1.83 (18)	1.77 (16)	2.04 (12)	2.15 (6)	1.36 (4)	2.03 (7)	1.74 (4)	2.19 (3)
10	1.99 (34)	2.29 (30)	2.22 (20)	2.69 (24)	2.70 (7)	1.86 (11)	2.22 (12)	2.00 (10)	1.88 (8)	1.50 (4)	2.30 (2)	2.42 (4)	2.75 (2)
11	1.96 (33)	2.31 (29)	2.26 (26)	2.90 (20)	1.67 (11)	2.02 (15)	2.10 (11)	1.59 (6)	1.40 (7)	1.71 (7)	2.39 (2)	1.42 (5)	1.68 (5)
12	2.16 (27)	2.44 (30)	2.47 (16)	2.00 (18)	2.22 (24)	2.37 (18)	1.67 (3)	2.06 (12)	1.88 (7)	1.64 (8)	2.51 (8)	2.35 (6)	1.95 (5)
13	2.01 (29)	1.77 (28)	2.27 (32)	1.73 (16)	2.45 (14)	2.10 (8)	2.07 (8)	1.48 (6)	1.95 (3)	1.69 (3)	1.65 (4)	1.87 (3)	1.29 (3)
14	1.72 (34)	2.32 (30)	2.40 (24)	2.67 (24)	1.95 (8)	1.81 (7)	1.53 (8)	1.67 (10)	1.89 (6)	2.29 (4)	1.45 (4)	1.12 (4)	1.16 (3)
15	1.87 (33)	2.14 (29)	2.21 (35)	2.51 (15)	1.64 (13)	1.37 (7)	2.19 (8)	2.17 (5)	2.10 (7)	2.48 (10)	2.32 (3)	1.62 (3)	1.68 (3)
16	1.94 (32)	2.32 (30)	2.59 (25)	2.23 (26)	2.12 (12)	2.30 (16)	2.18 (18)	2.72 (9)	2.34 (4)	1.49 (3)	1.72 (4)	1.45 (3)	1.12 (2)
17	2.01 (34)	2.06 (33)	2.54 (24)	2.04 (14)	1.94 (11)	2.26 (11)	2.63 (10)	1.54 (4)	1.77 (8)	2.16 (5)	1.91 (6)	1.40 (3)	1.37 (1)
18	2.05 (27)	1.96 (26)	1.83 (22)	2.43 (18)	1.71 (13)	2.14 (8)	1.81 (11)	2.02 (4)	1.35 (5)	1.34 (5)	.90 (3)	1.29 (5)	1.38 (5)
19	1.91 (31)	2.24 (32)	2.63 (31)	1.88 (20)	2.11 (13)	2.14 (13)	1.33 (6)	1.81 (5)	1.78 (9)	1.67 (5)	1.75 (2)	1.33 (1)	1.46 (3)
20	1.92 (35)	2.07 (30)	2.06 (18)	1.78 (18)	1.71 (11)	1.77 (11)	1.64 (8)	1.64 (9)	2.54 (9)	1.84 (7)	1.53 (4)	2.20 (6)	1.87 (3)
21	1.87 (36)	1.94 (25)	1.85 (24)	2.47 (21)	2.52 (9)	(0)	2.02 (6)	2.21 (8)	1.85 (5)	1.49 (4)	2.23 (3)	2.05 (4)	1.86 (7)

the first thirteen presentations of each of the twenty-one presentation sequences together with the number of observations in each average. The effect of a very short interpresentation interval on the average error latency is difficult to assess because the number of errors following a short interval is few if any. Also, the mean error latency curve becomes less and less stable across presentations as the number of errors decreases. On the whole, however, the error latencies seem not to show the decrease over presentations shown by the success latencies.

The general three-state conception of paired-associate learning discussed in Chapter 1 suggests an additional empirical analysis of the short-term effects of spacing. All models within the three-state framework imply that an error on any presentation, n, indicates that (1) the item presented was in the guessing state on trial N_n and (2) the item was never in the learned state prior to trial N_n. Hence, performance on an item prior to any error should be at chance level except for those cases when the subject responds correctly on the basis of short-term memory. The three-state notion thus implies that the short-term effects of the specific sequences of interpresentation intervals can be isolated by looking at performance prior to the last error.

Figure 7 shows the proportions of correct responses on presentations prior to the last error (forward stationarity curves) for each of the twenty-one presentation sequences. In this analysis the last error is defined with respect to a learning criterion of four or more successive

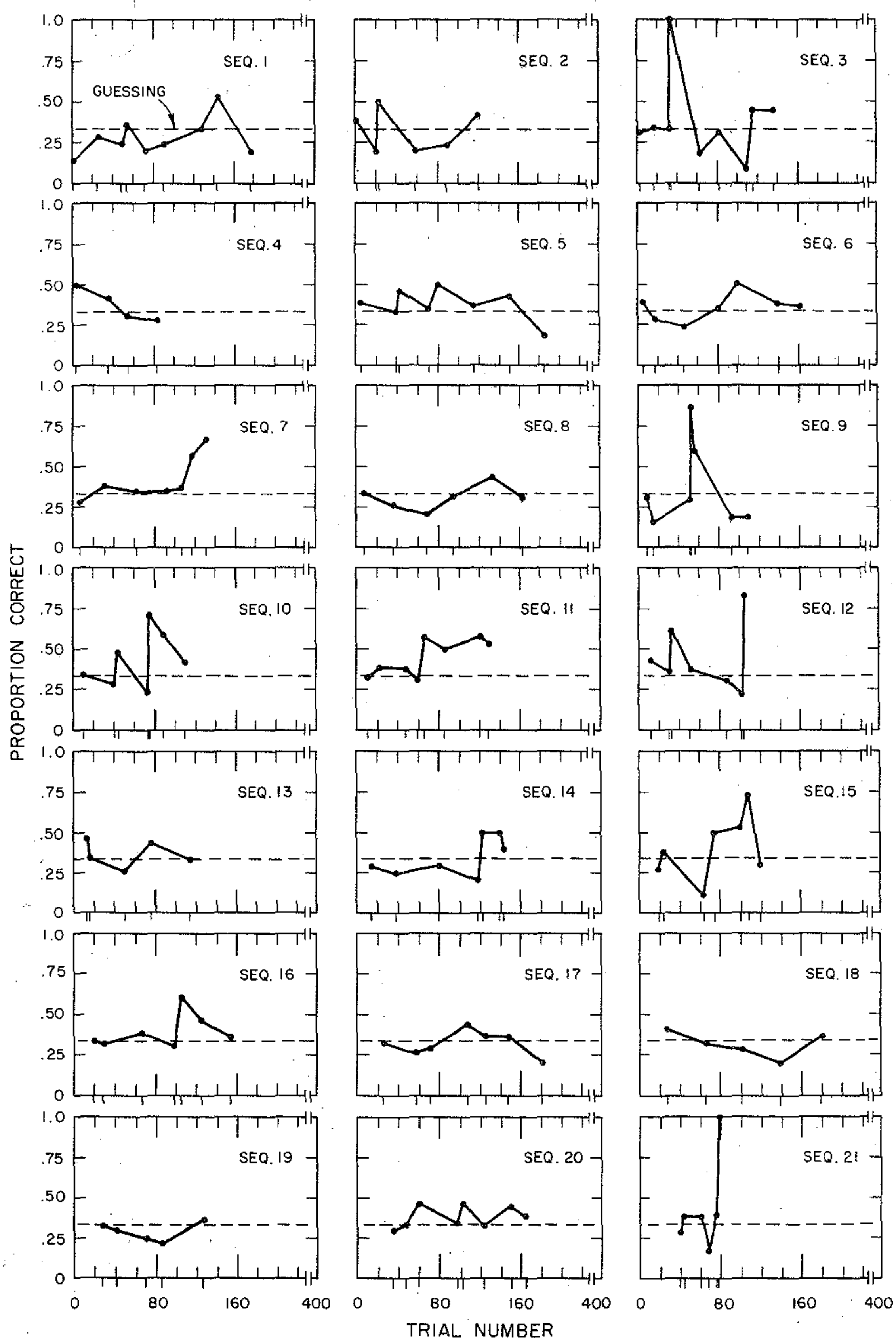


Figure 7. Proportions of Correct Responses on Presentations Prior to the Last Error for the Individual Presentation Sequences. (The hash marks below each abscissa designate the trials of the experiment on which the presentations occurred.)

correct responses on an item. Of the 1050 subject-item, error-success protocols, 1005 meet the learning criterion. The last error in the criterion protocols is defined as the error preceding the criterion run of four or more correct responses. In the 45 protocols which do not satisfy the learning criterion, the last error is defined as the actual last error made.

The number of observations per point in each of the forward stationarity curves decreases across presentations. In Figure 7 the curves are terminated when the number of observations declines below ten.

The short-term effects of spacing are particularly apparent in Figure 7. In general, the short-term effects superimposed on the learning curves in Figure 6 seem, as implied by the three-state conception of the learning process, to be isolated and magnified in Figure 7. Also, except for the short-term effects, the performance tends to vary around the chance level; i.e., the proportion of correct responses on a presentation following a long interval is, on the average, no greater than chance.

In Figure 8 a short-term retention curve is extracted from the forward stationarity analysis of Figure 7. The proportions of correct responses on presentations prior to the trial of last error are shown as a function of the interpresentation interval (lag). That is, the numbers of correct and incorrect responses following any given lag are summed for all occurrences of that lag in the precriterion data, and the total proportion of correct responses is plotted. There are two features of this curve which merit explicit comment. (1) The curve decreases

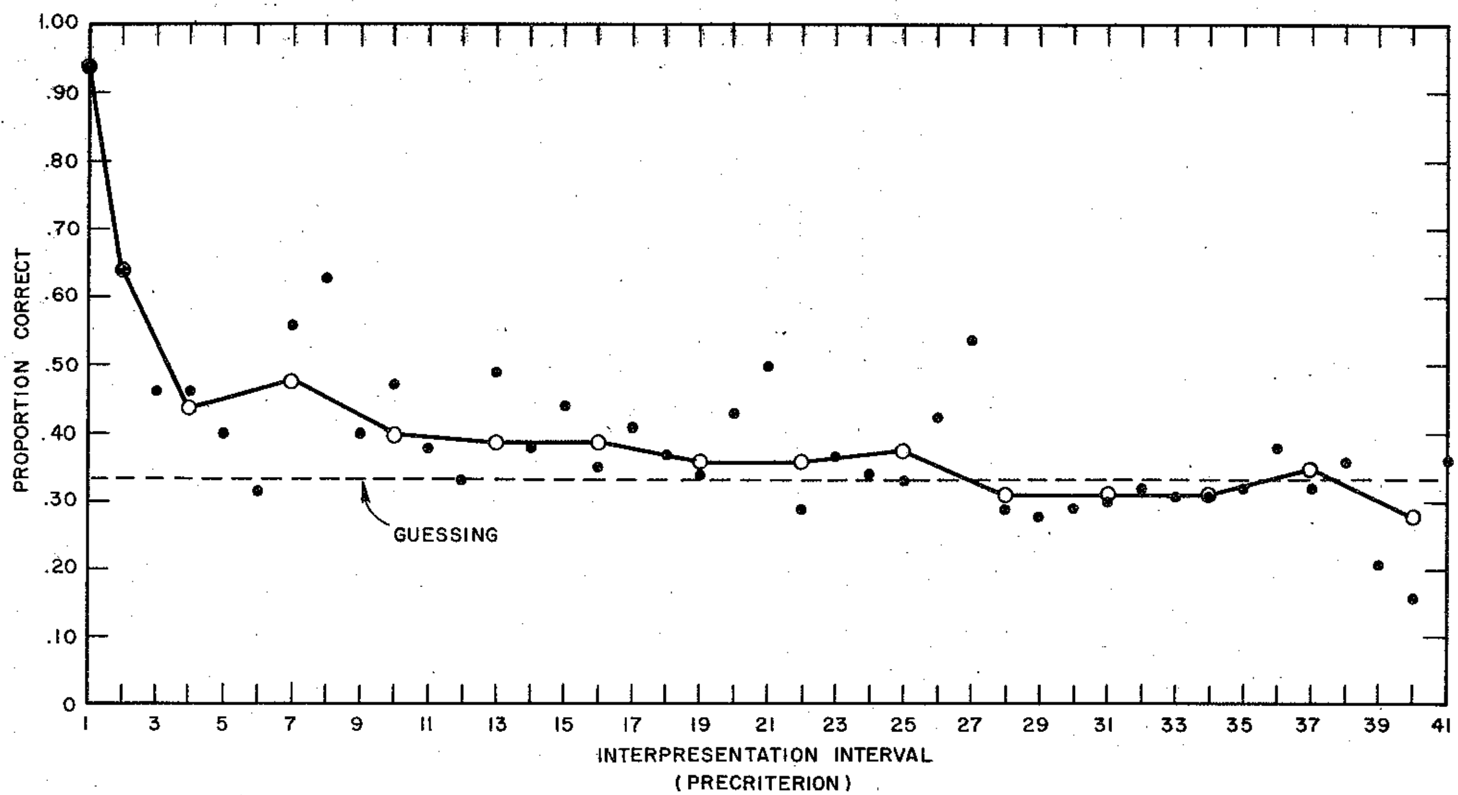


Figure 8. Proportions of Correct Responses Prior to the Last Error as a Function of Interpresentation Interval. (Each open circle plots the average of three successive points.)

more sharply during the first few trials and more slowly during later trials than the best fitting geometric curve. (2) When the precriterion lag is larger than about 21 trials, performance appears to be no better than chance. On the face of it, this chance asymptote of the retention curve offers general support for the notion of all-or-none learned state. If there were any non-transient partial learning of items on the presentations prior to the last error, one would expect the curve in Figure 8 to remain somewhat above chance for all values of the interpresentation interval.

The average latencies of the precriterion successes and precriterion errors are shown in Table 3 for the first seven presentations of each presentation sequence. In contrast to the forward stationarity curves, neither the average latency of a precriterion success nor the average latency of a precriterion error proves very sensitive to interpresentation interval. In Figure 9 the average precriterion success and error latencies are plotted as a function of interpresentation interval. There does appear to be a decrease in the mean latency of a precriterion success when the preceding interval is of length one or two. The precriterion error latency curve also drops at $t_n = 1$, but the $t_n = 1$ point is suspect since only four errors were made following an interval of length one.

A similar analysis of postcriterion data is also shown in Figure 9; that is, the average latency of successes occurring after the last error is also plotted as a function of interpresentation interval. The average latency of a postcriterion success appears to be somewhat more sensitive to the preceding interval than that of a precriterion error or success.

Table 3

Mean Precriterion Success (S) and Error (E) Latencies on Presentations 1 through 7 of the Individual Presentation Sequences.

Present. Sequence	Presentation Number													
	1		2		3		4		5		6		7	
	S	E	S	E	S	E	S	E	S	E	S	E	S	E
1	1.97 (6)	1.63 (37)	1.94 (10)	2.21 (25)	1.92 (7)	1.98 (22)	2.44 (9)	2.48 (16)	1.49 (4)	2.07 (16)	1.41 (4)	1.97 (13)	2.71 (5)	1.73 (10)
2	2.51 (14)	1.87 (23)	3.01 (6)	1.99 (25)	1.99 (15)	1.89 (15)	3.39 (4)	1.87 (16)	2.26 (3)	1.85 (10)	2.33 (5)	1.82 (7)	2.14 (7)	1.80 (2)
3	1.90 (12)	2.11 (27)	2.72 (11)	2.08 (21)	2.59 (10)	2.27 (20)	1.47 (30)	(0)	2.37 (4)	2.70 (18)	2.75 (5)	1.62 (11)	1.97 (1)	2.14 (10)
4	1.91 (21)	1.90 (21)	2.48 (14)	1.77 (19)	1.85 (8)	2.06 (18)	2.06 (5)	2.46 (13)	1.53 (3)	2.19 (5)	1.66 (3)	1.35 (3)	2.03 (2)	.77 (1)
5	1.65 (16)	1.87 (25)	1.75 (11)	2.05 (22)	2.26 (13)	2.46 (15)	1.95 (8)	2.29 (15)	2.07 (9)	1.98 (9)	1.89 (6)	2.21 (10)	2.54 (6)	3.08 (8)
6	1.69 (15)	1.95 (23)	3.03 (10)	2.62 (24)	1.96 (7)	1.90 (22)	1.88 (8)	2.34 (15)	1.81 (8)	2.12 (8)	1.96 (5)	2.33 (8)	1.65 (4)	2.06 (7)
7	2.03 (11)	2.59 (28)	1.51 (11)	2.73 (18)	2.08 (8)	2.27 (15)	2.10 (6)	2.65 (11)	2.76 (6)	2.28 (10)	1.91 (8)	2.43 (6)	1.71 (8)	2.17 (4)
8	1.91 (14)	1.88 (27)	2.08 (9)	2.06 (26)	2.49 (6)	2.23 (22)	2.25 (6)	1.98 (13)	2.11 (7)	1.75 (9)	3.24 (4)	1.60 (9)	2.47 (1)	1.63 (7)
9	2.12 (12)	1.71 (25)	1.83 (5)	1.95 (24)	2.30 (7)	1.93 (16)	1.21 (20)	1.47 (3)	2.13 (12)	1.73 (8)	2.26 (3)	1.92 (12)	2.28 (2)	1.69 (8)
10	2.07 (14)	2.12 (26)	3.01 (10)	2.13 (24)	2.54 (14)	2.09 (15)	3.14 (5)	2.44 (17)	2.02 (15)	2.02 (6)	1.66 (10)	1.58 (7)	1.71 (5)	2.53 (7)
11	1.80 (14)	1.95 (29)	1.90 (14)	2.35 (23)	1.97 (11)	2.44 (18)	1.80 (9)	2.51 (16)	1.87 (14)	1.66 (10)	1.69 (9)	2.06 (9)	2.12 (7)	2.08 (5)
12	1.85 (18)	2.24 (24)	2.16 (12)	2.48 (21)	2.02 (17)	2.54 (11)	2.65 (10)	2.02 (17)	2.57 (5)	1.55 (12)	1.78 (3)	2.17 (9)	1.04 (10)	1.94 (2)
13	1.83 (19)	1.88 (22)	2.06 (13)	1.75 (24)	2.17 (5)	2.45 (19)	2.86 (8)	1.58 (10)	1.92 (4)	2.44 (8)	1.94 (5)	2.64 (4)	2.20 (2)	1.75 (4)
14	1.98 (11)	1.75 (27)	1.98 (8)	2.07 (24)	2.04 (8)	2.51 (19)	2.70 (3)	1.93 (11)	2.05 (6)	1.68 (6)	1.96 (5)	1.60 (5)	1.28 (4)	1.77 (6)
15	1.64 (12)	1.81 (32)	1.82 (15)	2.24 (23)	2.05 (3)	1.97 (22)	1.90 (10)	2.89 (10)	2.23 (8)	1.60 (7)	2.30 (8)	1.84 (3)	2.21 (3)	2.36 (7)
16	2.46 (14)	2.02 (28)	2.19 (12)	2.46 (26)	2.02 (13)	2.73 (21)	1.63 (8)	2.15 (19)	2.50 (15)	2.25 (10)	2.66 (9)	2.17 (10)	2.76 (4)	2.44 (7)
17	1.60 (13)	2.03 (28)	2.26 (8)	1.99 (22)	2.29 (6)	2.07 (15)	1.72 (7)	2.20 (9)	2.09 (5)	1.90 (9)	1.65 (4)	2.28 (7)	1.03 (2)	2.29 (8)
18	2.00 (14)	1.97 (21)	2.20 (9)	1.98 (19)	1.77 (6)	1.96 (15)	1.84 (3)	2.38 (12)	2.36 (4)	1.71 (7)	1.82 (4)	1.93 (4)	(0)	2.20 (6)
19	1.98 (14)	1.86 (29)	1.54 (11)	2.04 (26)	1.72 (6)	2.76 (18)	3.30 (4)	1.76 (14)	2.47 (5)	2.33 (9)	2.08 (2)	1.41 (7)	1.69 (4)	1.36 (4)
20	1.85 (12)	1.78 (28)	2.05 (10)	2.24 (19)	1.90 (11)	1.91 (13)	1.41 (6)	1.74 (11)	1.68 (7)	1.84 (8)	2.01 (4)	1.84 (8)	2.35 (5)	1.42 (6)
21	1.84 (11)	1.99 (27)	2.68 (13)	1.95 (20)	3.09 (9)	1.86 (14)	2.25 (2)	2.73 (10)	2.21 (4)	1.93 (6)	1.08 (10)	(0)	1.82 (5)	1.96 (3)

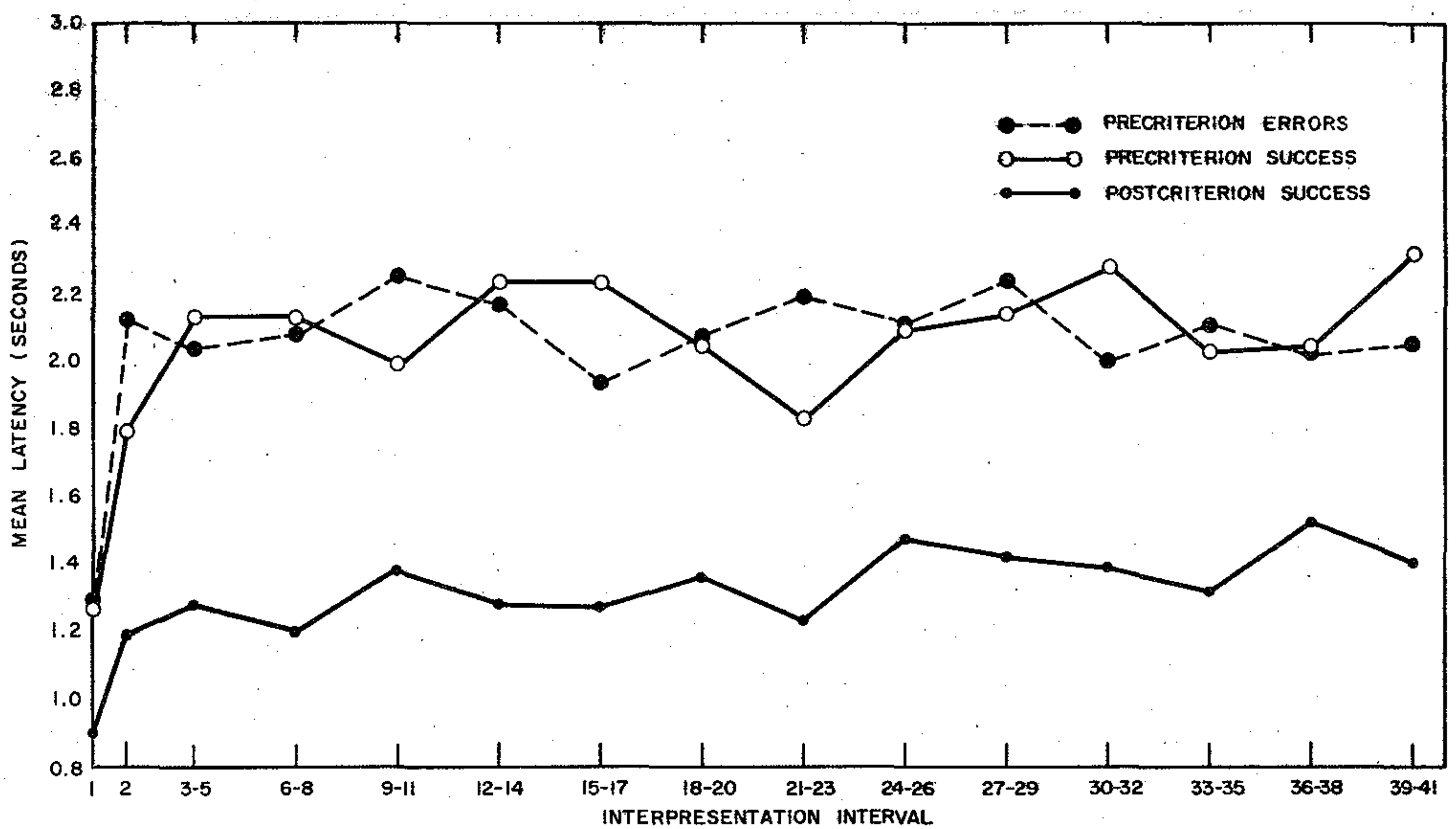


Figure 9. Mean Latencies of Precriterion Errors and Successes and Postcriterion Successes as a Function of Interpresentation Interval.

In the next section, when performance measures are averaged over all presentation sequences, the relationship between the latencies of the precriterion responses and the latencies of the postcriterion responses is discussed further.

Performance Averaged Over All Presentation Sequences

The analyses of the preceding section displayed a considerable variation in performance measures in relation to the individual presentation sequences. In the present section a number of standard analyses are performed on the results pooled over all presentation sequences. The smoothness of the curves in this section illustrates how effectively the striking short-term effects of spacing shown in the preceding section are averaged away in standard analyses.

In Figure 10 the average learning curve and the mean error and success latency curves are shown for all presentation sequences. There are no obvious differences between these curves and those commonly obtained for similar materials and list lengths with the standard anticipation method. The slight sigmoid shape of the learning curve is fairly characteristic of paired associate learning when the list is as long as twenty-one items. The observed pattern of the mean error and success latencies is also quite common.

If any property of the curves in Figure 10 distinguishes the average course of learning in this experiment from that characteristic of standard anticipation method experiments, it is the rate of learning. Compared to the learning rates in some comparable experiments, the learning in the present experiment seems somewhat slow. The difference in rate is slight,

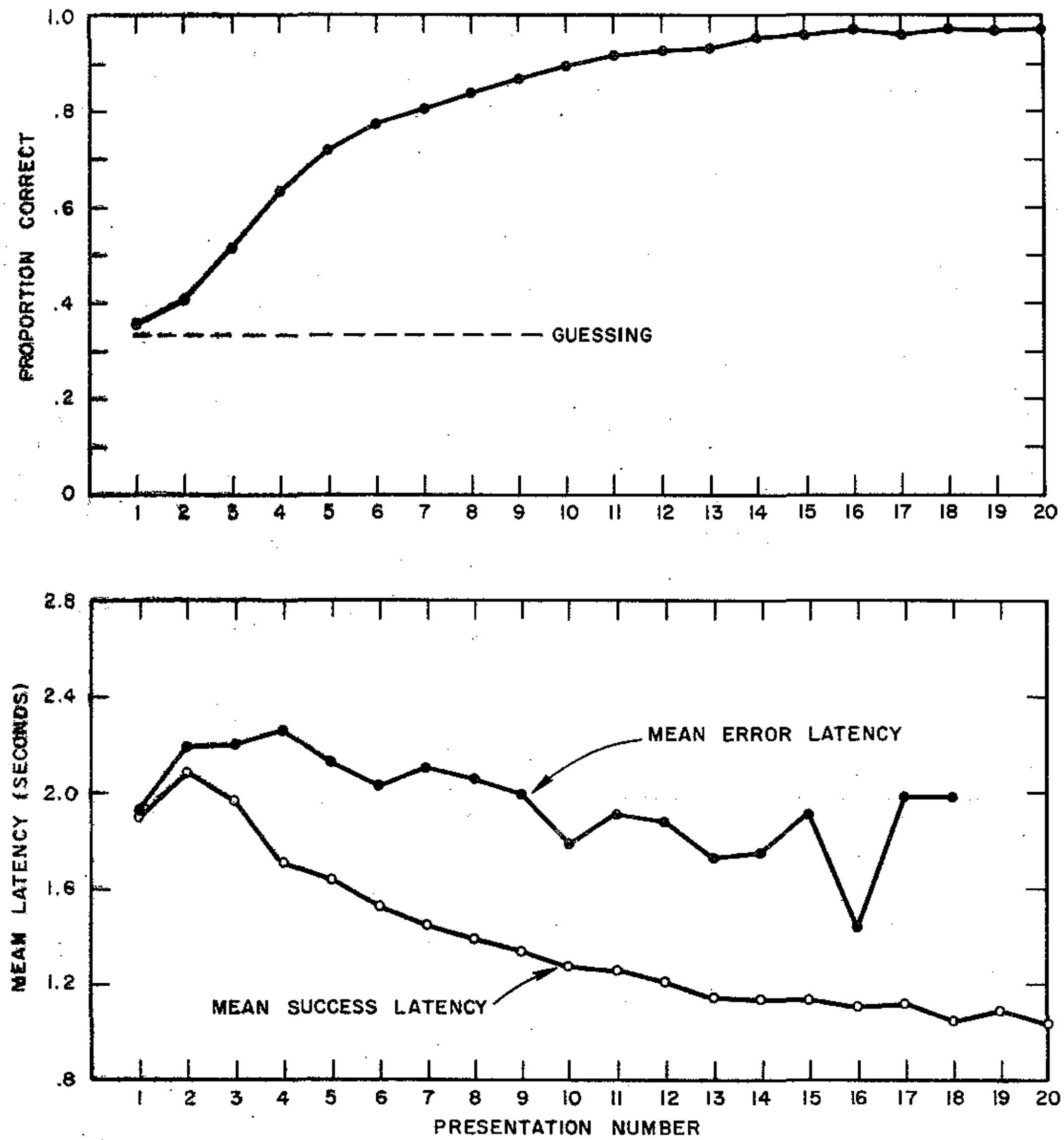


Figure 10. Mean Learning Curve; Mean Success Latency Curve.

but the increased frequency of short intervals in the present experiment might be expected, a priori, to improve overall performance. That is, the overall proportion of correct responses should be artificially elevated somewhat due to the increased number of cases in which correct responses are given to unlearned items on the basis of short-term memory. It could be, however, that the increased occurrences of very short intervals may in fact cause the slower learning. If as suggested by Greeno's results (Figure 1) there is negligible learning on a presentation following a short interval, whatever temporary benefits in performance may result are small compared to the cumulated detriment in performance over later presentations.

The average probability of an error on presentation n given an error on presentation n-1 is shown in Figure 11 for presentation 2-13. As discussed in Chapter 1, the observed decrease of this curve across presentations could reflect an improvement in the learning rate or an improvement in short-term memory or both. It is not clear whether the curve continues to decrease after the first few presentations. An initial decrease in the probability of an error given an error over the first few trials would be expected from the form of the learning curve in Figure 10. It appears as if the first presentation or two do not, proportionately, improve performance as much as later presentations.

Figure 12 exhibits the mean forward stationarity curve. The proportion of correct responses prior to the last error rises from chance on the first presentation (0.34) to 0.44 on the fifth presentation and appears to stay around that level out to the twelfth presentation.

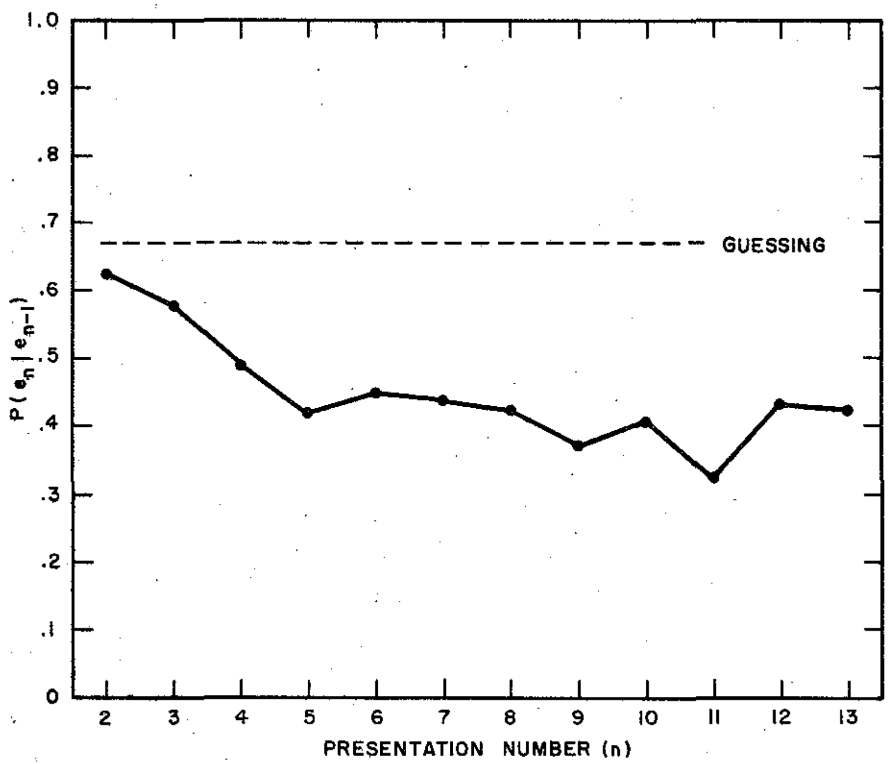


Figure 11. Proportion of Errors on Presentation n Given an Error on Presentation n-1.

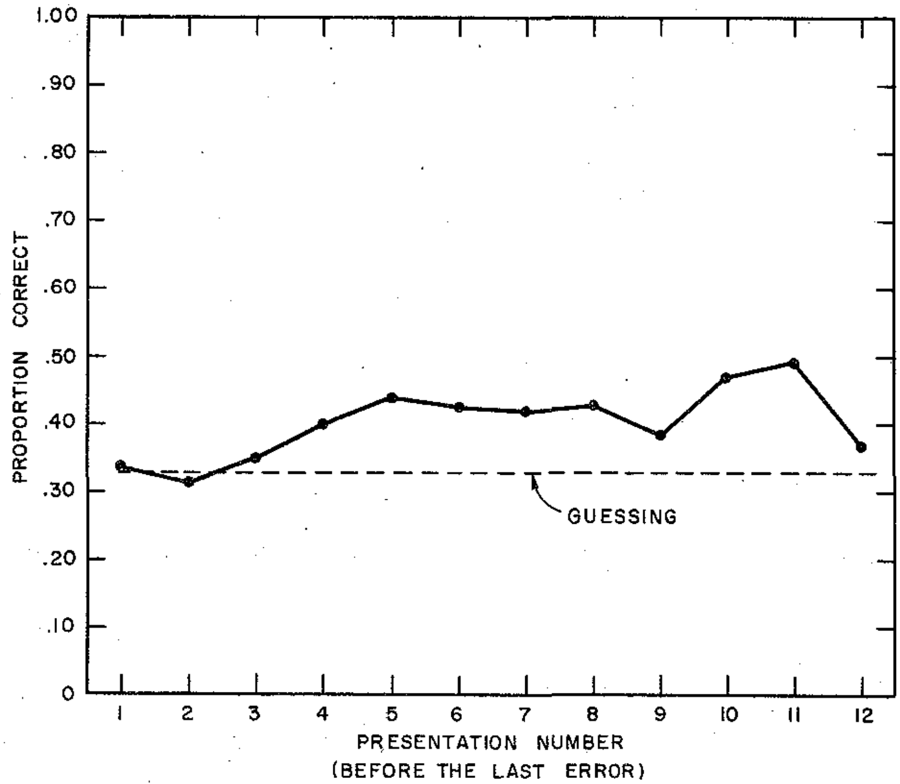


Figure 12. Mean Forward Stationarity Curve.

The forward stationarity curve in Figure 12 together with the learning curve in Figure 10 and the $P(e_n | e_{n-1})$ curve in Figure 11 have several implications with respect to the three-state conception of the learning process. Each of the curves seems to imply that learning or memory processes tend to change across the early presentations. One interpretation is that the state-to-state transition probabilities change over presentations. Another possibility is that the guessing state should be thought of as two states, a "forgotten" state into which items are lost from short-term memory and an initial "uncoded" state in which the item is unfamiliar. If all items start in the uncoded state, and if the probabilities of transition to the learned and short-term states are smaller from the uncoded state than from the forgotten state, one would expect the general pattern of results in Figures 10, 11, and 12.

The average latencies of precriterion errors and successes and the average latency of postcriterion successes across the first thirteen presentations are shown in Figure 13. Two features of the curves in Figure 13 are particularly apparent: (1) the similarity between the precriterion success latency and error latency curves and (2) the striking difference between the precriterion success latency curve and the postcriterion success latency curve. After an initial rise, the mean latency of the precriterion responses decreases only slightly if at all across presentations. The average latency of the postcriterion responses also shows an initial rise, but decreases steadily thereafter and, beyond the first few presentations, remains well below the precriterion curves.

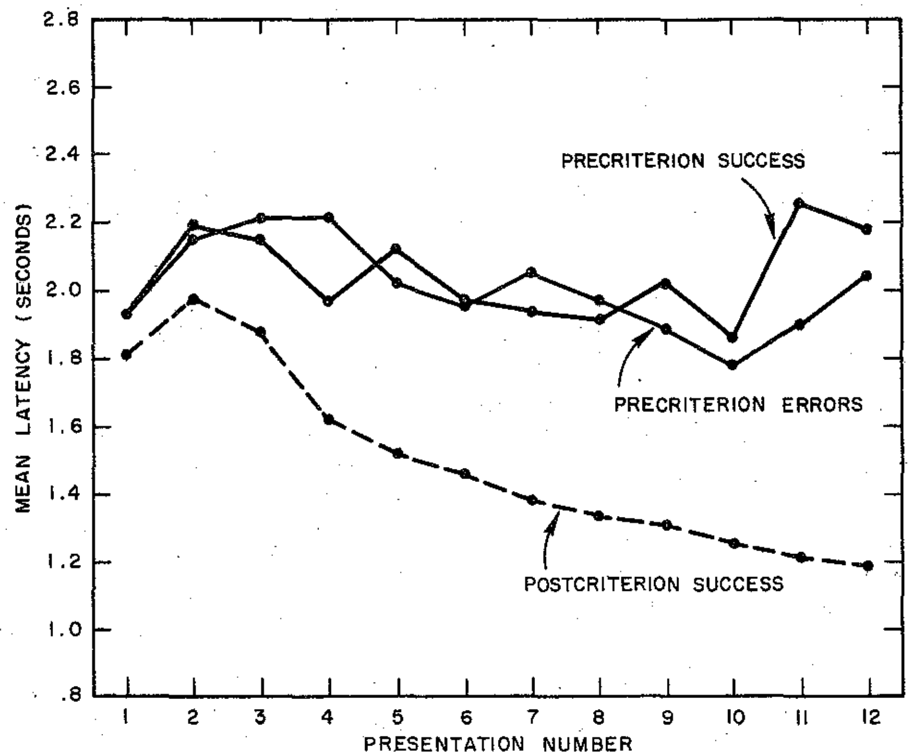


Figure 13. Mean Latency Curves for Precriterion Errors and Successes and Postcriterion Successes.

CHAPTER 4

THEORETICAL ANALYSIS OF THE RESULTS

The results in the preceding chapter suggest that short-term memory plays a significant role during the acquisition of a list of paired associates. They appear to support the distinction between a transient memory system and a relatively permanent memory system. The data also indicate that the early trials differ from the later trials in their effects on performance.

The first section of this chapter introduces a model for the learning of individual paired associates and compares the predictions of the model with the observed results. The second section presents an evaluation of the model along with some comments on short-term memory.

An All-or-None Forgetting Model

Assumptions of the model:

(1) At the start of any trial an item is in exactly one of four states: learned (L), short-term memory (S), forgotten (F), or uncoded ($\emptyset$).

(2) On the first presentation of an item it is in state $\emptyset$ with probability one.

(3) The probability of a correct response is one in states L and S, and g in states F and $\emptyset$. (Usually, $g = 1/r$, where r is the number of response alternatives.)

(4) The probabilities of transition from state to state as a result of a reinforcement (presented below) are constant and independent of path.

(5) The probabilities of transition from state to state as a result

of an intervening trial (presented below) are constant and independent of path:

EFFECT OF A REINFORCEMENT

$$\begin{array}{c} \\ L_N \\ S_N \\ F_N \\ \phi_N \end{array} \begin{array}{c} \begin{array}{cccc} L_{N+1} & S_{N+1} & F_{N+1} & \phi_{N+1} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ a & 1-a & 0 & 0 \\ c & 1-c & 0 & 0 \\ db & d(1-b) & 0 & 1-d \end{bmatrix} \end{array} \quad \begin{array}{c} P(\text{correct} \mid \text{row state}) \\ \begin{bmatrix} 1 \\ 1 \\ g \\ g \end{bmatrix} \end{array}$$

EFFECT OF AN INTERVENING TRIAL

$$\begin{array}{c} \\ L_N \\ S_N \\ F_N \\ \phi_N \end{array} \begin{array}{c} \begin{array}{cccc} L_{N+1} & S_{N+1} & F_{N+1} & \phi_{N+1} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-f & f & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

This model modifies the one-element forgetting model discussed in Chapter 1 in two ways. First, the model allows for different probabilities of transition from state S to state L and from state F to state L. Second, the model assumes that all items start in an initial uncoded[1] state, ϕ. Upon reinforcement of an item in state ϕ,

---

[1] In this feature the model is similar to the LS-3 model postulated by Atkinson and Crothers (1964). It is not intended in naming this state to specify its psychological properties (e.g., Lawrence, 1963). There are a number of possible processes which could explain the differential effects of the early trials. At this point, it does not seem possible on the basis of available evidence to choose among them.

the item leaves state $\emptyset$ with probability d and remains in state $\emptyset$ with probability 1-d. This is equivalent to saying that over all items and subjects the waiting time in state $\emptyset$ is geometrically distributed with parameter d. Given that an item leaves state $\emptyset$ on a presentation, it goes to state L with probability b and to state S with probability 1-b.

The model is consistent, qualitatively, with the results in Chapter 3. It appeared in Chapter 3 that no three-state model with fixed transition probabilities was adequate to account for all the observed results. The results which seemed most inconsistent with such models were the form of the learning curve, the forward stationarity curve, and the $P(e_n|e_{n-1})$ curve across the first several presentations. The assumption in the present model of an initial state $\emptyset$ out of which the probabilities of transition to the learned and short-term states are reduced compared to those out of state F appears, in general, to account for all the differential effects of the early presentations on performance. The sigmoid shape of the learning curve (Figure 6), the decrease in the $P(e_n|e_{n-1})$ curve (Figure 7), and the rise in the forward stationarity curve (Figure 8) might all reflect the existence of such an initial state.

As postulated, the all-or-none forgetting model has five parameters, a, b, c, d, and f. It may be that a two-, three-, or four-parameter special case of the model is more compelling than the general version. If learning out of the short-term memory state is either negligible (a=0) or the same as learning out of the forgotten state (a=c), the number of parameters is reduced to four. Alternatively, if $d = 1$ or $b = c$ the number of parameters is reduced to four. And, of course, various two- and three-paramater special cases can result from combinations of such equivalences; e.g., when $d = 1$ and $a = b = c$, the model reduces to the one-element forgetting model. The reason for postulating the model in the general form is to allow the best-fitting estimates of the parameter values to suggest any of the possible special cases.

Fit of the model: The best simultaneous fit of the all-or-none forgetting model to the twenty-one learning curves in Figure 6 was determined by a least squares method. There are a total of 400 points on the observed learning curves in Figure 6. For any set of values of its parameters the model generates a predicted value for each of the 400 points. One measure of the correspondence between the observed and predicted values is the squared deviation between the observed and predicted values summed over all 400 points. For the all-or-none forgetting model this sum had its minimum (1.016) when $a = 0.000$[1], $b = 0.285$, $c = 0.245$, $f = 0.170$, and $d = 0.410$.

The predicted and observed proportions of correct responses are shown in Table 4 for the first thirteen presentations of each presentation sequence. At the bottom of the Table are the average predicted and observed proportions of correct responses over all presentation sequences. In general, the model accounts quite well for the observed learning curves. The average deviation between pairs of observed and predicted values is 0.050. This average deviation does not seem large, especially

[1]The parameter values were estimated to the nearest 0.005.

Table 4

Predicted (Pre.) and Observed (Obs.) Proportions

of Correct Responses

Present. Sequence		Presentation Number												
		1	2	3	4	5	6	7	8	9	10	11	12	13
1	Pre.	.333	.414	.510	.730	.658	.727	.777	.840	.865	.898	.922	.940	.986
	Obs.	.220	.340	.440	.600	.540	.680	.760	.820	.800	.840	.820	.880	.920
2	Pre.	.333	.420	.726	.541	.637	.715	.795	.828	.864	.896	.924	.939	.956
	Obs.	.360	.380	.680	.480	.640	.820	.900	.820	.900	.940	.940	.960	.980
3	Pre.	.333	.432	.521	.863	.606	.702	.759	.885	.843	.875	.933	.929	.939
	Obs.	.360	.440	.560	1.000	.480	.640	.640	.860	.860	.820	.920	.900	.900
4	Pre.	.333	.412	.517	.595	.677	.748	.826	.845	.972	.892	.922	.951	.959
	Obs.	.520	.440	.500	.580	.680	.880	.900	.940	.980	.980	.900	.960	.960
5	Pre.	.333	.412	.663	.561	.717	.707	.773	.825	.865	.901	.966	.929	.945
	Obs.	.400	.400	.600	.600	.720	.740	.780	.740	.740	.860	.940	.940	.920
6	Pre.	.333	.441	.500	.591	.686	.739	.802	.842	.884	.915	.930	.943	.956
	Obs.	.400	.440	.460	.580	.700	.780	.820	.760	.960	.880	.960	.980	.920
7	Pre.	.333	.415	.505	.597	.703	.778	.807	.831	.873	.900	.924	.950	.957
	Obs.	.320	.440	.580	.660	.780	.840	.880	.780	.900	.940	.960	.920	.940
8	Pre.	.333	.413	.506	.600	.677	.747	.802	.853	.887	.929	.942	.937	.952
	Obs.	.380	.360	.420	.560	.740	.760	.760	.880	.860	.940	.900	.920	.920
9	Pre.	.333	.495	.485	.863	.826	.630	.734	.776	.836	.869	.899	.931	.941
	Obs.	.320	.360	.560	.940	.760	.640	.680	.760	.880	.920	.860	.920	.940
10	Pre.	.333	.412	.663	.560	.877	.705	.739	.798	.842	.920	.923	.919	.960
	Obs.	.320	.400	.600	.520	.860	.780	.760	.800	.840	.920	.960	.920	.960
11	Pre.	.333	.447	.500	.639	.753	.721	.773	.873	.855	.889	.919	.934	.984
	Obs.	.340	.420	.480	.600	.780	.700	.780	.880	.860	.860	.960	.900	.900
12	Pre.	.333	.422	.726	.555	.632	.733	.940	.782	.867	.897	.935	.906	.932
	Obs.	.460	.400	.680	.640	.520	.640	.940	.760	.860	.840	.840	.880	.900
13	Pre.	.333	.550	.472	.575	.658	.748	.798	.909	.969	.861	.919	.938	.933
	Obs.	.420	.440	.360	.680	.720	.840	.840	.880	.940	.940	.920	.940	.940
14	Pre.	.333	.415	.504	.596	.822	.726	.892	.796	.842	.914	.902	.923	.942
	Obs.	.320	.400	.520	.520	.840	.860	.840	.800	.880	.920	.920	.920	.940
15	Pre.	.333	.510	.481	.640	.653	.793	.805	.874	.842	.879	.917	.928	.945
	Obs.	.340	.420	.300	.700	.740	.860	.840	.900	.860	.800	.940	.940	.940
16	Pre.	.333	.461	.493	.589	.760	.728	.781	.833	.869	.956	.915	.930	.947
	Obs.	.360	.400	.500	.480	.760	.680	.640	.820	.920	.940	.920	.940	.960
17	Pre.	.333	.412	.506	.614	.689	.745	.796	.842	.878	.907	.929	.953	.957
	Obs.	.320	.340	.520	.720	.780	.780	.800	.920	.840	.900	.880	.940	.980
18	Pre.	.333	.411	.505	.597	.678	.791	.792	.839	.876	.911	.931	.943	.957
	Obs.	.460	.480	.560	.640	.740	.840	.780	.920	.900	.900	.940	.900	.900
19	Pre.	.333	.432	.502	.613	.669	.741	.799	.858	.876	.905	.962	.980	.941
	Obs.	.380	.360	.380	.600	.740	.740	.880	.900	.820	.900	.960	.980	.940
20	Pre.	.333	.441	.533	.583	.775	.721	.777	.840	.862	.895	.920	.939	.954
	Obs.	.300	.400	.640	.640	.780	.780	.840	.820	.820	.860	.920	.880	.940
21	Pre.	.333	.528	.499	.675	.705	.952	.837	.780	.958	.809	.849	.894	.914
	Obs.	.280	.500	.520	.580	.820	1.000	.880	.840	.900	.920	.940	.920	.860
Mean	Pre.	.333	.443	.539	.627	.708	.743	.800	.836	.877	.896	.923	.935	.950
Mean	Obs.	.361	.408	.517	.634	.720	.775	.807	.838	.872	.896	.919	.926	.931

when it is compared to the average deviation of 0.059 between the observed guessing performance on the first presentation of each of the twenty-one items and the predicted guessing performance, g = 0.333 .

The estimated probability of learning out of the short-term state, a , is essentially zero. This finding is similar, indirectly, to Greeno's results in Figure 1 (1964). Also, the value of c , 0.245, and the value of b , 0.285, are quite close.

The short-term memory perturbations of the individual observed learning curves discussed in Chapter 3 are predicted quite well by the model. Some examples of the close correspondence between the predicted and observed spikes in the learning curves occur on the fourth presentation of sequence 9, the third and fifth presentations of sequence 10, and the third and seventh presentations of sequence 12.

There does not seem to be a clear systematic nature to the deviations between the predicted and observed values in Table 4. Pooled over all the presentation sequences, the average predicted and average observed proportions of correct responses are quite close.

Another comparison by which to test the model's adequacy is the predicted versus observed forward stationarity curves. Unfortunately, the exact predicted forward stationarity curve cannot be computed. Since last errors are more likely to follow a long interpresentation interval than a short one, it is not reasonable to assume that the distribution of precriterion intervals is uniform. In general, however, the predicted forward stationarity curve starts at chance on the first presentation and negatively accelerates, at a rate governed by the value of d, to an asymptote slightly higher than

$$g + (1-g) \sum_{k=1}^{41} P(t_n=k)(1-f)^{k-1} = g + (1-g)(1/41) \sum_{k=1}^{41} (1-f)^{k-1}$$

$$\approx g + (1-g)(1/41)(1/f) .$$

With $f = 0.17$, this value is 0.42. There seems to be a good qualitative correspondence between the observed forward stationarity curve in Figure 7 and the approximate predicted curve.

A learning curve can be thought of as giving the probability of each of two error-success one-tuples on each presentation, i.e., the marginal probability of an error and of a success. The next finer level of analysis, and one that is more sensitive to the first order transition probabilities, is to investigate the probabilities of the four possible error-success two-tuples on presentations n and $n+1$. It is possible for a model to fit an observed learning curve well and to fit the observed error-success two-tuples poorly.

For presentations 2 through 13 of each of the twenty-one presentation sequences, the observed frequency of each of the four two-tuples were extracted from the subject-item protocols. That is, the observed frequencies of the four events, $e_n e_{n+1}$, $e_n s_{n+1}$, $s_n e_{n+1}$, $s_n s_{n+1}$, were computed for $n = 2, 3, \ldots, 12$. These observed proportions were compared with those predicted by the all-or-none forgetting model. A minimum chi-square procedure was used to find the best-fitting parameter values of the model. For a given set of parameter values, the quantity,

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

was computed where O and E are the observed and predicted two-tuple frequencies and the sum is over all two-tuples, trials, and presentation sequences. The minimum value reached by this quantity and the values of the paramaters which yielded the minimum are shown below.

$$\min \chi^2 = 793.7 \text{ ; d.f.} = 688 \text{ .}$$

$$a = 0.00 \text{ , } c = 0.21 \text{ , } b = 0.48 \text{ , } d = 0.29 \text{ , and } f = 0.08 \text{ .}$$

It is a misnomer to call the above quantity a minimum chi-square. Since the components of the sum are not independent, the sum is not chi-square distributed. Also, many of the predicted frequencies are so small that the sum is artificially elevated. The procedure was used primarily to estimate parameters in order to compare the predicted and observed two-tuple frequencies pooled over all presentation sequences. In Table 5 the average predicted and observed values of $P(s_n s_{n+1})$, $P(s_n e_{n+1})$, $P(e_n s_{n+1})$, and $P(e_n e_{n+1})$ are shown for $n = 2, 3, \ldots, 12$. The observed and predicted conditional probabilities of an error on presentation n ,

$$P(e_{n+1} | e_n) = \frac{P(e_{n+1} e_n)}{P(e_n)}$$

are also shown in Table 5.

The predicted and observed proportions in Table 5 are quite close. There do, however, appear to be systematic deviations. The predicted $P(s_n s_{n+1})$ values start above the observed $P(s_n s_{n+1})$ values, but from $n = 5$ to $n = 12$ they are slightly but consistently below the

Table 5

Predicted (Pre.) and Observed (Obs.) Proportions

of Error-Success Two-tuples

	Presentations n, n+1										
	2,3	3,4	4,5	5,6	6,7	7,8	8,9	9,10	10,11	11,12	12,13
$P(s_n s_{n+1})$											
Pre.	.290	.403	.509	.585	.646	.698	.747	.785	.824	.852	.869
Obs.	.267	.389	.510	.621	.691	.727	.770	.820	.850	.880	.889
$P(s_n e_{n+1})$											
Pre.	.168	.145	.121	.119	.093	.086	.070	.062	.047	.049	.038
Obs.	.141	.129	.125	.099	.084	.080	.068	.052	.047	.039	.037
$P(e_n s_{n+1})$											
Pre.	.257	.227	.195	.154	.138	.118	.100	.086	.077	.055	.053
Obs.	.250	.246	.210	.154	.115	.111	.102	.076	.070	.046	.043
$P(e_n e_{n+1})$											
Pre.	.284	.225	.175	.142	.123	.097	.083	.067	.052	.043	.039
Obs.	.342	.237	.155	.126	.110	.082	.060	.051	.034	.035	.031
$P(e_{n+1} \mid e_n)$											
Pre.	.525	.500	.475	.481	.467	.453	.451	.441	.396	.433	.422
Obs.	.578	.491	.425	.449	.489	.423	.371	.409	.329	.436	.424

observed values. Conversely, the predicted $P(e_n e_{n+1})$ values start below the observed $P(e_n e_{n+1})$ values, but for $n = 4$ to $n = 12$ they are slightly but consistently above the observed values. The deviation between the predicted and observed values of $P(s_n e_{n+1})$ and $P(e_n s_{n+1})$ do not appear to be systematic.

The parameter values estimated from the two-tuple frequencies are quite different from those estimated from the learning curves. Even though one would not expect the values to be the same since the estimation procedures were quite different, the variance between the sets of values still seems large. In a sense, however, the values are not as different as they appear. There are obvious trading relationships between several of the parameters in the model. For example, though the values of b, 0.285, and d, 0.410, estimated from the learning curves are quite different from the values of b, 0.48, and d, 0.29, estimated from the two-tuples, the products of b and d are quite close (0.117 versus 0.138). That is, for each set of values of b and d, the probability, bd, of learning out of state $\emptyset$ is comparable.

When the parameter values estimated from the learning curves are used to predict the two-tuple frequencies, the resulting chi-square is 917.8. This value is quite large compared to the minimum value, 793.7. One reason for the increase is that the systematic deviations between the predicted and observed values of $P(s_n s_{n+1})$ and $P(e_n e_{n+1})$ are increased. These deviations are not reflected in the learning curve analysis because they tend to cancel each other.

When the parameter values estimated from the two-tuple frequencies are used to predict the learning curves, the resulting sum of the squared deviations is 1.191, which means that the average deviation between pairs of predicted and observed values is 0.055.

Evaluation of the Model: Some Comments on Short-Term Memory

It is somewhat surprising that the all-or-none forgetting model accounts as well as it does for the empirical results. For several reasons the short-term memory structure of the model seems inadequate to account for all the effects of spacing.

(1) The model assumes that with each intervening trial there is some probability, f, that an item in short-term memory is forgotten. This all-or-none forgetting assumption is very tractable, mathematically, but constrains the short-term retention curve to be geometric in shape. Empirical short-term memory curves tend not to be geometric in form.

(2) The model predicts that long-term performance improves monotonically as the spacing between two presentations is increased. This runs counter to the results obtained both by Peterson and by Young, as shown in Figure 3. Their results suggest that there is a limit to the improvement in long-term performance with the spacing of two presentations, and that after a certain point long-term performance tends to decline with an increase in the interval.

(3) The model predicts not only that long-term performance improves with the interval between presentations, but also that short-term performance improves with the interval between two presentations. Peterson's results in Figure 2 suggest that although long-term performance is better with spaced presentations, short-term performance is better

with massed presentations. The crossed curves in Figure 2 imply a more elaborate short-term memory mechanism than that assumed by the all-or-none forgetting model.

All of the above findings which are inconsistent with the general implications of the model come from miniature experiments. The experimental task facing the subject in these experiments is quite different from the learning of a list of paired associates. There is a constant introduction of new items and disappearance of old items in these experiments, and the typical range of intervals utilized is less than in the present experiment. It is, nonetheless, likely that many of the effects of spacing revealed by miniature experiments are present in paired-associate list learning. The empirical results of the present study in fact suggest that short-term retention effects during paired-associate list learning are similar in their general properties to those in short-term retention experiments.

It is likely, however, that the small second-order effects of spacing shown in Figures 2 and 3, though present in this experiment, may be insignificant compared to the obvious first-order effects. Thus, although there is evidence for some complex effects of spacing not postulated by the model, these effects may not occur in any magnitude in paired-associate list learning. As a model for the learning of the individual members of a list of paired associates, the all-or-none forgetting model seems to account for the basic features of memory and learning during the acquisition of the list.

APPENDIX

THE UNIFORM LAG ALGORITHM

This appendix presents a procedural algorithm designed to generate a series of M trials on the members of a list of L items such that the distribution of interpresentation intervals is uniform over the range, $t_n = 1$ to $t_n = 2L-1$.

Description

To facilitate the description of the algorithm, it is necessary to add to the notation introduced at the start of Chapter 2.

$N_{k,n}$:	The trial number of the n^{th} presentation of the k^{th} item.
$(N_{k,1}, N_{k,2}, \ldots, N_{k,n})$:	The presentation sequence of the k^{th} item.
$[N+1, N+2, \ldots, N+(2L-1)]$:	The block of $2L-1$ trials following trial N .

The algorithm assigns a presentation sequence to each member of the list in turn. That is, all presentations of item k are determined before any of the presentations of items $k+1, k+2, \ldots, L$ are determined. The basic steps of the procedure are summarized in the following outline.

I. For $k = 1, 2, \ldots, L-1$,

A. $N_{k,1}$ = the first open trial[1] in the block $[1, 2, \ldots, 2L-1]$.

[1]An "open" trial is one not already assigned to present (filled with) one of the preceding $k-1$ items.

B. For $n = 2, 3, \ldots,$ $N_{k,n}$ is determined as follows:

1. Some trial, N, is chosen at random from the block $[N_{k,n-1}+1, N_{k,n-1}+2, \ldots, N_{k,n-1}+(2L-1)]$.
2. $N_{k,n} = N$ if every block of $2L-1$ trials containing N has $L-k$ or more open trials, not counting trial N.[1]
3. If the condition in (2) is not satisfied, another trial N is chosen and so on until the condition is satisfied.

C. When for some value of n, $N_{k,n} > M$, then the presentation sequence for item k is complete, terminating with $N_{k,n-1}$.

II. The last item in the list, $k = L$, fills all the remaining open trials; i.e., $N_{L,1}, N_{L,2}, \ldots$ equal, respectively, the trials from 1 to M not assigned to one of the items, 1, 2, ..., $L-1$.

A program was written by Robert Miller at the Institute for Mathematical Studies in the Social Sciences, Stanford University, to computer-implement the algorithm. Given a list of items, the program will generate any specified number of series of trials. For each series, the program outputs (1) the series itself, (2) the number of presentations of each item, (3) the sequence of intervals for each item, (4) the distribution of intervals between presentations n and $n+1$ of all items, and

---

[1] This condition assures that trials are available for the presentations of later items in the list. Since no interpresentation can exceed $2L-1$ trials, there must be at least one presentation of each item in any block of $2L-1$ trials. If, after the k^{th} item has been assigned a presentation sequence, some block of $2L-1$ trials contains less than $L-k$ open trials, it is impossible for each of the remaining $L-k$ items to be presented in that block of trials.

(5) the total distribution of interpresentation intervals.

Uses and Distributional Properties

The algorithm can be used in a number of experimental situations in addition to list learning by the anticipation method. For example, it can be used to generate a series of study (R) trials and test (T) trials for the study of list learning by the RT procedure; the members of a given presentation sequence can be arbitrarily designated as study trials or test trials on a particular item. The procedure is also applicable to a variety of steady-state procedures in the study of short-term recall and recognition memory. Katz (1966) used the algorithm to generate a series of anticipation trials in which the response reinforced to a particular stimulus was, with a high probability, changed on each successive occurrence of the stimulus in the series. Subjects were required to remember the current response for each stimulus. Herman Buschke (personal communication) at the Stanford University Medical Center has also made use of the algorithm in a steady-state experiment. In his experiment subjects were required to estimate, at each stimulus presentation, how many trials had elapsed since the last presentation of that stimulus.

The exact distributional properties of the algorithm are very difficult to ascertain. It is not clear how to solve for the distributional properties analytically. In practice, the algorithm generates series of trials with the following descriptive properties.

(1) The total distribution of interpresentation intervals appears uniform.

(2) There seems to be a negligible correlation between the lengths

of any two successive intervals, t_n and t_{n+1}, for a given item.

(3) The last item, on the average, is presented more often than the other items; it tends to have more short than long interpresentation intervals. This is the only apparent bias in the procedure. Generally, the last item is presented 10 to 30 per cent more often than the average number of presentations of the other items. Thus, since the total distribution of intervals is uniform, the other items have slightly more long intervals than short intervals. The difference, however, does not seem noticeable.

REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. Psychol. Monogr., 1960, 74 (10, Whole No. 497).

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

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

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

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

Estes, W. K. New developments in statistical behavior theory: differential tests of axioms for associative learning. Psychometrika, 1961, 26, 73-84.

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

Holland, P. W. A variation on the minimum chi-square test. Tech. Rep. No. 85, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965.

Izawa, C and Estes, W. K. Reinforcement-test sequences in paired-associate learning. Tech. Rep. No. 76, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965.

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

Lawrence, D. H. The nature of a stimulus: some relationships between learning and perception. In S. Koch (Ed.) *Psychology: a study of a science*, Vol. 5. New York: McGraw-Hill, 1963. Pp. 179-212.

Peterson, L. R., Hillner, K. and Saltzman, D. Supplementary report: time between pairings and short-term retention. *J. exp. Psychol.*, 1962, *64*, 550-551.

Peterson, L. R., Wampler, R., Kirkpatrick, M. and Saltzman, D. The effect of spacing presentations on retention of a paired associate over short intervals. *J. exp. Psychol.*, 1963, *66*, 206-209.

Yellott, J. I. Some effects of noncontingent success in human probability learning. Tech. Rep. No. 89, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965.

Young, J. L. Effects of intervals between reinforcements and test trials in paired-associate learning. Tech. Rep. No. 101, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1966.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Stanford University
Institute for Mathematical Studies in the Social Sciences, Stanford, California 94305

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE
Learning and Short-Term Retention of Paired Associates in Relation to Specific Sequences of Interpresentation Intervals

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report, August, 1966

5. AUTHOR(S) (Last name, first name, initial)
Bjork, Robert A.

6. REPORT DATE
August 11, 1966

7a. TOTAL NO. OF PAGES
63

7b. NO. OF REFS
15

8a. CONTRACT OR GRANT NO.
Nonr-225(73)

b. PROJECT NO.
NR 154 218

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
Technical Report No. 106

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Office of Naval Research
Personnel and Training Branch, Code 458
Washington, D. C. 20360

13. ABSTRACT
This dissertation reports a study of short-term retention in paired-associate list learning. One purpose of the study was to ascertain empirically the extent to which short-term memory influences performance during the acquisition of a list of paired associates; a second was to gather evidence with respect to the conceptual relationship of short-term memory and learning.

The particular experimental behavior chosen for study was the learning of a list of paired associates by means of a series of anticipation trials. An anticipation trial starts with the presentation of a stimulus to which the subject attempts to anticipate the correct response and ends with a presentation of the correct response. During an experimental session, the trials on any one item (presentation sequence) are characterized by a sequence of interpresentation intervals $(t_1, t_2, \ldots, t_n)$; that is, any two successive presentations $(n, n+1)$ of an item are separated by some number of trials (t_n) on other items.

The experimental design modified the standard anticipation procedure in two ways. (1) The series of trials was generated by a computer-implemented algorithm designed to yield a uniform distribution of interpresentation intervals; for a list of length L, t_n takes on the values $1, 2, \ldots, 2L-1$ equally often. (2) All subjects had the same series of trials in the sense that each had the same set of presentation sequences. That is, every subject had exactly one item in the list assigned to each of L specific $(t_1, t_2, \ldots, t_n)$ sequences. The confounding of item differences with the effects of the presentation sequences was avoided by counterbalancing across subjects the assignment of items to presentation sequences.

DD FORM 1473 1 JAN 64 (cont. on other page)

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Paired-associate learning Short-term retention Interpresentation interval All-or-none learning Initial uncoded state						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ______________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ______________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ______________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

13. ABSTRACT (cont.)

Fifty subjects (college freshmen) were each given a series of 400 trials on a list of 21 items. Nonsense syllables served as stimuli, and the three digits, 3, 5, and 7, served as responses. Each of the responses was paired with seven of the stimuli.

The experimental design allowed performance measures to be examined as a function of the 21 specific presentation sequences. In general, the observed performance on a presentation was quite sensitive to the preceding inter-presentation interval. The 21 learning curves showed striking temporary increments due to short-term memory. The temporary nature of these increments was emphasized by their spike-like appearance; if an interval of even moderate length followed a very short interval, the sudden increase in the learning curve was followed by a distinct decrease.

The latency of correct responses was similarly sensitive to the inter-presentation interval. The 21 mean success latency curves tended to show a negative spike whenever the learning curve showed a positive spike. In contrast, the mean error latencies did not seem clearly sensitive to the interpresentation interval.

Consideration of the results led to formulation of an all-or-none forgetting model for paired-associate learning which distinguishes two memory systems, one transient and the other permanent. Several analyses of performance prior to the last error were carried out to test the implication of the model that, prior to the last error, performance should be only at the chance level except for short-term memory effects. The observed learning appeared to support the distinction between the all-or-none learned state and a short-term memory state.

A number of measures of the obtained learning, in particular the 21 learning curves, were found to be consistent with the predictions of the all-or-none forgetting model.