8-1-1970

A Three State Markov Model for Paired Associate Learning

Thomas R. Linscheid

A THREE STATE MARKOV MODEL FOR
PAIRED-ASSOCIATE LEARNING

by
Thomas R. Linscheid

Bachelor of Arts, Jamestown College 1966
Master of Arts, University of North Dakota 1969

A Dissertation

Submitted to the Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota

August
1970

This dissertation submitted by Thomas R. Linscheid in partial fulfillment of the requirements for the Degree of Philosophy from the University of North Dakota is hereby approved by the Faculty Advisory Committee under whom the work has been done.

(Chairman)

Dean of the Graduate School

Permission

Title A THREE STATE MARKOV MODEL FOR PAIRED-ASSOCIATE LEARNING

Department Psychology

Degree Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the Chairman of the Department or the Dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Signature ____________________

Date ____________________

ACKNOWLEDGMENTS

The author wishes to express his appreciation to the members of his committee, Drs. Robert S. Beecroft, Ralph H. Kolstoe, L. A. Stone, John B. Carman and A. W. Sturges. A very special thanks is extended to Dr. Beecroft for his generous offering of time and advice throughout the course of this project. His interest and encouragement helped to make this paper enjoyable as well as instructive.

I also wish to thank my wife, Linda, for her encouragement, understanding and self-sacrifice, not only during the preparation of this paper, but throughout my graduate career. I wish to dedicate this paper to my parents, Dr. and Mrs. E. C. Linscheid, whose patience and understanding are in great part responsible for my academic progress.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The continuity-noncontinuity issue remains a focus for both the theoretical and experimental inquiry into the fundamental nature of the learning process. Although experimental studies of paired-associate learning have often disclosed continuities, mathematical models incorporating all-or-none processes have generally fit the data quite well. The most basic of these is the one-element model proposed by Bower in 1961. The model assumes only two states, a learned state and an unlearned state. Transitions from the unlearned to the learned state occur with a fixed probability which is constant across trials. Extensions and modifications of the Bower model have consisted basically of the addition of intermediate states which have their owned fixed transition probabilities. Theoretical explanations of these intermediate states include short-term memory stores, discrimination processes, and recognition-recall differences.

Previous evidence has shown that Bower's model loses its accuracy of prediction with difficult and/or long lists. In an effort to predict accurately both quickly and slowly learned lists an all-or-none three state model was built. The model, called the paired-associate recognition-recall (PARR) model, was based upon established differences between recognition and recall learning. The first state is a nonrecognition-nonrecall state in which the probability of a correct response is zero. The intermediate state or recognition

state contains paired-associate items which can be recognized but are not yet recalled. While in the recognition state an item may be selected for rehearsal with probability p, in which case a correct response will be given. The third state is the recall or learned state in which pairs are correctly recalled on every trial. Unlike many previous models direct transitions from the first state to the third state are possible. Also, the probability of moving into the recall state from the recognition state is independent of the probability that an item is rehearsed. Predictions for the learning curves, errors before the first correct response (J), total errors (T), and last error trial (L) were derived and tested against obtained data. Predictions from the Bower model and from a model for discrete performance levels by Bower and Theios were also compared to the data.

List difficulty was varied by manipulating stimulus term meaningfulness. CVC's selected from Archer's 1960 list were used to build low, medium, and high meaningfulness lists. Response terms were the digits 1-16 for each list.

None of the models tested adequately described data from the three meaningfulness conditions. In all cases the models predicted a more rapid rate of learning than was observed. The Bower-Theios and one-element models made very similar predictions about the learning curves but were very dissimilar in their predictions of the probability distributions of J, T, and L. Data from the high meaningfulness list indicated that an intermediate state did not exist. Since the Bower-Theios and PARR models are intermediate state models only the one-element model was used in the consideration of the high meaningfulness

data. Surprisingly, the one-element model provided a very bad fit to the data; predicting a much more rapid learning rate than was observed.

Results were discussed in terms of parameter estimates, the failure of the one-element model, and with regard to the conventional two-stage analysis of paired-associate learning.

CHAPTER I

INTRODUCTION

The formulation of learning theories is by no means new. In the years between the turn of the century and the 1950's numerous theoretical approaches to learning arose. The most well known was Clark Hull's (1943) comprehensive and elaborate theory of learning. Theories of learning, while they are formulated for the purpose of understanding learning, find their ultimate test and usefulness in whether or not they can predict data. Hull's system with its numerous definable but unmeasurable hypothetical constructs (e.g., ${}_{s}H_{r}$, I_{r}) could be used only to make, at best, limited ordinal predictions about data.

Reaction to the Hullian-type theories came mostly from men like B. F. Skinner and W. K. Estes. Skinner (1950) argued against the formulation of elaborate theories which are based upon undemonstrable hypothetical constructs. He suggested that psychologists concentrate on studying observable stimulus-response relationships and that response probability should be the basic dependent variable. This emphasis upon response probability had a profound effect upon later formulation of learning theory in mathematical terms. Estes (1950) felt that progress toward general agreement among learning theorists would be slow as long as theories were built upon verbally defined hypothetical constructs. He suggested "the possibility of agreement on a theoretical framework, at least in some intensely studied areas, may be maximized by defining concepts in terms of experimentally manipulable variables, and developing the consequences of assumptions by strict mathematical reasoning" (p. 94). He described several studies which were used to develop a statistical theory of elementary learning processes. Since Estes 1950 article mathematical models have thrived and been extended to many other areas of psychology.

At this point it would be wise to examine just what a mathematical theory of learning really is. Atkinson, Bower and Crothers (1965) define mathematical learning theory as "the conduct of theorizing and research on learning by explicit mathematical means" (p. 1). They explain that if mathematics is the application of rigorous logical thinking then any scientist who states and derives his theory with precision and logic is applying mathematics to his own science. They point out that the word mathematical refers to the method of theorizing and not to the substantive ideas expressed in a theory. To Bush and Estes (1959), mathematical learning theory "denotes the growing body of research methods and results concerned with the conceptual representation of learning phenomenon, the mathematical formulation of assumptions or hypotheses about learning, and the derivation of testable theorems" (p. 3). This definition provides a better description of what the theorist actually does. Ideas, hypotheses, or just plain guesses about the learning process are translated into mathematical language and are used to derive predictions about data obtained in learning experiments. If the predictions are found to be accurate the theorist has gained information as to the validity of his original formulations; if the predictions are found to be inaccurate the theorist must reformulate or reconsider his view of the learning process.

Almost all mathematical learning theories assume an underlying stochastic process. A stochastic process is characterized by a temporal sequence of events that can be analyzed by using probability theory. A distinction should be made here between probability theories and deterministic theories. With a deterministic theory if certain circumstances are met an event will occur. With a probability theory, however, only the likelihood of an event can be predicted regardless of the amount of information available. There are basically two types of stochastic processes: the independent trials Bernoulli process and the Markov chain process. A sequence of responses is a Bernoulli process if the probability that a given event in the sequence will occur is independent of the outcomes of preceding trials and independent of the trial number. A Markov chain process is satisfied if the probability of the occurrence of an event in the sequence is dependent only upon the preceding event or trial in the sequence.

Mathematical learning theory has been used widely by both sides in the long standing controversy over whether learning is all-or-none or incremental. The all-or-none position maintains that the learning of an association occurs all-at-once on a single trial whereas the incremental or continuity position states that an association is learned by the gradual build up of associative strength. A mathematical theory of paired-associate learning based upon the all-or-none position was proposed by Bower (1961). In this model an item is in one of two states, an unlearned state or a learned state. In the unlearned state the probability of a success is the same as the probability of a correct guess (g) since the item is yet unlearned. On each trial an item either moves into the learned state with probability, c, or remains in the unlearned state. Once an item is in the learned state its probability of being correct is one. With this simple formulation Bower developed predictions concerning such "fine grain" aspects of the data as mean trial of last error, distribution of total errors, distribution of last error trials, and many others.

Despite its success the Bower model has some difficulties, such as accounting for experimentally-produced variations in learning rate. The purpose of this paper is to examine the logical extensions of Bower's model and to build an all-or-none based Markov model that will describe data regardless of the learning rates under which the data are obtained. Recent findings concerning the role of recall and recognition in learning are considered in the formulation of the model to be built and tested.

CHAPTER II

REVIEW OF THE LITERATURE

Bower's One-Element Model

A variation of the traditional stimulus sampling theory developed by Estes (1950, 1959) provides the basis for all the models which will be described in this chapter. Stimulus sampling theory assumes that a stimulus is composed of discrete elements and postulates that an organism draws a sample from this population on each trial during learning. Only a certain number of these elements are selected and conditioned to the response on each trial. Response probability is related to the proportion of elements that have been selected and conditioned. A special case of this theory involves regarding the stimulus as just one element (Estes, 1960) and then assuming that it is sampled on every trial and conditioned to the response with some fixed probability. Although Bush and Mosteller (1959) were the first to use this one-element idea to build a model for learning, they did not fully investigate the properties of the model. It was Bower (1961) who formalized and applied the one-element idea to paired-associate learning. Bower felt that the true test of a model was its ability to describe the response sequences of paired-associate items. His intent was to build a model which was theoretically simple but which would allow for the derivation of an extensive number of predictions about response sequences. Bower's formulation was a two-state model characterized by the following axioms:

1. Each item may be represented by a single stimulus element which is sampled on every trial.
2. This element is in either of two conditioning states: C_1 (conditioned to the correct response) or C_0 (not conditioned).
3. On each reinforced trial the probability of a transition from C_0 to C_1 is a constant, c, and the probability of a transition from C_1 to C_1 is 1.
4. If the element is in state C_1 then the probability of a correct response is 1; if the element is in state C_0, then the probability of a correct response is 1/N, where N is the number of response alternatives.
5. The probability, c, is independent of the trial number and the outcomes of preceding trials (p. 258).

It can be seen that the first axiom gives the model its one-element characteristic and that the remaining axioms make the model all-or-none since an item is either learned completely or not at all. Using these axioms, transition probabilities can be determined and are given in the following matrix:

		Trial N+1	
		C_1	C_0
Trial N	C_1	1	0
	C_0	c	1-c

Examining the matrix it can be seen if an item is in state C_0 on trial N then with probability c it will move into state C_1 on trial N+1; otherwise with probability 1-c it will remain in C_0. Once an item is in C_1 it remains there and cannot return to C_0. From this basic design Bower derived expressions to predict many of the characteristics of the response sequences. For example, his formula for predicting the mean last error trial is, $E(L) = \frac{(1-g)b}{c}$, where $b = \frac{c}{1-g(1-c)}$. The formula for the probability distribution of total errors is, $Pr(T=k) = (1-gb)(1-b)^{k-1}b$, (notations and formulas are from Atkinson, Bower and Crothers, 1965, Ch. 3).

The utility of Bower's model lies in the numerous predictions about response sequences which can be derived using only one parameter, c. This learning rate constant is estimated very simply by proper manipulation of the formula $\overline{T} = (1-g)/c$ where $\overline{T}$ is equal to the observed mean total number of errors for each subject-item and g is equal to the probability of a correct guess (one divided by the number of response alternatives). Bower demonstrated a very satisfactory fit of his model to paired-associate data obtained from ten item lists which used consonant pairs as stimuli and the integers "one" and "two" as responses. However, in a comparison of seven models Atkinson and Crothers (1964) found that the one-element model provided an unsatisfactory fit to paired-associate data. Atkinson and Crothers used, for the most part, longer lists and more response alternatives than did Bower. Calfee and Atkinson (1965) also failed to demonstrate a satisfactory fit of the Bower model when list length was 9, 15, or 21 items and responses were three consonant-vowel-consonant (CVC) trigrams. Because of these and other failures of the Bower model, theorists have extended and modified it in various ways. The remainder of this section will be devoted to reviewing models of paired-associate learning which are direct extensions or modifications of the one-element model just described.

Extensions of Bower's One-Element Model

An explicit prediction of the all-or-none position which the Bower model epitomizes is that the probability of a correct response is stationary or constant before learning occurs. Suppes and Ginsberg (1963) by vincentizing the pre-criterion response sequences demonstrated that the probability of a correct response was not stationary over trials; but response probability increased as trials increased. A two element model was devised by Suppes and Ginsberg to handle this problem. The transition matrix for their model is given below:

		Trial N+1			Pr (Correct)
		C_2	C_1	C_0	
	C_2	1	0	0	1
Trial N	C_1	b	(1-b)	0	g'
	C_0	0	a	(1-a)	g

In this model the stimulus is thought of as being composed of two elements. In state C_0 the response is not conditioned to either of the elements and the probability of a correct response is g. With a fixed probability, a, the correct response becomes conditioned to one of the elements and moves into state C_1. In state C_1 the probability of a correct guess, g', is greater than g but less than one. This formulation allows Suppes and Ginsberg to account for an increased probability of correct responding as trials increase. The model, however, does not give up its all-or-none properties since movement between states is still seen to occur in an all-or-none manner with fixed probabilities. This model is really nothing more than two one-element models placed end to end with different c and g values. The two one-element models are overlapped with the terminal probability for the first model being the starting probability for the next.

Atkinson and Crothers (1964) objected to the Suppes and Ginsberg model for two reasons. First, while g given by 1/N is a reasonable estimate of the guessing probability in state C_0 there is no convincing experimental interpretation given for the value of g', the guessing probability in state C_1. Secondly, Atkinson and Crothers demonstrated that when g' is estimated from data several predictions from the two element model are inaccurate. Atkinson and Crothers produced their own model which assumes four states L, S, F, and U. Learning consists of encoding a stimulus and then associating the encoded stimulus to the correct response. Before encoding the stimulus is in state U (uncoded); the subject responds by guessing randomly with probability g. Once the stimulus is encoded it can become associated to the correct response. When this happens the item moves into the L or learned state and has a correct response probability of one. F and S are intermediate states which represent events assumed to occur between encoding and learning. When an item is in state S it is in a short-term memory store and the probability of a correct response is also one. However, an item in short-term memory may be forgotten in which case it will move into state F wherein the probability of a correct response is again g. The transition matrix for this model is given below.

		Trial N+1				Pr (Correct)
		L	S	F	U	
Trial N	L	1	0	0	0	1
	S	a	(1-a)(1-f)	(1-a)f	0	1
	F	a	(1-a)(1-f)	(1-a)f	0	g
	U	ca	c(1-a)	c(1-a)f	1-c	g

The probability that an item is encoded on any trial is c and the probability an item is both encoded and associated on a single trial is ca. It is interesting to note that Atkinson and Crothers postulate the same probability of moving into the L state following a correct response in the S state as following an error in the F state. The model has the desirable feature of predicting increasing response probability over precriterion trials and is also qualitatively in accord with over-learning data since trials past the criterion serve to allow more transitions from S to L. Atkinson and Crothers propose two forms of the above model. The model as it was described is termed the LS-3 model because of the three parameters a, c, and f. The LS-2 model is a special case in which c equals one, which would mean that all stimulus items become encoded on the first trial.

Calfee and Atkinson (1965) propose a model which is quite similar to the Atkinson and Crother LS model. The trial dependent forgetting (TDF) model has three states (L, S, U) rather than the four found in the LS models. In the TDF model when an item has been in the S or short-term memory state and then is forgetten it returns to the U state rather than to an F state. The probability of returning to the U state from the S state is a function of the number of items that remain to be learned on any given trial. The transition matrix for the TDF model follows:

		Trial N+1			Pr (Correct)
		L	S	U	
	L	1	0	0	1
Trial N	S	a	$(1-a)(1-F_n)$	$(1-a)F_n$	1
	U	a	$(1-a)(1-F_n)$	$(1-a)F_n$	g

The parameter F_n is the probability of returning to the U state and is dependent upon the number of items that remain to be learned on trial N. As in the Atkinson and Crothers model the probability of learning following an error or a correct response is the same. Calfee and Atkinson also describe a revised version of the TDF model in which the probability, a, of moving into the learned state following a response is state S is different from the probability, b, of moving into the learned state following a response in state U. Minimum χ^2 estimates of these parameters showed the probability of getting into the learned state following a response in the S state is about four times greater than following a response in the U state. One serious drawback of this model is that it is difficult to determine just how many items are yet to be learned on each trial in order to estimate F. When a correct response is given it is impossible to tell whether it is correct because the item is in state L or because it is in state S. It will be evident as this review progresses this difficulty of parameter estimation increases as the models become more complex.

Greeno (1967) also uses the idea of a short-term memory store as an important state in a model. Grenno's contribution to the short-term memory store type models is his emphasis upon the effects of consolidation processes which occur while the item is in the short-term store. His model allows an item to drop out of the short-term memory state back into an unlearned state with the probability of this occurring being a function of the length of time between successive presentations of the same item. Greeno has a parameter, h, which is the probability of going into the short-term state and a is the parameter describing the probability of achieving long term storage or learning during the interval between successive presentations of the same item and is related to the length of the interval.

Bower and Theios (1964) developed a model for learning which separates an intermediate state into two states, one of which is an error state (E) and the other a success state (S). They adopt this formulation not because of theoretical assumptions such as short-term memory effects but rather as an aid in assessing the effectiveness of reinforcement following responses in the E or S state. The model is given by the following transition matrix.

		Trial N+1				
		1	S	E	0
	1	1	0	0	0
	S	s	$p(1-s)$	$q(1-s)$	0
Trial N	E	ε	$p(1-\varepsilon)$	$q(1-\varepsilon)$	0
	0	0	cp	cq	$1-c$

If s were equal to ε then the probability of going into the learned state after a success would be the same as after an error. In this case the S and E states could be collapsed and the model would become a three-state model. Bower and Theios feel that the values of s and ε obtained from the data are very instructive as to the learning process. This model is very interesting because it exemplifies how mathematical models can be used other than as direct tests of theories. Bower and Theios have set up the model's framework so that the parameters obtained from the data are indicative of the learning process and provide information as to differential effects of reinforcement.

A mathematical model built from a nonassociative point of view was proposed by Restle (1964). His model is based on trace theory of learning. The basis of this theory is the assumption that each time a stimulus-response pair is seen it is permanently recorded as an engram or memory trace. Learning is the process of adopting "strategies" which are used as aids in recalling the engram. After a strategy has been adopted it remains in use until the next presentation, if the subject is successful in recalling the stimulus then he maintains the strategy but if he is unsuccessful in recalling the engram he discards the strategy and adopts a new one. Once the subject has adopted a successful strategy he will thereafter respond correctly. This conceptualization is similar to the hypothesis explanation of concept learning. Explicit in this theory is the requirement that learning can occur only after an error unless the subject adopts a successful strategy on the first trial. The subject must make an error in order to force him to discard an ineffective strategy and resample from the population of strategies. If θ is the proportion of effective strategies in the pool then with probability θ the subject will sample an effective strategy and move into the learned state.

Restle is concerned with the problem of stimulus similarity. When stimuli are similar the subject may adopt the same strategies for both or he may become confused due to the similarity of stimuli and use the inappropriate strategy. Restle's model was built to describe paired-associate learning when the stimuli are similar and the problem of discrimination was introduced. The model requires the subject to select not only a successful learning strategy but also a successful discrimination strategy. Errors due to inappropriate discrimination strategies Restle calls confusion errors. The transition matrix for the model is given below.

		Trial N+1				Pr (Correct)
		S_3	S_2	S_1	S_0	
Trial N	S_3	1	0	0	0	1
	S_2	d	q(1-d)	p(1-d)	0	0
	S_1	0	q	p	0	1
	S_0	Θd	$q\Theta(1-d)$	$p\Theta(1-d)$	$1-\Theta$	g

Theta is the probability that the subject will select a strategy that leads to recall, d is the probability that he selects a successful discrimination strategy, p is the probability that he will make a correct response in the discrimination learning phase, and q is the probability of an error in the discrimination learning phase.

Two very similar models which were introduced at about the same time assume an all-or-none elimination of incorrect responses. Nahinsky (1964) and Millward (1964) both built their models upon the assumption that subjects can learn to eliminate incorrect response alternatives on their way to the learned state. These models can account for increasing response probability before learning because the subject can increase his guessing probability by learning to eliminate some of the wrong responses. This theoretical approach assumes simultaneous operation of two learning processes: (1) learning which is the correct response (2) learning which are not correct responses.

A model which views the learning of paired-associate lists as a decrease, with repetition, in the probability of forgetting was described by Bernbach (1965). The transition matrix is given below.

		Trial N+1				Pr (Correct)
		C'	C	G	E	
Trial N	C'	1	0	0	0	1
	C	θ	$(1-\theta)(1-\delta)$	$(1-\theta)\delta$	0	1
	G	0	$[1-\beta(1-g)](1-\delta)$	δ	$\beta(1-g)(1-\delta)$	g
	E	0	$(1-\beta)(1-\delta)$	δ	$\beta(1-\delta)$	0

Bernbach uses four states in his theory. An error state (E), a state in which a correct guess is made (G), a state which is similar to a short-term memory state which he calls C and a learned state (C') in which the probability of forgetting is zero. As learning progresses, the probability of forgetting, δ, decreases to an asymptote of zero. The relationship between δ and the number of trials is given by a stepwise function. Steps or changes in the probability of forgetting occur in an all-or-none manner only when the item is in state C and occur with probability θ. Bernbach also postulates the possibility of proactive inhibition which may operate after a subject makes an error and is then shown the correct response. The result of this proactive inhibition will be to produce an increase in the subjects tendency to repeat the incorrect response upon the next presentation of the stimulus. The probability of this occurring is β. It is interesting that Bernbach requires all items to be in the C state for at least one trial before final learning occurs. In other words, items cannot go directly from the error or from the guessing states into the learned state as they can in many of the other models reviewed.

The lack of stationarity observed in paired-associate data was a troublesome point for all-or-none theorists for many years. Recently, however, Polson and Greeno (1969) have demonstrated nonstationary data can be produced by an all-or-none process. They demonstrated that sequential probabilities can be manipulated by the effects of short term memory when randomized lists are used. They also describe biasing factors such as forgetting by the subject of the response which the experimenter gave as correct. These and other factors can produce data which is nonstationary even though the underlying association process is all-or-none. While Suppes and Ginsberg (1963) used an intermediate state to account for nonstationarity most authors, especially since the Polson and Greeno article, make use of intermediate states not as an explanation for nonstationarity but as theoretical steps through which items must pass on their way to the learned state.

Polson, Restle and Polson (1965) discuss the reasons for introducing intermediate states into all-or-none models. They propose the idea that the number of intermediate states needed to described paired-associate learning is proportional to the number of sources of difficulty in the list. When learning involves only the association of a stimulus with a response that is familiar to the subject then it can be described by a simple formulation such as Bower's one-element model. If other sources of difficulty are introduced into the list (e.g., stimulus similarity, response learning) then intermediate states must be added to models to account for these difficulties. To emphasize their contention they demonstrate that data obtained from lists made up of easily discriminated stimulus items and familiar response terms could be quite adequately described by Bower's model. However, when an additional source of difficulty was added by introducing similar pairs of stimulus items a three state model was needed to adequately predict the data. The three state model they used is described by the following matrix.

		S_2	S_1	S_0	Pr (Correct)
	S_2	1	0	0	1
Trial N	S_1	Qd	1-Qd	0	Q
	S_0	cd	c(1-d)	(1-c)	0

When a subject learns the correct response to one of the pair of similar stimuli he goes into the intermediate state state S_1 and makes a confusion error with probability Q. It will be noted that this model is merely a simpler formulation of Restle's model described earlier. Polson <u>et al</u>. conclude that unitary learning is an all-or-none process but most experiments require multiple processes for solution. These processes, however, are each all-or-none and with sufficient experimental control can be separated.

The following two models were built to describe free-recall learning rather than paired-associate learning but are included in the review because of their influence upon the formulation of the model which will be developed in the next chapter.

Waugh and Smith (1962) proposed a complex five-state model for free-recall. Three processes are proposed by Waugh and Smith to account for learning. The first process is labeling, which is finding a mnemonic device for associating the item. The next process is selecting or sampling items to be rehearsed and the third process is fixing which is analogous to learning to the point of recall on every trial.

Labeling occurs with probability λ, selecting with probability σ, and ϕ is the probability that an item will be fixed. The matrix follows:

		Trial N+1				
		5	4	3	2	1
	5	1	0	0	0	0
	4	σ	$1-\sigma$	$\sigma(1-\phi)$	0	0
Trial N	3	σ	$1-\sigma$	$\sigma(1-\phi)$	0	0
	2	σ	0	$\sigma(1-\phi)$	$(1-\sigma)$	0
	1	$\lambda\sigma$	0	$\lambda\sigma(1-\phi)$	$(1-\sigma)$	$1-\lambda$

State 1 is the not yet labeled state in which the probability of a correct response is zero. State 2 is the labeled but not selected state, here also the probability of a correct response is zero. State 3 is the labeled, selected but not yet stored state. When an item is in state 3 it is recalled with probability one. State 4 is for items which have been labeled and selected on previous trials but are not selected on trial N. State 5 is the absorbing learned state in which recall occurs on each trial. Waugh and Smith assume initial recall depends on a dual process, labeling and selecting. They do provide, however, for both these processes and the fixing process to occur on a single trial so that an item may move into the learned state on the first trial with probability $\lambda\sigma$.

The traditionally recognized difference between recognition and recall forms the basis for a model proposed by Kintsch and Morris (1965). The model assumes recognition can be described by Bower's simple model and that recall can also be described by a two-state process once the items to be recalled have moved into the recognition state. In this conceptualization the learned state for recognition, C_1, is the initial state of the recall learning model. Their matrix is presented below.

		Trial N+1			Pr (Correct)
		C_2	C_1	C_0	
	C_2	1	0	0	1
Trial N	C_1	θ	$(1-\theta)$	0	1-r
	C_0	0	c	(1-c)	0

C_0 is the non-recognition, nonrecall state and C_1 is the recognition-but-not-recall state. It should be noted that Kintsch and Morris propose different learning rates for the two initial stages θ and c. The model is very similar to the Bower and Theios model of paired-associate learning, except Kintsch and Morris do not propose differential learning probabilities following successes and errors. Kintch and Morris also do not provide for a direct transition from C_0 to C_2 which means an item cannot be learned (state C_0) on the first trial. Kintch and Morris had subjects learn lists of nonsense syllables by the methods of recall and recognition. They found that a two state model would describe data from the recognition learning but that a three-state model was needed to describe the free recall data. They also found, as they predicted, that once a list had been learned by the recognition method a two state model would describe the data from that point in learning until the list was learned to a recall criterion.

In summary, Bower's model has been extended and modified in many ways while still retaining its all-or-none and Markov properties.

Also, greatly different theoretical assumptions have used it as a starting point for their mathematical expression. It has proven useful not only as an incentive to research but also as a basic tool which can be used by differing theoretical views (e.g., cognitive, associative) of the learning process. It has also been used to test more subtle differences in subprocesses from similar theoretical positions such as the question of whether or not an item must go trrough intermediate states on its way to the learned state. However, a note of caution should be sounded. Too many processes and subprocesses may have been elaborated. It is highly possible that paired-associative learning may be simpler than many of these models would postulate.

CHAPTER III

A THREE STATE MARKOV MODEL

Data from the author's Master's Thesis was used to test Bower's one-element model. It was found that the model provided a satisfactory fit to data from easily learned lists but failed to describe adequately data from slowly learned lists (Linscheid, 1969). Other examples of the failure of Bower's model when applied to paired-associate data from difficult lists have been reported in the literature (e.g., Atkinson and Crothers, 1964; Calfee and Atkinson, 1965). In an effort to describe data regardless of the learning rate a three state Markov model was developed. The model is based on theoretical formulations from both the Waugh and Smith (1961) paper and the Kintsch and Morris (1965) paper.

The model assumes three states. An unlearned state, A, a recognition state, B, and the learned or recall state C. Transition probabilities for the model are given below:

		Trial N+1			Pr (Correct)
		C	B	A	1
	C	1	0	0	1
Trial N	B	b	1-b	0	p
	A	ab	a(1-b)	(1-a)	0

The learning of paired-associates is viewed as a recognition-recall process. Subjects first learn to recognize a stimulus-response pair and then to recall it. In state A a pair is neither recognized nor recalled. Once a pair is recognized it moves into state B. While in state B a recognized pair may be selected for rehearsal with probability p. If the recognized pair is rehearsed the correct response is given with probability one; if the recognized pair is not rehearsed the probability of a correct response is zero. The transition from the B state to the C state (recognition to recall) occurs with probability b and is independent of whether or not an item has been selected for rehearsal on any previous trial. Like the Kintsch and Morris model this model views recognition and recall as two separate processes which are each described by two-state models. Unlike the Kintsch and Morris formulation the present model allows for a direct transition from the unlearned to the learned state. The model is designed for use with the train-test method. The train-test method allows subjects to view the correct pairs before the first test so that the probability of a correct response on the first trial is not the guessing probability. The model was designed for longer lists with each stimulus paired with a different response. In such lists the guessing probability is negligible and it has been ignored in the formulation of the model. However, with minor modifications the model could be adapted for use with the anticipation method or to accommodate larger guessing probabilities.

The first step in deriving predictions from a Markov model is to determine the state probabilities. The notation $W_{S,n}$ will be used to denote the probability that an item is in state S on trial n. The probability of being in the A state on the n^{th} trial is the probability of having not left the A state for n trials or

$$W_{A,n} = (1-a)^n \tag{1}$$

The probability of being in the B state on Trial n is the sum of the separate pathways an item can take in getting to the B state on Trial n [e.g., $W_{B,3} = Pr(A_1A_2A_3) + Pr(A_1B_2B_3) + Pr(B_1B_2B_3)$]. In general form it is given by

$$W_{B,n} = a(1-b)^n + a(1-b)(1-a) \sum_{k=0}^{m-2} (1-a)^k(1-b)^{n-2-k} \tag{2}$$

The probability of being in the C state is obtained by subtraction

$$W_{C,n} = 1 - (W_{A,n} + W_{B,n}) \tag{3}$$

The next derivation of interest is the learning curve or probability of an error which is denoted by q_n. An error can occur in two of the three states. If an item is in state A an error occurs with probability 1; if the item is in state B an error occurs with probability q, which is the probability of an item in the B state not being rehearsed. The formulation of the error probability formula is quite simple.

$$q_n = W_{A,n} + q\, W_{B,n} \tag{4}$$

The variable J will be defined as the number or errors before the first correct response. To derive the probability distribution of J two paths must be considered. A subject may make k errors in state A and then move either into state B and make a correct response with probability p or he may move directly into state C and be correct with probability 1. In either case, the correct response ends the error run. A subject also may make i errors in state A (where i J) and then move into state B and make J-i errors before making a correct response. Alpha (α) will be defined as the probability that an error follows an error in the intermediate state or $\alpha = q(1-b)$ and w will be defined as the probability of making a correct response upon leaving the A state and is given by $w = b + p(1-b)$. The distribution of J can now be written as follows:

$$Pr(J=k) = (1-a)^k aw + \sum_{i=o}^{k} (1-w)(1-a)^i a(\alpha)^{k-i}(1-\alpha) \quad (5)$$

or in general form

$$Pr(J=k) = (1-a)^k aw + \frac{a(1-w)(1-\alpha)}{\alpha-(1-a)} [\alpha^k-(1-a)^k] \quad (6)$$

It can be seen from inspection of equation 5 that Pr(J=k) is equal to the probability that all J errors are made in state A plus the probability that i errors are made in state A followed by a string of J-i errors in state B. The formula for the mean number of errors before the first success, E(J), is:

$$E(J) = \frac{1}{a} + \frac{q}{b + qb} \quad (7)$$

This is simply the mean number of errors in the A state plus the mean number of errors before a success in the intermediate state.

The next formulation to be considered is the probability distribution of the total errors. To do this we break up the total errors into the total errors made in state A and the total errors made in state B or $T = t_A + t_B$:

therefore $\Pr(t_A=k) = (1-a)^k a$ (8)

and $\Pr(t_B=j) = \begin{cases} u & (\text{for } j=0) \\ (1-u)aE(1-E)^j & (\text{for } j\leq 1) \end{cases}$ (9)

where E is equal to the probability of no more errors starting in the intermediate state and is given by $E = b/\ 1-p(1-b)$. The quantity u is the probability of no more errors upon leaving the A state. The formula for us is; $u = b + pE(1-b)$. Conceptually the overall distribution of T may be given by

$$\Pr(T=k) = [\Pr(t_A+K)][\Pr(t_B+))] + \sum_{i=1}^{k} [\Pr(t_A=i)][\Pr(t_B+k-i)] \quad (10)$$

Summing and simplifying we obtain

$$\Pr(T=k) = (1-a)^k au + \frac{aE(1-u)}{a-E} [(1-E)^k-(1-a)^k] \quad (11)$$

The mean total errors is the sum of the average number of errors in state A and state B and is given by

$$E(T) = 1/a + q/b \quad (12)$$

The final statistic to be developed is the probability distribution of the last error trial. In order for an error to occur an item must be either in state A or state B. If an item is in state A on trial n then with probability aw it will leave A and no more errors will be made. If an item is in state B on trial n an error will be made with probability q and no further errors will be made with probability E. Therefore the probability distribution of the last error trial is given by

$$Pr(L=k) = W_{A,n}\ a\ w + W_{B,n}\ q\ E \tag{13}$$

and the mean last error trial is given by

$$E(L) = \frac{1}{a} + \frac{q}{b(q+pb)} \tag{14}$$

Parameter Estimates

Bower and Theios (1964) in the formulation of their model describe the estimation of parameters. Because of the similarity of the present model to the Bower-Theios model several of the same methods will be used. Their c is comparable to a in the present model since it is the probability of leaving the initial state. Since the present model does not postulate differential learning rates following successes and errors the best estimate of the b value is an average of Bower-Theios error and success learning probabilities or $b = \frac{s + \varepsilon}{2}$. The probability of an error in the intermediate state, q, will be taken directly from the data by counting the number of errors between the first success and the last error and dividing by the number of intermediate trials.

To produce the differential learning rates needed to test the model stimulus term meaningfulness will be manipulated. Numerous studies which have held response meaningfulness constant and varied stimulus meaningfulness (Goss and Nodine, 1965, pp. 90-92) have demonstrated differential and reliable effects upon learning rates. Response term meaningfulness will be held constant by using digits as responses and stimulus meaningfulness will be manipulated by using CVC's from Archer's (1961) list.

CHAPTER IV

METHOD

The subjects were 90 introductory psychology students (65 males, 25 females) from the University of North Dakota who participated in the experiment as a course requirement. They were run in groups of 8 to 15 until 30 had been run for each of the three lists.

The task for all subjects was to learn a paired-associate list of 16 pairs. The train-test method was used. The items were presented on slides using a Kodak 800 slide projector. The projector was programmed to change slides every four seconds. The changing time was slightly under one second so the actual exposure time was just over three seconds. During the training phase the subjects studied the stimulus-response slides. At the end of the training phase a blank slide appeared for four seconds followed by the test phase in which subjects were shown the stimulus items and were required to supply the response. Exposure time was the same during the test phase. Following the test phase two successive blank slides signaled the beginning of the next trial. There were four random orders of each list, both for the training and test phases. Subjects wrote their responses in booklets and after each trial (i.e., during the exposure of the two blank slides) turned the page so that a blank answer sheet was exposed. This prevented them from studying previously given responses. Subjects were required to copy down the stimulus and supply the response during the test phase. They were instructed to copy the stimulus even if they did not know the response. The following instructions were read to the subjects before each experimental session: This is a learning experiment in which you will learn to associate or "hook up" a three letter sequence like RTX with a number. You will be presented a series of slides. On each slide there will appear a three letter sequence and number. The numbers will be the integers from 1 to 16. Following the presentation of the three-letter sequence and number slides you will see a blank slide. After the blank slide you will be shown a series of slides which have only the three letter sequences on them. As these slides are presented you are to copy down the three letter sequence and supply the number that was paired with it previously. You may guess if you wish but you are not required to do so. You should write quickly because you will not have unlimited time to answer. Write the letter sequences and number in your answer booklet, one sequence and number per line. At the end of each trial, which will be signaled by two successive blank slides, turn the page of your answer booklet so that you can no longer see the answers which you have just written. The same sequence of events will then be repeated. Do not go back to previous pages of your answer booklet and do not attempt to study from your previous answers. Remember. You must copy the three letter sequence, and if you can, supply the number which goes with it. Also, remember to turn to the next blank page of your answer booklet each time you see the two successive blank slides. Are there any questions?

The stimulus items were CVC trigrams selected from Archer's (1961) list. To control for stimulus similarity the following restrictions were placed on the items. No consonant could appear in each list more than twice in the first position of the trigram nor could it appear more than twice in the third position. No vowel (y included) could appear more than four times in each list and no CV or VC combination could appear more than once in each list. Three lists were chosen in this manner. The low meaningfulness (LM) list was made up of trigrams whose mean association value (AV) was 8.0 (range 1 to 15). The mean AV for the medium meaningfulness (MM) list was 45.25 (range 43 to 47) and the high meaningfulness list (HM) had a mean of 99.9 (range 99 to 100). The response items for all three lists were the integers 1 through 16. Each was paired with a CVC and was used only once in each list. The lists are shown in Table 1.

TABLE 1

LOW, MEDIUM, AND HIGH MEANINGFULNESS LISTS

	Meaningfulness	
Low	Medium	High
XIH-16	FUJ-10	BAN-3
QUJ-7	GYT-12	BIT-16
GYQ-1	SIQ-5	CAT-10
TEJ-14	DEG-15	GUM-5
WUQ-8	GOZ-6	DEN-4
JYH-6	HAQ-16	FAR-15
MYV-9	JOH-8	FOX-6
RYW-13	PEM-1	GAS-11
ZOS-5	QIC-11	HIM-12
ZUF-3	WYM-3	HOP-7
GEX-2	CYK-2	WIG-1
VOF-12	DYS-14	JUG-8
VUB-10	LIX-4	WED-14
FEP-15	CIB-13	MUD-2
BIW-11	MOG-9	TUB-13
XAZ-4	PAJ-7	SIR-9

The LM group was run for 24 trials, the MM group for 18 trials, and the HM group for 16 trials. The number of trials each list was run was adequate to insure virtually all items were learned to a criterion of five successive correct responses.

CHAPTER V

RESULTS

For the analysis of stimulus meaningfulness effects four subjects (three from the LM group and one from the MM group) were excluded because they failed to learn to the criterion of two perfect recitations of the list. Using trials to criterion as the dependent variable a one way analysis of variance yielded a significant effect for meaningfulness ($F = 37.66$, $df = 2$ and 83, $P < .001$). Table 2 is the analysis of variance summary table and Table 3 shows the means and standard deviations.

TABLE 2

ANALYSIS OF VARIANCE SOURCE TABLE

Source	df	SS	MS	F
Meaningfulness	2	817.045	403.523	37.662 P<.001
Error	83	900.268	10.847	
Total	85	1717.313	20.204	

TABLE 3

MEANS AND STANDARD DEVIATIONS

Group	M	SD
LM	12.481	4.069
MM	8.380	3.429
HM	4.900	1.938

The means are in the expected direction and show the usual marked relationship between stimulus meaningfulness and trials to criterion. The learning curves for the three groups are shown in Figure 1. The HM curve is remarkably smooth, an almost picture-perfect curve. The MM and LM curves are also quite smooth for paired-associate data. It is evident that the desired differential learning rates were obtained and the results indicate once again the potent effects of meaningfulness in verbal learning.

The criterion for considering an individual subject-item correct was five successive correct responses. For the HM group all 480 subject-items met this criterion. The MM group had 471 items and the LM had 450 items. The maximum loss of items (6.2%) occurred in the LM group. The data obtained was compared to predictions from Bower's one-element model, the Bower-Theios model, and to the model developed in Chapter III hereafter called the paired-associate recognition-recall (PARR) model. While many predictions about the data could be made, the three models will be evaluated by comparing predictions of probability of correct response (p_n), errors before the first success (J), total errors per item (T), and trial of last error (L) with their respective observed values. Both the means and distributions of these statistics will be considered.

Fits of Models to LM Data

Data from the LM list will be considered first. Since the Bower-Theios model was designed for use with the anticipation method a modification was required. The learning curve and probability distributions were all shifted one trial so predictions from Bower-Theios were in line with the actual trial number of the data. For example, the prediction

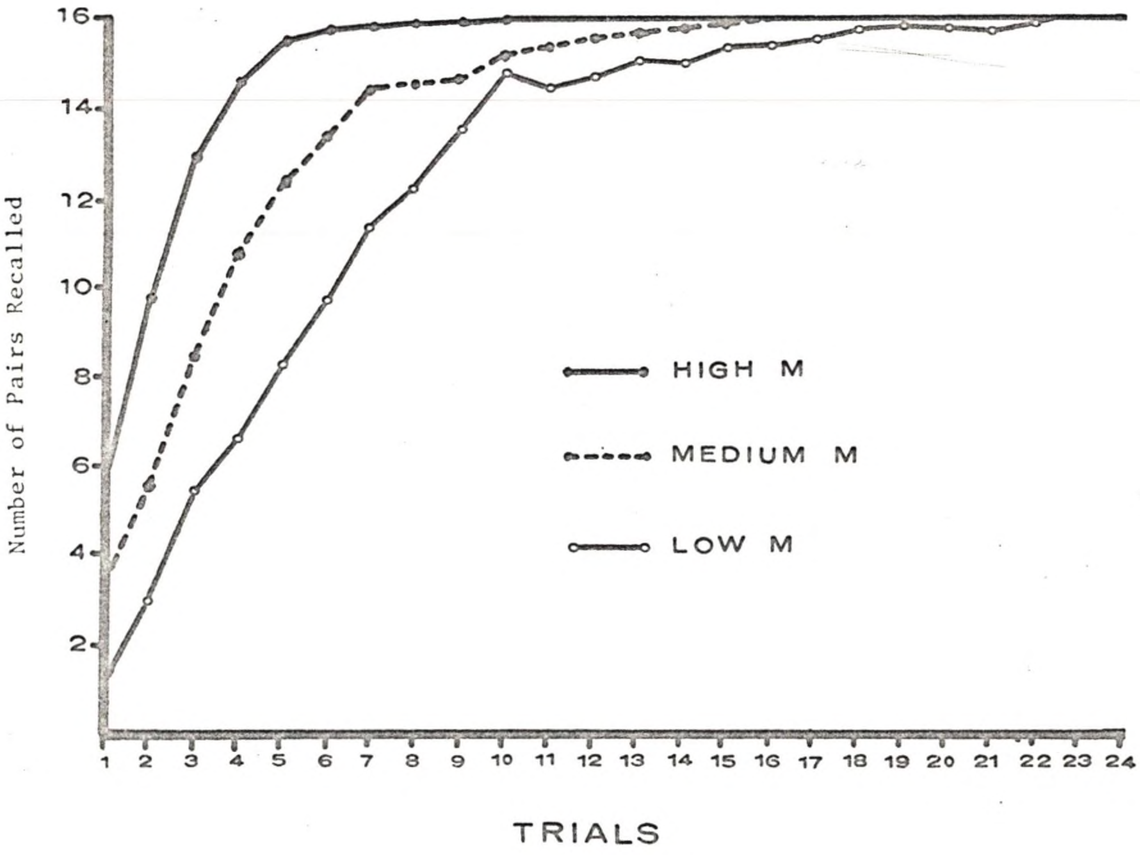

Fig. 1. Learning curves for the LM, MM, and HM lists.

for the probability of being correct on the second trial was actually the prediction of the first trial correct response probability. The same modification was used for the probability distributions. The one-element model was also modified since it too was designed for the anticipation method. The guessing probability aspect was removed from the one-element model to make it more in line with the other models. The modified formula used to calculate the predicted learning curve for the one-element model is given below:

$$\Pr(\text{correct}) = 1-(1-c)^n$$

The estimate of the c parameter for the Bower one-element model was .1974. This was obtained by dividing one by the mean total errors for the LM list or 5.0666. The estimate of the Bower-Theios c value was .3238 and the estimates of s and ε were .3070 and .2470 respectively. The PARR estimate for a was .3238 or the same value as the Bower-Theios c value. The b parameter used in the PARR model was .2745 and was found by averaging the s and ε estimates from Bower-Theios. The p value was determined to be .4570.

The observed and predicted learning curve values are shown in Table 4. Figure 2 shows the predicted learning curves and the observed learning curve. Only fifteen trials are shown because after that point the predictions from the several models are virtually the same. All three models predict a more rapid learning rate than is observed. Because it has only one constant it is easiest to appraise the failure of the one-element model to predict the learning curve by saying that c is roughly twice what it should be. A c value of around .10 would fit fairly well, although it would do quite badly in predicting other

TABLE 4

PREDICTED AND OBSERVED LEARNING CURVES FOR THE LM LIST

Trial	Pr(correct) Observed	One-Element	Bower & Theios	PARR
1	.0688	.1974	.1478	.1962
2	.1813	.3558	.2963	.3639
3	.3188	.4830	.4316	.5028
4	.4000	.5851	.5485	.6150
5	.4939	.6670	.6460	.7043
6	.5938	.7327	.7253	.7744
7	.6875	.7855	.7886	.8289
8	.7188	.8278	.8384	.8708
9	.8188	.8618	.8772	.9028
10	.8500	.8891	.9072	.9272
11	.8875	.9110	.9301	.9456
12	.8938	.9286	.9476	.9595
13	.9188	.9427	.9608	.9699
14	.9295	.9540	.9708	.9777
15	.9438	.9631	.9783	.9835
16	.9500	.9704	.9839	.9878
17	.9750	.9763	.9881	.9910
18	.9906	.9809	.9912	.9934
19	.9956	.9847	.9935	.9951
20	.9906	.9877	.9952	.9964
21	.9906	.9902	.9965	.9974
22	.9956	.9921	.9974	.9981
23	1.0000	.9937	.9981	.9986
24	1.0000	.9949	.9986	.9990

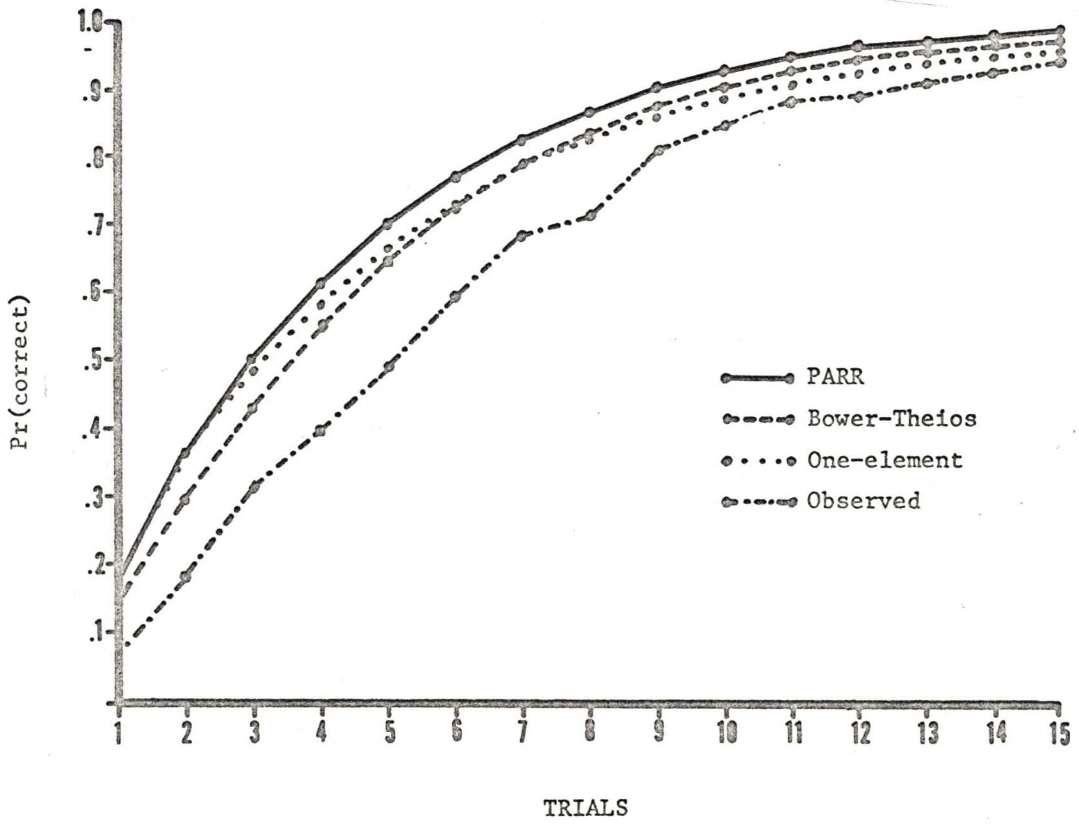

Fig. 2. Observed and predicted learning curves for the LM list.

statistics (e.g., it would predict a mean of 10 errors compared to an observed mean of 5 errors). Similar considerations apply to the other models. Thus, all the models predict that the first stage of learning (the only stage in the one-element model) progresses more rapidly than what is observed. The models do not have to predict such fast learning, in the sense that the first stage constants can be very small, but given the particular estimates they are seriously in error in predicting the LM learning curve.

Despite its failure to predict the LM learning curve, the PARR model does very well in its predictions of the mean number of errors before first correct response, mean total errors, and mean last error trial. These predictions are shown in Table 5. The Bower-Theios model

TABLE 5

OBSERVED AND EXPECTED MEANS OF J, T AND L FOR THE LM LIST

	Observed	One-Element	Bower & Theios	PARR
$\bar{J}$	4.0066	5.0666	4.0066*	3.8942
$\bar{T}$	5.0666	5.0666*	5.0666*	5.0665
$\bar{L}$	6.0911	5.0666	5.9836	6.0477

*Values used in parameter estimation

predicts the mean last error trial well, the other two means are used in parameter estimation. The one-element model fails in predicting the means. The means imply the typical LM pair is learned with four consecutive errors and five total errors with the last error made on the sixth trial. Judging from the closeness of the J and L means the transition of an item from unlearned to learned is abrupt, but not as abrupt as the one-element model would have it.

The probability distributions of number of errors (trials) before the first success (J) are shown in Table 6 and graphically represented

TABLE 6

OBSERVED AND PREDICTED PROBABILITY DISTRIBUTIONS OF J FOR THE LM LIST

J	Pr(J) Observed	One-Element	Bower & Theios	PARR
0	.0688	.1974	.1478	.1962
1	.1400	.1584	.2040	.2100
2	.1666	.1272	.1804	.1725
3	.1288	.1021	.1394	.1286
4	.1155	.0819	.1014	.0917
5	.1244	.0657	.0714	.0639
6	.0822	.0528	.0495	.0439
7	.0511	.0423	.0340	.0300
8	.0466	.0340	.0232	.0204
9	.0311	.0272	.0157	.0138
10	.0177	.0219	.0107	.0094
11	.0066	.0176	.0072	.0063
12	.0044	.0141	.0049	.0043
13	.0022	.0113	.0033	.0029
14	.0000	.0091	.0022	.0020
15	.0044	.0073	.0015	.0013
16	.0022	.0058	.0010	.0009
17	.0000	.0047	.0007	.0006
18	.0044	.0038	.0005	.0004
19	.0022	.0030	.0003	.0003

in Figure 3. It should be noted that when the guessing probability is assumed to be zero the probability distributions of J, T, and L will be the same for the one-element model since there can be no correct

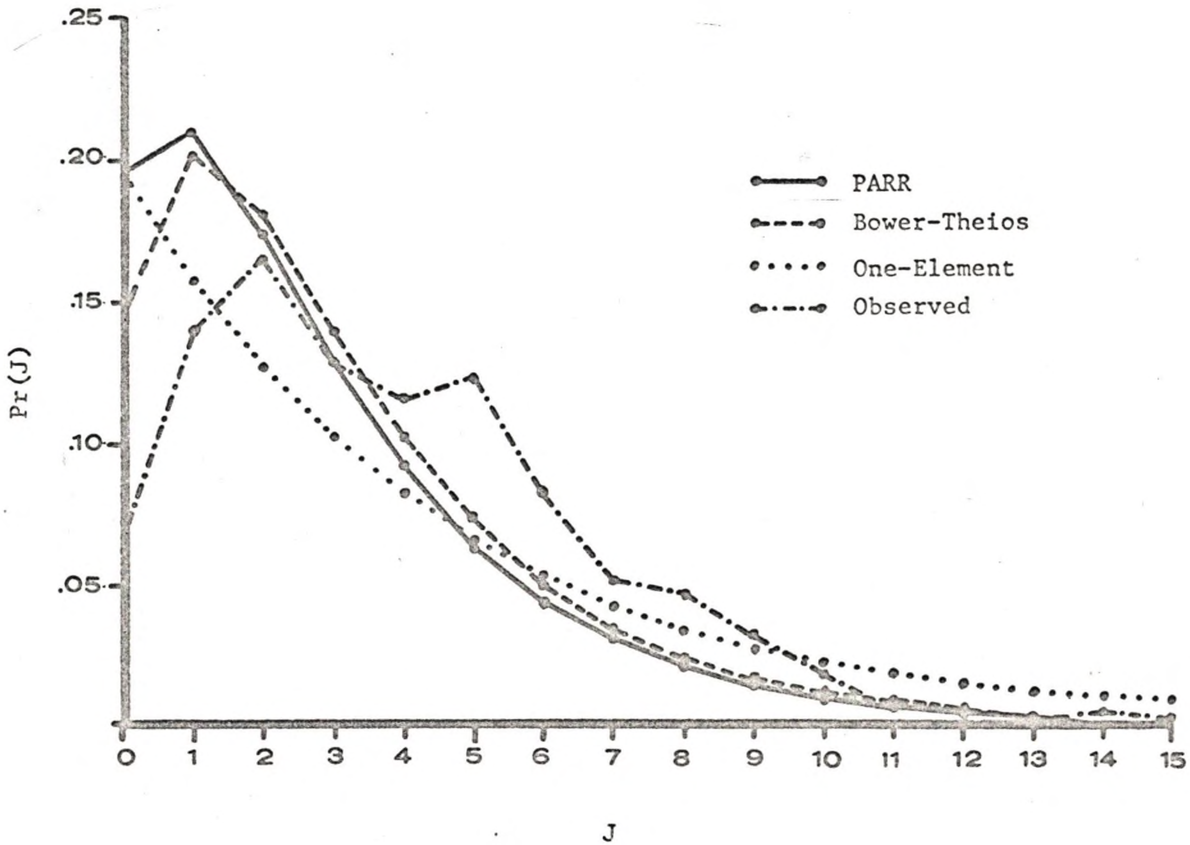

Fig. 3. Observed and predicted probability distributions of J for the LM list.

responses before learning. The mode of the obtained J distribution is 2. The one-element model predicted a mode of 0 and the PARR and Bower-Theios model both predicted a mode of 1. All three models predicted a higher proportion of small J values than were observed. The obtained J distribution differs from the predicted distributions in two other ways. First, it has a larger tail. This means that some pairs are not recalled for the first time until practice is rather well along. Although it may be that some pairs are of greater intrinsic difficulty than others, it seems certain that any heterogeneity whatever, regardless of the source, will stretch the tail of the J distribution. Second, the models grossly overestimate the proportion of items which will be recalled on the first test trial. Paired-associate learning of LM trigrams started out very slowly, so much so that from the obtained J distribution we may deduce the initial positive acceleration in the learning curve (see Table 4). This positive acceleration is not present in the MM and HM curves to be considered later.

The T probability distributions are shown in Figure 4 and Table 7. The obtained distribution does not have a well defined mode. The geometric distribution predicted by the one-element model is not realized. The Bower-Theios model appears to have an edge on the PARR model, particularly for small values of T. The L distributions are shown in Figure 5 and Table 8. The obtained L distribution is more irregular. Although the Bower-Theios model appears to predict the shape, the point-by point discrepancies are fairly sizeable. None of the models fit well, although it may be recalled that both the Bower-Theios and PARR models predicted mean L very well.

TABLE 7

OBSERVED AND PREDICTED PROBABILITY DISTRIBUTIONS OF T FOR THE LM LIST

		Pr(T)		
T	Observed	One-Element	Bower & Theios	PARR
0	.0244	.1974	.0665	.1330
1	.1088	.1584	.1483	.1683
2	.1311	.1272	.1621	.1600
3	.1244	.1021	.1466	.1354
4	.1044	.0819	.1213	.1076
5	.1288	.0657	.0953	.0822
6	.0933	.0528	.0723	.0612
7	.0733	.0423	.0537	.0446
8	.0711	.0340	.0391	.0321
9	.0355	.0272	.0281	.0229
10	.0288	.0219	.0201	.0161
11	.0200	.0176	.0142	.0113
12	.0155	.0141	.0099	.0079
13	.0155	.0113	.0069	.0055
14	.0044	.0091	.0048	.0038
15	.0200	.0073	.0033	.0026
16	.0000	.0058	.0023	.0018
17	.0000	.0047	.0016	.0012
18	.0000	.0038	.0011	.0008
19	.0000	.0030	.0007	.0006

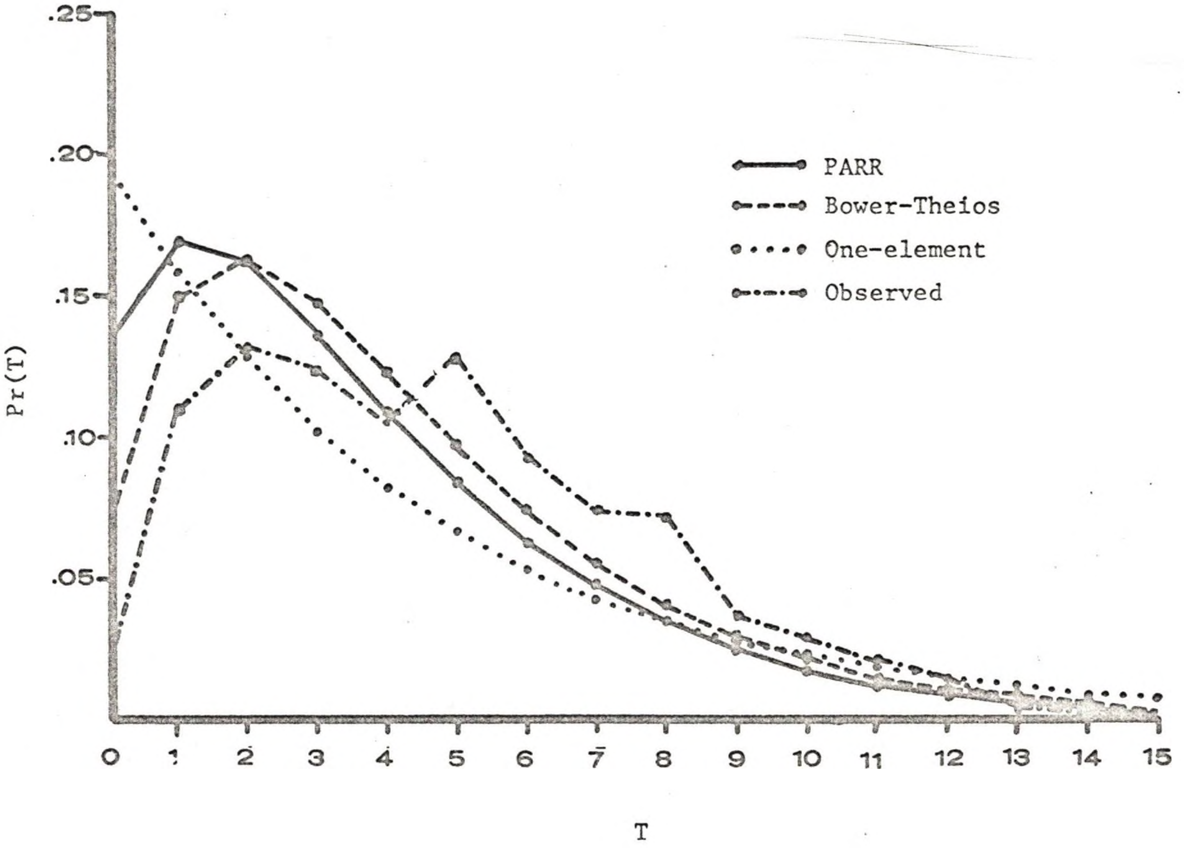

Fig. 4. Observed and predicted probability distributions of T for the LM list.

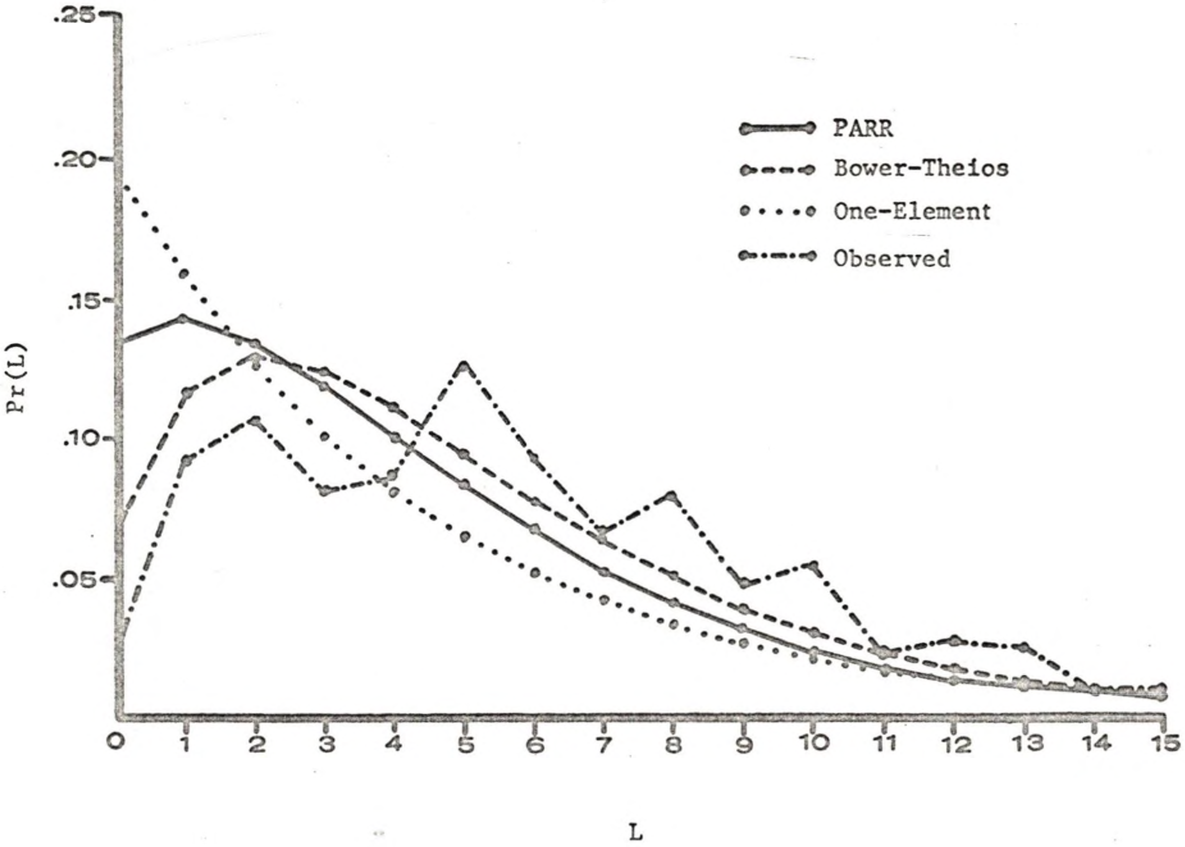

Fig. 5. Observed and predicted probability distributions of L for the LM list.

TABLE 8

OBSERVED AND PREDICTED PROBABILITY DISTRIBUTIONS OF L FOR THE LM LIST

L	Pr(L) Observed	One-Element	Bower & Theios	PARR
0	.0244	.1974	.0665	.1329
1	.0911	.1584	.1156	.1423
2	.1066	.1272	.1294	.1342
3	.0800	.1021	.1247	.1183
4	.0866	.0819	.1113	.1000
5	.1266	.0657	.0948	.0821
6	.0933	.0528	.0783	.0661
7	.0666	.0423	.0633	.0523
8	.0800	.0340	.0502	.0409
9	.0488	.0272	.0394	.0317
10	.0555	.0219	.0306	.0243
11	.0244	.0176	.0235	.0186
12	.0288	.0141	.0180	.0141
13	.0266	.0113	.0137	.0106
14	.0111	.0091	.0103	.0080
15	.0111	.0073	.0078	.0060
16	.0088	.0058	.0058	.0045
17	.0088	.0047	.0044	.0033
18	.0111	.0038	.0033	.0025
19	.0066	.0030	.0024	.0018

Fit of the Model to MM Data

The fit of the models to the MM list data will be considered next. Parameter estimates yielded .3209 for the one-element c value, .6889 for the Bower-Theios c and .4819 and .2938 for s and ε respectively. The PARR estimates were .6889 for a and .3879 for b. The value of p was .3889. The observed and predicted learning curves are given in Figure 6 and Table 9.

All three models, as they did for the LM list, overpredict the learning rate. They predict a higher initial correct response probability and continue to predict more rapid learning across all trials. There is very little difference in the learning rates predicted by the Bower-Theios model and the one-element model. The PARR model provides the worst predictions about the learning rate. The shape of the PARR learning curve is very similar to the other models but is elevated considerably. Ironically, the constants calculated for the LM list are about right for all three models to predict the MM learning curve. Again this seems to be a fault of the estimates.

Table 10 gives the mean J, T, and L values. The one-element model, with its across-the-board prediction of the equality of these means is less in error than with the LM data because the spread of the MM means is reduced. The PARR predictions of the J and T means are good but the model is outpointed by the Bower-Theios model in predicting the L mean. The J distributions are shown in Table 11 and Figure 7. Approximately 60 percent of the items are recalled correctly on the first three test trials; thereafter, the proportion of newly recalled items declines. All three models overpredict the

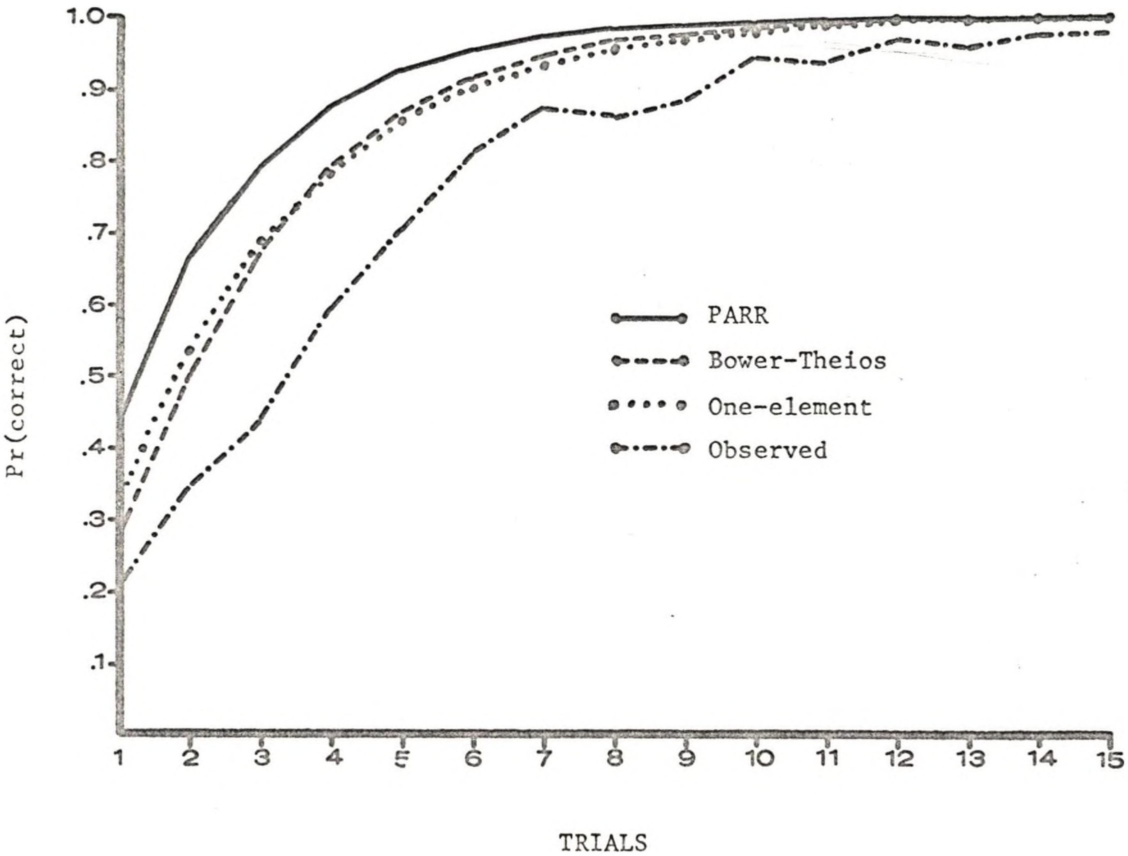

Fig. 6. Observed and predicted learning curves for the MM list.

TABLE 9

OBSERVED AND PREDICTED LEARNING CURVES FOR THE MM LIST

Trial	Observed	One-Element	Bower & Theios	PARR
1	.2044	.3209	.2679	.4313
2	.3456	.5388	.5058	.6653
3	.4400	.6868	.6776	.7994
4	.5963	.7873	.7929	.8785
5	.7056	.8556	.8680	.9260
6	.8144	.9019	.9161	.9548
7	.8769	.9334	.9468	.9724
8	.8606	.9548	.9663	.9831
9	.8831	.9693	.9787	.9897
10	.9425	.9792	.9864	.9937
11	.9356	.9859	.9914	.9961
12	.9668	.9904	.9946	.9976
13	.9581	.9935	.9966	.9986
14	.9738	.9956	.9978	.9991
15	.9794	.9970	.9986	.9995
16	.9856	.9980	.9991	.9997
17	1.0000	.9986	.9995	.9998
18	1.0000	.9990	.9997	.9999

TABLE 10

OBSERVED AND EXPECTED MEANS OF J, T AND L FOR THE MM LIST

	Observed	One-Element	Bower & Theios	PARR
$\bar{J}$	2.5265	3.1167	2.5265*	2.4279
$\bar{T}$	3.1167	3.1167*	3.1167*	3.0270
$\bar{L}$	3.5647	3.1167	3.5368	3.9685

*Used in parameter estimation

TABLE 11

OBSERVED AND PREDICTED PROBABILITY DISTRIBUTIONS OF J FOR THE MM LIST

J	Observed	One-Element	Bower & Theios	PARR
0	.2038	.3209	.2679	.4313
1	.2080	.2179	.3227	.2955
2	.1995	.1480	.2036	.1522
3	.1380	.1004	.1079	.0699
4	.0828	.0682	.0528	.0302
5	.0509	.0463	.0247	.0125
6	.0509	.0315	.0113	.0051
7	.0191	.0214	.0051	.0020
8	.0084	.0145	.0022	.0008
9	.0084	.0099	.0010	.0003
10	.0084	.0067	.0004	.0001
11	.0106	.0045	.0002	.0000
12	.0042	.0031	.0001	.0000
13	.0063	.0021	.0000	.0000

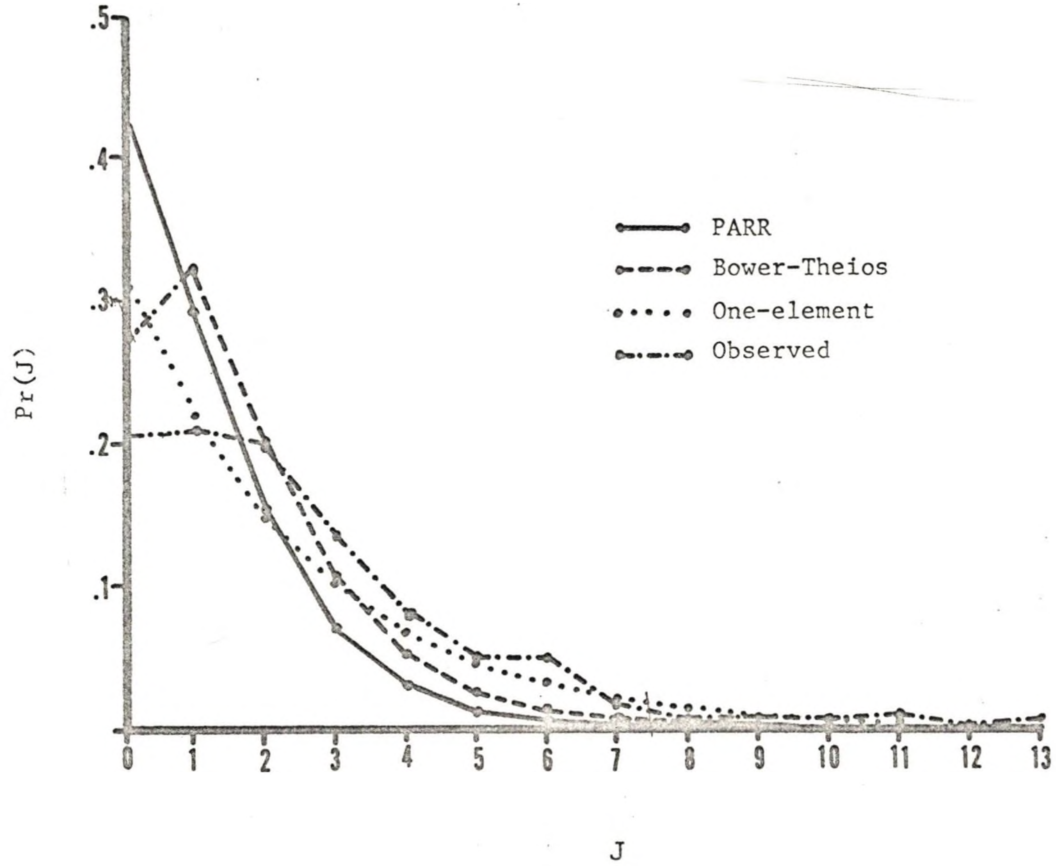

Fig. 7. Observed and predicted probability distributions of J for the MM list.

probabilities of first and second trial recalls, then underestimate the larger J values. The T distributions are found in Figure 8 and Table 12. The Bower-Theios model predicts this distribution rather well except for erroneously forecasting a peaked mode of one error. The PARR model does very poorly. Here the fault seems to be the formulation that an item can go straight through from the A state to the C state with probability ab, in which case no errors are made. This probability is (.6889)(.3879) = .2672 alone, and this is not the only error-free pathway available to the subject item. The L distributions shown in Table 13 and Figure 9 are similar to the T distributions just described and for much the same reasons.

Fit of the Models to HM Data

Inspection of data from the HM list indicated the absence of an intermediate state. Only 24 of the 480 subject-items showed response sequences which indicated intermediate states. While the guessing probability has not been considered to this point it is possible that chance guessing produced what looked like intermediate trials for the HM list. The correct response probability before learning was determined by the formula, $p = (\overline{L} - \overline{T})/(\overline{L})$, which is the average number of correct responses before learning divided by the average number of trials to learn. The formula yielded a p value of .0643. With the 16 item lists used, the probability of a correct guess would be .0625 which is very close to the obtained p value. This bit of evidence also argues against the existence of intermediate states for the HM subject-items. Intermediate states postulate greater than chance correct response probabilities before learning and are not needed when correct response

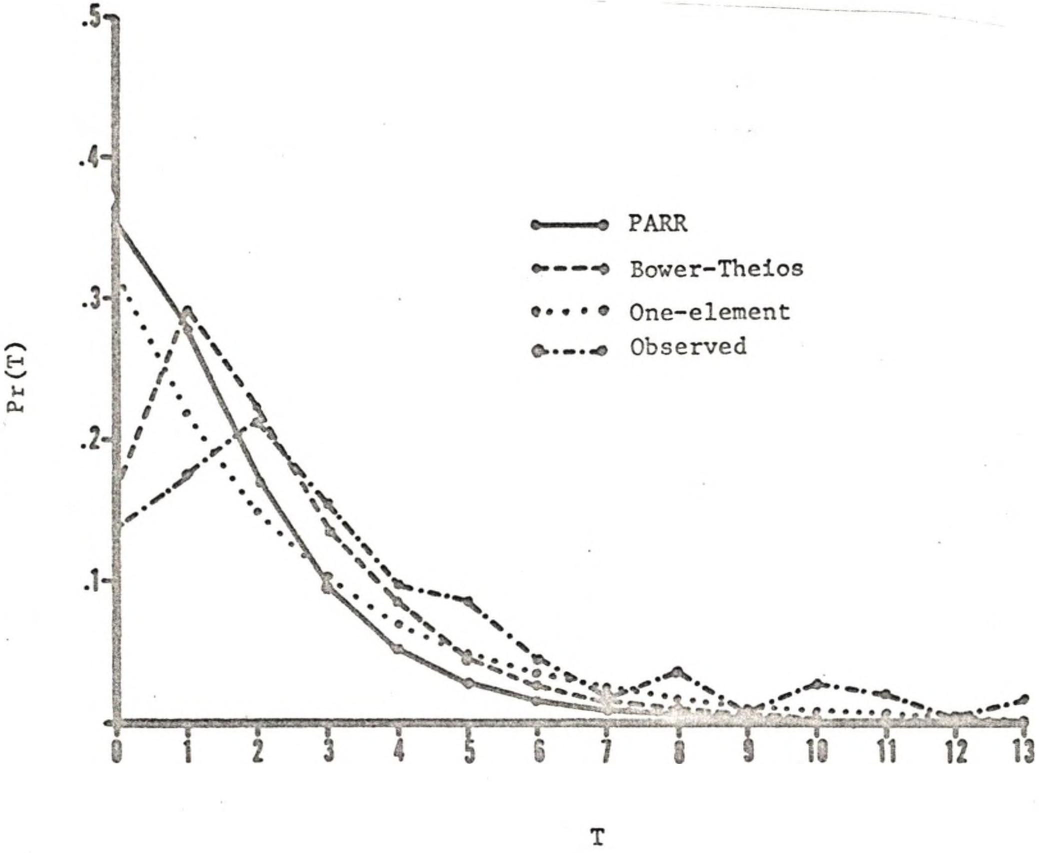

Fig. 8. Observed and predicted probability distributions of T for the MM list.

TABLE 12

OBSERVED AND PREDICTED PROBABILITY DISTRIBUTIONS OF T FOR THE MM LIST

T	Observed	One-Element	Bower & Theios	PARR
0	.1380	.3209	.1617	.3507
1	.1719	.2179	.2926	.2813
2	.2123	.1480	.2220	.1720
3	.1528	.1004	.1398	.0950
4	.0934	.0682	.0817	.0499
5	.0764	.0463	.0461	.0255
6	.0424	.0315	.0255	.0128
7	.0169	.0214	.0140	.0064
8	.0339	.0145	.0076	.0032
9	.0084	.0099	.0041	.0016
10	.0233	.0067	.0022	.0008
11	.0191	.0045	.0012	.0004
12	.0000	.0031	.0007	.0002
13	.0106	.0021	.0004	.0001

TABLE 13

OBSERVED AND PREDICTED PROBABILITY DISTRIBUTIONS OF L FOR THE MM LIST

L	Observed	One-Element	Bower & Theios	PARR
0	.1380	.3209	.1617	.3507
1	.1464	.2179	.2438	.2403
2	.1825	.1480	.1983	.1550
3	.1549	.1004	.1392	.0974
4	.0955	.0682	.0924	.0604
5	.0764	.0463	.0598	.0372
6	.0573	.0315	.0383	.0228
7	.0191	.0214	.0244	.0140
8	.0276	.0145	.0155	.0086
9	.0233	.0099	.0098	.0053
10	.0148	.0067	.0062	.0032
11	.0339	.0045	.0039	.0020
12	.0084	.0031	.0025	.0012
13	.0212	.0021	.0016	.0007

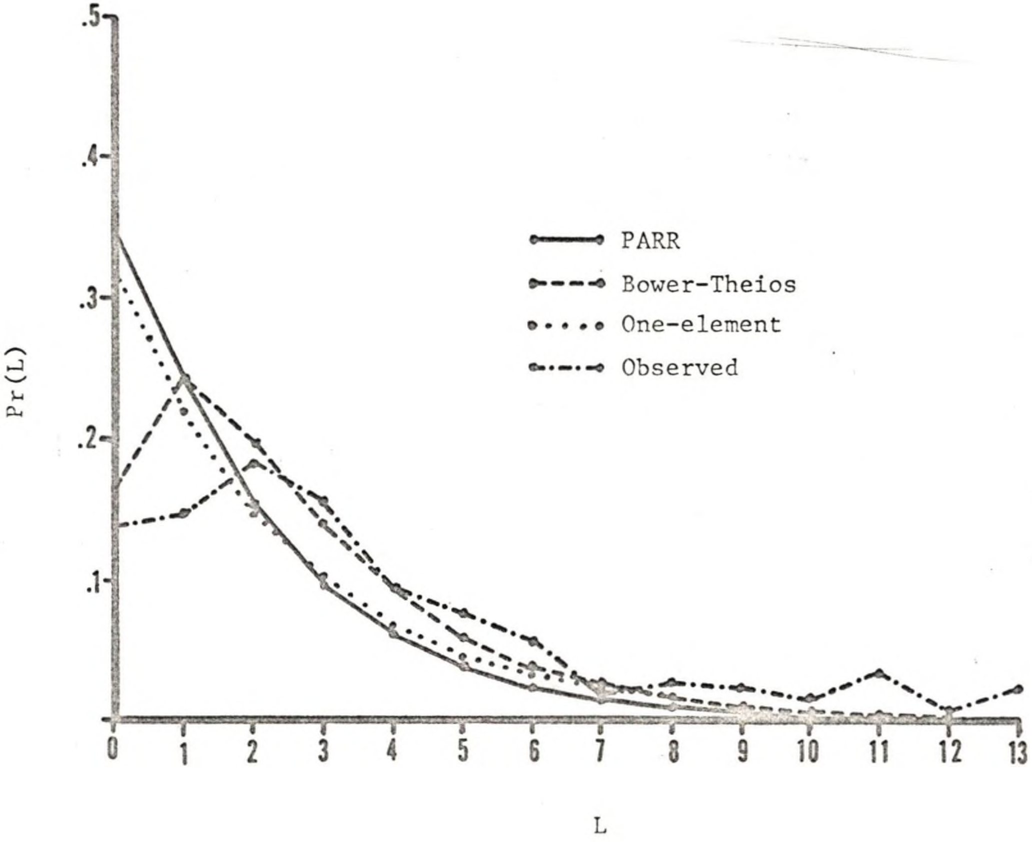

Fig. 9. Observed and predicted probability distributions of L for the MM list.

probabilities before learning are only chance. The Bower-Theios model and the PARR model are both intermediate state models and are therefore inappropriate for the HM list. For this reason only the one-element model's fit to the HM data will be examined.

The c parameter for the one-element model was determined to be .7037. The observed and predicted learning curves are shown in Figure 10 and Table 14. Again, the one-element model predicts a more rapid learning rate than is observed. It is evident that the estimate of c is much too high. The correct response probability on the first trial is predicted to be just slightly more than twice what is observed. The one-element model predicts that by the second trial over 90 percent of the items will be recalled while in reality it takes four trials before 90 percent of the items have been recalled. This gross inaccuracy of the one-element model is surprising in light of the success the model has had with rapidly learned lists.

Predictions of the mean J, T and L from the one-element model are compared to the observed means in Table 15. Again, the one-element model predicts the same mean for J, T, and L but in this case is rather accurate because of the small differences between the observed means of J, T, and L. Even though the probability of correct response before learning is only about chance for the HM list the correct guesses do not appear to occur at random during the precriterion trials. If correct guesses occurred at random the observed $\overline{J}$ should have been about half of the $\overline{L}$. The fact that it is considerably greater than half indicates that correct guesses tended to be made later in the precriterion trials.

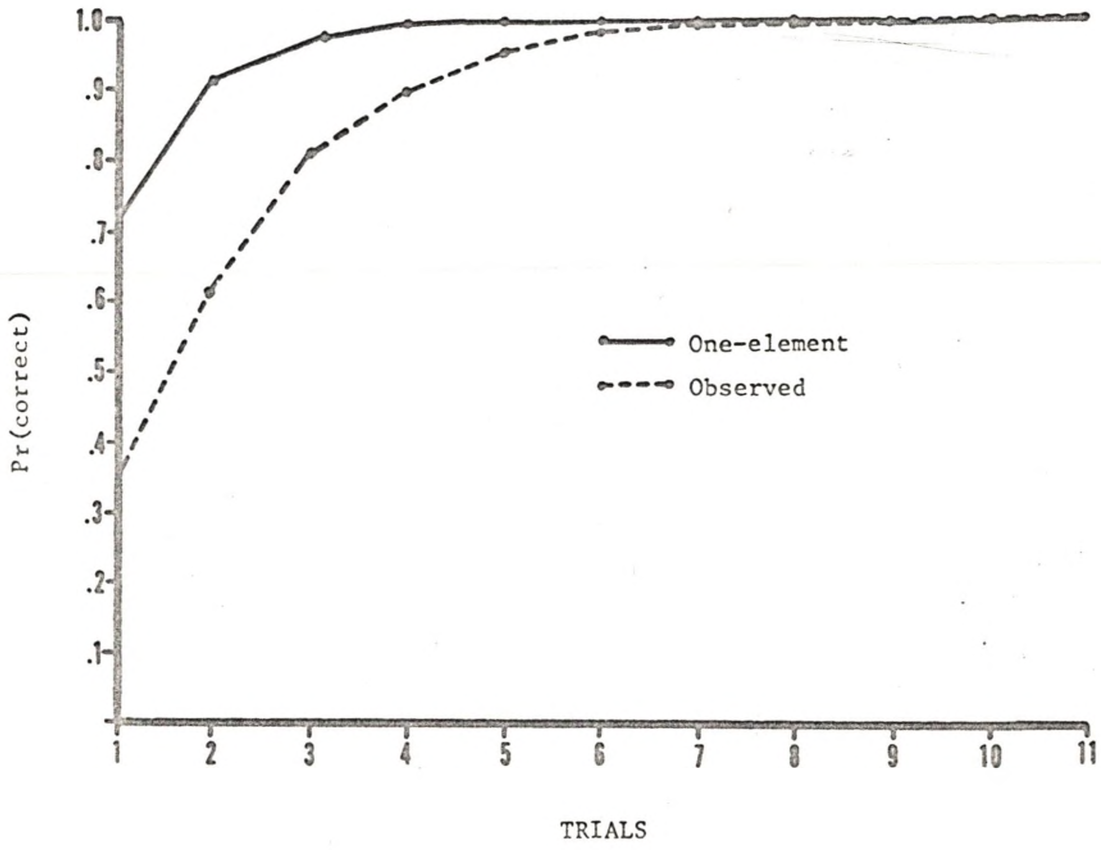

Fig. 10. Observed and predicted learning curves for the HM list.

TABLE 14

OBSERVED AND PREDICTED LEARNING CURVES FOR THE HM LIST

Trials	Pr(Correct) Observed	One-Element
1	.3413	.7037
2	.6063	.9122
3	.8063	.9740
4	.9009	.9923
5	.9569	.9977
6	.9812	.9993
7	.9925	.9998
8	.9975	.9999
9	.9975	.9999
10	1.0000	.9999

TABLE 15

OBSERVED AND EXPECTED MEANS OF J, T, AND L FOR THE HM LIST

	Observed	One-Element
J	1.2688	1.4210
T	1.4210*	1.4210
L	1.5187	1.4210

*Used in parameter estimation.

The predicted and observed probability distributions of J for the HM list are shown in Table 16. The observed J distribution shows that the probability that J=0 is about equal to the probability that J = 1, whereas the one-element model predicts a vast difference between these two points. The one-element model's prediction for J = 0 is clearly too high and its prediction of all other J's is clearly too low.

TABLE 16

OBSERVED AND PREDICTED PROBABILITY DISTRIBUTIONS OF J FOR THE HM LIST

J	Pr(J) Observed	One-Element
0	.3417	.7037
1	.3000	.2035
2	.2000	.0618
3	.1000	.0183
4	.0375	.0054
5	.0125	.0016
6	.0063	.0005
7	.0000	.0001
8	.0000	.0000
9	.0021	.0000
10	.0000	.0000
11	.0000	.0000

Table 17 shows the obtained and predicted distributions of T for the HM list and Table 18 shows the obtained and predicted distributions of L. The obtained T and L distributions are very similar; both are

TABLE 17

OBSERVED AND PREDICTED PROBABILITY DISTRIBUTIONS OF T FOR THE HM LIST

T	Pr(T) Observed	One-Element
0	.2895	.7037
1	.3145	.2085
2	.2104	.0618
3	.1062	.0183
4	.0500	.0054
5	.0145	.0016
6	.0125	.0005
7	.0000	.0001
8	.0000	.0000
9	.0020	.0000
10	.0000	.0000
11	.0000	.0000

TABLE 18

OBSERVED AND PREDICTED PROBABILITY DISTRIBUTIONS OF L FOR THE HM LIST

L	Pr(L) Observed	One-Element
0	.2895	.7037
1	.2750	.2085
2	.2145	.0618
3	.1062	.0183
4	.0645	.0054
5	.0229	.0016
6	.0125	.0005
7	.0041	.0001
8	.0000	.0000
9	.0020	.0000
10	.0000	.0000
11	.0000	.0000

smooth for T and L equal to 0 and 1. The distributions indicate that in actuality T and L are distributed very evenly over the first few points. The one-element model predicts a much more skewed distribution than is observed. The one-element prediction that approximately 70 percent of the items show correct responses on the first trial, are learned with no errors and are learned on the first trial is far too high. The one-element model was formulated with the assumption that the probability of a correct guess should be the same for each precriterion trial. However, it was shown earlier that the probability of a correct guess seemed to increase as practice increased. This would have a tendency to string out the obtained J, T, and L distributions and may be responsible for the bad fit of the one-element model to the distributions.

CHAPTER VI

DISCUSSION AND CONCLUSIONS

It is evident that none of the models tested provided what could be considered a good fit to the data. In all the cases the models predicted more rapid learning rates than were observed. The one-element model and the Bower-Theios model were remarkably similar in their predictions of the learning curve for the LM and MM lists. They predicted, however, quite different probability distributions of J, T, and L. The agreement of the Bower-Theios model and the one-element model in predicting the learning curves and their contrasting disagreement in predicting the probability distributions exemplifies the advantage of the intermediate state models. Models with transition states do not necessarily lose accuracy in learning curve predictions but they have the advantage of being more flexible in the prediction of the fine grain aspects of the data.

The fact that the s and ε values obtained for the Bower-Theios model were not equal ($s = .3070$ and $\varepsilon = .2470$ for the LM list and $s = .4819$ and $\varepsilon = .2938$ for the MM list) indicates that the probability of moving into the learned state following an error is not as great as following a success. These differential learning probabilities make the PARR assumption that recall is independent of correct response probability seem tenuous.

Two of the models (Bower-Theios and PARR) showed some of the flexibility needed to describe the data but were shown to be inaccurate in predicting obtained values. Better methods of parameter estimation may be a solution to this problem. It is suspected that the parameter estimates obtained from the Bower-Theios model and then used in the PARR model were not appropriate. It would seem that the estimates of a and b were not as descrepant as the recognition-recall reasoning would suggest. With recognition shown to be a more simple learning process than recall one would expect a rather wide range between estimates of the recognition and recall learning probabilities.

An interesting feature of the results is that the one-element model did no better in predicting data from the easily learned list than it did in predicting the learning of the difficult list. The one-element model has had its greatest success with easily learned lists, however, these lists generally were shorter and had higher guessing probabilities. In Bower's (1961) original test of the model only two response alternatives were used, making the guessing probability .50. It is of interest to note that if we were to obtain an estimate of c for the LM list by assuming the guessing probability to be .50 we would obtain a c value of about .10 which, it may be remembered, would have provided a much better fit to the data. The same holds true for the MM and HM lists. It is likely that the one-element model is not effective unless the guessing probability is high. The presence of more correct responses before learning, as a result of a higher guessing probabilities would tend to increase the probability of small J's, T's, and L's in the obtained data. This would be more in line with the predictions from the model. This idea may partially explain the one-element model's surprisingly bad fit to the HM data.

The general failure of the PARR to predict many details of the data satisfactorily tends to reflect on the adequacy of the conventional two-stage analysis of paired-associate learning (Underwood and Schulz, 1960). Two stage theory posits a response learning phase which precedes or occurs simultaneously with an associative stage. The PARR model has similar formal properties. Recognition is assumed to precede or occur concomitantly with recall. An advocate of two-stage analysis might argue that no response learning phase was involved in the present experiment because the response terms are the already-known numbers 1 to 16. If that argument is accepted, then the one-element model is the Markov model, which is analagous to a one-stage analysis of paired-associate learning. One could make a case for the obtained distributions being multiply determined, so that more parameters may be needed if Markov models are to be successfully applied to paired-associate data. Aside from the computational problems posed by such models, there is some reason to believe that paired-associate learning is a very simple process, at least with HM pairs. It would be a shame if mathematical models of learning had to be more complex than psychological reality in order to work.

REFERENCES

REFERENCES

Archer, J. A. A re-evaluation of the meaningfulness of all possible CVC trigrams. Psychological Monographs, 1960, 74, 1-23.

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

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

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

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

Bower, G. H., & Theios, J. A learning model for descrete performance levels. In R. C. Atkinson (Ed.), Studies in Mathematical Psychology. Stanford: Stanford University Press, 1964, 1-31.

Bush, R. R., & Estes, W. K. (Eds.) Studies in mathematical learning theory. Stanford: Stanford University Press, 1959.

Bush, R. R., & Mosteller, F. A comparison of eight models. In R. R. Bush and W. K. Estes (Ed.), Studies in mathematical learning theory. Stanford: Stanford University Press, 1959, 293-307.

Calfee, R. C., & Atkinson, R. C. Paired-associate models and the effect of list length. Journal of Mathematical Psychology, 1965, 2, 255-267.

Estes, W. K. Toward a statistical theory of learning. Psychological Review, 1950, 57, 94-107.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), Psychology: a study of a science. Vol. II. New York: McGraw-Hill, 1959, 380-491.

Estes, W. K. Learning theory and the new mental chemistry. Psychological Review, 1960, 67, 207-223.