JOURNAL OF MATHEMATICAL PSYCHOLOGY: 2, 254-265 (1965)

Paired-Associate Models and the Effects of List Length[1]

ROBERT C. CALFEE

University of Wisconsin, Madison, Wisconsin

AND

RICHARD C. ATKINSON

Stanford University, Stanford, California

A two-process Markov model for paired-associate learning is presented in which stimulus-response associations may pass through an intermediate or short-term memory state before learning is complete. In the short-term state, forgetting may occur, and in the trial-dependent-forgetting (TDF) model, the likelihood that forgetting takes place on any trial is postulated to be a function of the number of S-R pairs remaining to be learned on that trial. To determine the quantitative accuracy of the model, a paired-associate experiment was conducted in which list length was varied. Specific response-sequence frequencies from experimental lists of 9, 15, and 21 items were reasonably well predicted by the TDF model. A much better account of the data was obtained by a revision of the model, in which it was assumed that the probability of learning on a given trial depended on whether an item was still in short-term memory or had been forgotten. Comparative predictions from the linear and all-or-none models, as well as an alternative two-process Markov model, are also presented.

In a recent article Atkinson and Crothers (1964) applied several variations of a three-state Markov model to data from a number of different paired-associate learning experiments. Common to most of the models was the assumption that paired-associate learning is a two-stage process in which a given stimulus item may be viewed as initially moving from the original unconditioned state to an intermediate short-term storage state. The stimulus item next moves from the intermediate state to the absorbing state, or long-term storage. As an example of a specific model based on this type of two-stage process, we consider the LS-3 model, for which the transition matrix and response probability vector are as follows:

$$\begin{array}{c} \\ L_n \\ S_n \\ U_n \end{array} \begin{array}{c} \begin{array}{ccc} L_{n+1} & S_{n+1} & U_{n+1} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ a & (1-a) & 0 \\ ca & c(1-a) & 1-c \end{bmatrix} \end{array} \qquad \begin{array}{c} \Pr(\text{Correct} \mid \text{row state}) \\ \begin{bmatrix} 1 \\ 1-f+fg \\ g \end{bmatrix} \end{array}. \tag{1}$$

[1] Support for this research was provided by the National Aeronautics and Space Administration.

Each item is assumed to begin in state U; changes in state from one trial to the next are governed by the transition probabilities given above. In the unconditioned state correct responses occur randomly with probability g, usually assumed to be equal to the reciprocal of the number of response alternatives. In the long-term state, the probability of a correct response is 1. If an item is in state S following the nth reinforced presentation, it is assumed that with probability f, the item is forgotten before the $(n + 1)$st presentation. Thus, for an item in the short-term state, the probability of a correct response is equal to the probability that there is no forgetting, or that forgetting takes place and a correct response occurs by chance, hence

$$\Pr(Correct \mid State\ S) = 1 - f + fg.$$

For detailed derivations from this model, the reader is referred to the original article by Atkinson and Crothers (1964).

The trial-dependent-forgetting (TDF) model is an alternative formulation considered by Atkinson and Crothers which is similar to the LS-3 model in many respects. Each item in a list may be in one of three states: (a) state U is an unlearned state, in which the subject guesses at random from the set of response alternatives, (b) state S is a short-term memory state, and (c) state L is a long-term state. The subject will always give a correct response to an item if it is in either state S or state L. However, it is possible for an item in state S to be forgotten, i.e., to return to state U, whereas once an item moves to state L it is learned, in the sense that it will remain in state L for the remainder of the experiment. In the TDF model, forgetting involves a return from state S to state U, and the probability of this return is postulated to be a function of the number of other items that remain to be learned on any given trial.

More specifically, two types of events are assumed to produce transitions from one state to another in the TDF model: (a) the occurrence of a reinforcement, i.e., the paired presentation of the stimulus item together with the correct response alternative and (b) the presentation of an unlearned stimulus-response pair (an item which is not in state L) between successive occurrences of a particular item. The associative effect of a reinforcement is described by matrix $\mathbf{A}$ below:

$$\mathbf{A} = \begin{array}{c} \\ L \\ S \\ U \end{array} \begin{array}{c} \begin{array}{ccc} L & S & U \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ a & 1-a & 0 \\ a & 1-a & 0 \end{bmatrix} \end{array}. \tag{2}$$

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

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

$$\mathbf{F} = \begin{array}{c} \\ L \\ S \\ U \end{array} \begin{array}{c} \begin{array}{ccc} L & S & U \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-f & f \\ 0 & 0 & 1 \end{bmatrix} \end{array} . \tag{3}$$

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

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

$$\mathbf{T}_n = \begin{array}{c} \\ L_n \\ S_n \\ U_n \end{array} \begin{array}{c} \begin{array}{ccc} L_{n+1} & S_{n+1} & U_{n+1} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ a & (1-a)(1-F_n) & (1-a)F_n \\ a & (1-a)(1-F_n) & (1-a)F_n \end{bmatrix} \end{array} \tag{4}$$

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

Unfortunately, there is no way of determining from the data the exact value of ξ_n, the number of interpolated pairs (IP) that are not in state L. If an incorrect response is given to an IP, then it must certainly be in state U; if a correct response occurs, however, then the IP may be in either state S or L, or it may even be that it is in state U and a correct response is given by chance. The most precise estimate of ξ_n would involve keeping track of the number of IP's for each item, counting all those IP's with errors as unlearned IP's, and then estimating the number of unlearned items among the remaining IP's, keeping in mind the fact that successes might occur in any of the three states. However, no record was kept of specific inter-item presentation events in the experiment to be reported, and the amount of bookkeeping required to perform the analysis seems formidable in the absence of access to an on-line computer.

Since the exact value of ξ_n is indeterminate, the following approximation was used. In the typical paired-associate experiment, a trial consists of the random presentation of each item in the list of X items. Between nth and the $(n + 1)$st presentations of a given item i from the list, $(j + k)$ IP's may intervene; j on trial n and k on trial $n + 1$, where $j, k = 0, 1 \ldots X - 1$. The probability of j IP's on trial n is the probability that item i is in position $X - j$, which is $1/X$; whereas the probability of k IP's on trial $n + 1$ is the likelihood that i is in position $k + 1$, which also is $1/X$. Thus for each combination of j and k, the probability of the combination occurring is $1/X^2$. For each of these combinations the *average* value of ξ_n will be $j(1 - l_n) + k(1 - l_{n+1})$, where l_n is the probability of being in state L on trial n. Using this average as an approximation,

$$F_n = 1 - \sum_{j=0}^{X-1} \sum_{k=0}^{X-1} \frac{1}{X^2} (1 - f)^{[j(1-l_n)+k(1-l_{n+1})]} \tag{5}$$

$$= 1 - \frac{1}{X^2} \left\{ \frac{1 - (1 - f)^{X(1-l_n)}}{1 - (1 - f)^{(1-l_n)}} \right\} \left\{ \frac{1 - (1 - f)^{X(1-l_{n+1})}}{1 - (1 - f)^{(1-l_{n+1})}} \right\}.$$

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

$$\Pr(e_{n+1} \mid e_n) = (1 - a)(1 - g)F_n. \tag{6}$$

In other words, $\Pr(e_{n+1} \mid e_n)$ is predicted to decrease over trials, a finding which has been reported by a number of investigators.

Another interesting characteristic of the TDF model is that it is able to account quantitatively for variations in list length. For example, longer lists are generally more difficult to learn, taking the mean number of errors per item as a criterion. From the matrix $\mathbf{T}_n$, it is easy to show that the probability of an error on trial $n + 1$ is

$$\Pr(e_{n+1}) = (1 - g)(1 - a)^n F_n. \tag{7}$$

For fixed values of a and f, the amount of forgetting before transition to state L increases with list length; as X increases, so does F_n, as may be seen by referring to Eq. 5. Thus increasing list length will increase $\Pr(e_{n+1})$, and so mean total errors also will increase. Data from the experiment to be described below, in which list length was varied over three values, were used to evaluate this feature of the model.

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

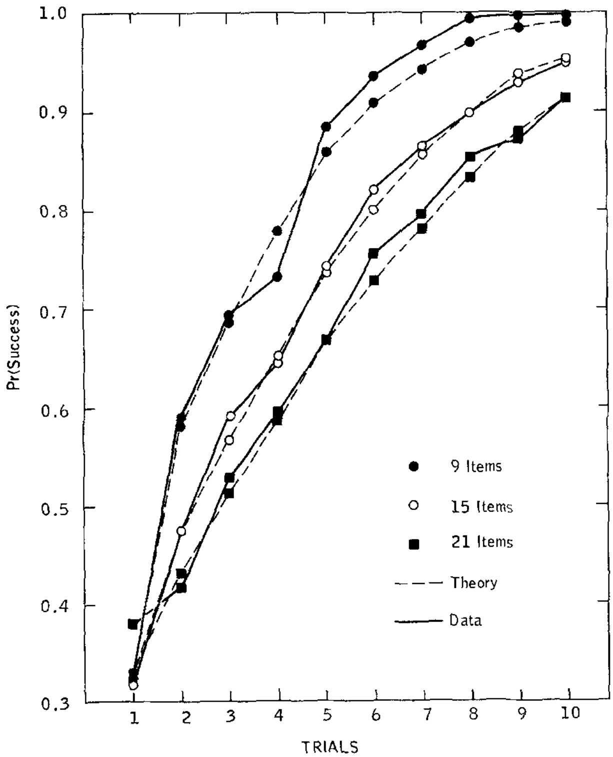

FIG. 1. Average probability of a success on trial *n* for three groups with different list lengths. See text for description of theoretical curves.

Figure 1 presents the mean learning curves for the three experimental groups. The curves are ordered on the list length variable, with the longer lists producing a slower rate of learning. The conditional error curves, $\Pr(e_{n+1} \mid e_n)$,are shown in Fig. 2, and also are ordered according to list length. It is apparent that the conditional probability is decreasing over trials.

Parameter estimates for the LS-3 and TDF models were obtained by applying the chi-square minimization method described by Atkinson and Crothers (1964). The data used in parameter estimation were the sequences of successes and errors from trials 2 through 5 and trials 6 through 9. The 16 possible combinations of correct responses (c) and errors (e) for a four-trial block are listed in Table 1 together with the observed frequencies of each combination for the three experimental groups. Thus, the sequence consisting of four errors (*eeee*) on trials 2 through 5 was observed in 6 of 225 item protocols in Group *9*, in 30 out of 375 protocols in Group *15*, and in 55 out of the 525 protocols in Group *21*. The sequences for trials 6 to 9 are listed in Table 2. In all of the theoretical analyses g was set equal to $\frac{1}{3}$, the reciprocal of the number of response alternatives.

For each of the models, the theoretical expression for the probability of a four-trial sequence was obtained. Following the notation of Atkinson and Crothers, let $0_{i,j,n}$ be the ith four-tuple in Table 1 for Group j ($j = 9, 15, 21$) where the sequence begins at trial n. Let $\hat{N}(0_{i,j,n})$ be the observed frequency of this four-tuple, and let $\Pr(0_{i,j,n}; p)$ be the predicted probability for a particular choice of the parameters p of the model. The expected frequency may be obtained by taking the product of $\Pr(0_{i,j,n}; p)$ with T, the total number of item protocols in Group j. We then define the function

$$\chi^2_{i,j,n} = \frac{[N(0_{i,j,n}; p) - \hat{N}(0_{i,j,n})]^2}{N(0_{i,j,n}; p)}. \tag{8}$$

A measure of the discrepancy between a model and the data from Group j is found by summing Eq. 8 over the sixteen possible sequences for both of the four-trial blocks; i.e.,

$$\chi^2_j = \sum_{i=1}^{16} \chi^2_{i,j,2} + \sum_{i=1}^{16} \chi^2_{i,j,6}. \tag{9}$$

For the LS-3 model, estimates of the three parameters a, f, and c were obtained by minimizing χ^2_j separately for each experimental group. Equation 9 was also used to obtain estimates of c and θ for the all-or-none and linear models, respectively, for each of the three experimental groups (these models are described in the paper by Atkinson and Crothers).

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

TABLE 1

Observed and Predicted Frequencies for Response Sequences from Trials 2 through 5

Trial 2345	9 Items Obs.	9 Items Linear	9 Items All-or-none	9 Items LS-3	9 Items TDF	9 Items TDF-Rev.	15 Items Obs.	15 Items Linear	15 Items All-or-none	15 Items LS-3	15 Items TDF	15 Items TDF-Rev.	21 Items Obs.	21 Items Linear	21 Items All-or-none	21 Items LS-3	21 Items TDF	21 Items TDF-Rev.
cccc	83	59.0	88.4	83.1	68.4	77.2	98	39.9	103.7	97.3	97.2	90.7	97	45.4	112.6	106.5	129.1	107.5
ccce	3	9.5	1.3	4.3	6.0	4.2	10	17.8	3.8	7.3	6.9	6.7	11	24.2	6.8	11.2	8.1	9.0
ccec	10	15.2	3.0	7.5	9.3	8.0	13	23.9	6.6	11.6	11.1	11.1	14	31.5	10.3	16.7	13.0	13.7
ccee	4	2.4	2.7	2.6	5.1	3.7	10	10.7	7.6	6.7	9.0	9.2	12	16.8	13.5	12.0	12.6	14.5
cecc	18	25.7	10.4	15.9	16.7	17.2	25	33.1	17.3	23.3	22.1	22.7	35	42.2	23.0	30.8	27.6	27.3
cece	2	4.1	2.7	3.1	5.9	4.4	4	14.8	7.6	7.5	9.9	9.9	14	22.5	13.5	13.2	13.5	15.1
ceec	10	6.6	6.1	5.9	9.2	8.5	7	19.8	13.3	13.0	15.8	16.5	17	29.3	20.7	20.5	21.6	23.3
ceee	3	1.1	5.3	2.9	5.0	3.9	12	8.9	15.2	11.9	12.8	13.6	20	15.6	27.1	22.0	21.0	24.5
eccc	40	48.3	41.9	44.4	35.3	39.5	58	48.7	57.3	60.8	55.1	54.6	78	59.4	67.6	73.1	74.6	67.6
ecce	3	7.8	2.7	4.5	6.7	4.9	6	21.8	7.6	9.5	10.7	10.5	15	31.7	13.5	16.0	14.1	15.6
ecec	12	12.5	6.1	8.3	10.4	9.4	16	29.2	13.3	16.0	17.1	17.4	22	41.2	20.7	24.4	22.6	24.0
ecee	2	2.0	5.3	3.5	5.7	4.4	12	13.0	15.2	12.3	13.9	14.3	30	22.0	27.1	22.6	22.0	25.3
eecc	14	21.1	20.8	20.0	18.8	20.2	31	40.5	34.6	35.7	34.1	35.4	47	55.2	46.0	48.6	47.9	47.6
eece	2	3.4	5.3	4.5	6.6	5.1	11	18.1	15.2	14.0	15.3	15.5	16	29.5	27.1	24.9	23.4	26.5
eeec	13	5.4	12.2	9.3	10.3	9.9	32	24.2	26.5	24.5	24.3	25.7	42	38.3	41.4	39.0	37.5	40.6
eeee	6	0.9	10.7	5.3	5.6	4.6	30	10.8	30.3	23.6	19.8	21.2	55	20.4	54.1	43.6	36.4	42.8
χ^2		73.5	42.5	13.7	18.8	11.0		173.2	30.3	14.1	22.3	21.7		180.5	21.8	12.2	27.6	17.0

TABLE 2

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

Trial	9 Items						15 Items						21 Items					
6789	Obs.	Linear	All-or-none	LS-3	TDF	TDF-rev.	Obs.	Linear	All-or-none	LS-3	TDF	TDF-rev.	Obs.	Linear	All-or-none	LS-3	TDF	TDF-rev.
cccc	205	177.7	192.2	195.1	166.2	197.2	271	156.3	263.9	262.2	250.6	260.3	319	178.1	309.7	312.7	340.9	317.1
ccce	0	5.3	0.3	2.1	4.2	1.1	6	26.1	1.6	5.3	5.2	3.3	8	39.5	3.5	9.6	5.8	5.2
ccec	0	7.9	0.7	3.4	6.3	2.6	8	32.8	2.7	8.0	8.2	6.6	13	48.4	5.4	13.7	9.2	9.2
ccee	0	0.2	0.6	0.9	1.9	0.3	2	5.5	3.1	2.6	4.0	2.6	4	10.7	7.1	5.9	6.0	6.1
cecc	12	5.0	2.5	6.1	10.3	6.4	13	41.6	7.1	13.8	14.4	14.4	27	59.8	12.0	22.5	17.4	19.2
cece	0	0.4	0.6	0.9	2.2	0.5	1	6.9	3.1	2.9	4.6	3.1	6	13.3	7.1	6.3	6.8	6.8
ceec	1	0.5	1.5	1.5	3.4	1.2	2	8.7	5.4	4.8	7.2	6.2	11	16.3	10.8	9.6	10.7	12.1
ceee	0	0.0	1.3	0.5	1.0	0.2	5	1.5	6.2	3.6	3.5	2.4	10	3.6	14.1	8.8	6.9	8.0
eccc	13	18.3	10.1	12.6	18.1	15.4	24	53.5	23.5	28.7	29.2	33.7	55	74.8	35.3	44.2	39.3	45.8
ecce	0	0.5	0.6	1.0	2.6	0.6	2	8.9	3.1	3.5	5.3	3.6	10	16.6	7.1	7.4	7.6	7.5
ecec	0	0.8	1.5	1.7	4.0	1.5	11	11.2	5.4	5.6	8.3	7.2	5	20.3	10.8	11.1	12.0	13.2
ecee	0	0.0	1.3	0.5	1.2	0.2	1	1.9	6.2	3.7	4.1	2.8	3	4.5	14.1	9.1	7.8	8.8
eecc	1	1.2	5.0	3.5	6.4	3.7	15	14.2	14.2	11.9	14.6	15.8	17	25.1	24.0	21.3	22.6	27.4
eece	0	0.0	1.3	0.6	1.4	0.3	5	2.4	6.2	4.2	4.7	3.4	7	5.6	14.1	10.0	8.9	9.8
eeec	0	0.1	2.9	1.2	2.1	0.7	5	3.0	10.9	7.3	7.3	6.8	11	6.8	21.6	15.6	14.0	17.3
eeee	0	0.0	2.6	0.5	0.6	0.1	4	0.5	12.4	6.9	3.6	2.7	19	1.5	28.3	17.1	9.1	11.5
χ^2		25.5	21.3	21.4	44.6	15.8		210.0	52.0	15.4	17.1	18.9		428.9	76.0	16.6	38.6	31.2

over all three groups, so that parameters a and f were found that minimized the function

$$\chi^2 = \chi_9^2 + \chi_{15}^2 + \chi_{21}^2 \tag{10}$$

where the χ_j^2 are defined in Eq. 9. The minimization was carried out by using a digital computer to search a grid on the parameter space, yielding parameter values accurate to three decimal places.

The χ^2 values obtained by minimizing Eqs. 9 or 10 do not have a chi-square distribution, since the frequencies in the two 4-trial sets are not independent. However, it can be shown that in general if one interprets the value obtained from this procedure as a true χ^2 then the statistical test will be conservative; i.e., it will have a higher probability of rejecting the model than is implied by the confidence level (for a discussion of this problem, see Atkinson, Bower and Crothers, 1965). In evaluating the minimum χ^2, each set of 16 sequences yields 15 df, since the predicted frequencies are constrained to add to the total number of protocols. Further, it is necessary to subtract one df for each parameter estimate. Thus, there are 27 df for each group in the case of the LS-3 model or 81 df over groups, and 88 df over the three groups for the TDF model.

Tables 1 and 2 present the predicted frequencies of each response sequence for the four models using the minimum χ^2 parameter estimation procedure. Table 3 presents the minimum χ^2 values and the parameter estimates. The LS-3 model, using nine

TABLE 3

PARAMETER ESTIMATES FOR VARIOUS MODELS AND TOTAL χ^2 VALUES OVER GROUPS

Model	Parameter	9 Items	15 Items	21 Items	χ^2 value		
					Trials 2–5	Trials 6–9	Total
Linear	θ	0.32	0.17	0.15	427.2	664.4	1091.6
All-or-none	c	0.30	0.20	0.15	94.6	149.3	243.9
LS-3	a	0.31	0.27	0.24			
	c	0.47	0.27	0.21	40.0	53.3	93.3
	f	0.42	0.27	0.28			
TDF	a	0.16	—	—	68.7	100.4	169.1
	f	0.23	—	—			
TDF revised	a	0.42	—	—			
	b	0.11	—	—	49.6	65.9	115.5
	f	0.19	—	—			

parameters overall ($\chi^2 = 93.3$), does a considerably better job than the TDF model with two parameters ($\chi^2 = 169.1$), even taking into account the difference in number of parameters. However, the TDF model gives a better account of the data than the one-element model where the predictions of the latter model are based on three parameter estimates ($\chi^2 = 243.9$). The fit of the linear model, again based on a separate parameter estimate for each group, is even less adequate ($\chi^2 = 1091.6$).

On the basis of statistical considerations, the LS-3 model might be favored over the alternative representations, but the possibility that behavioral effects of list length variations could be handled within the basic structure of a model rather than by means of a proliferation of parameters seems worth further investigation. Consequently, consideration was given to a revision of the TDF model in which the probability of transition to state L differed according to whether an item is in the state S or state U at the time of reinforcement. If an item is still in state S at the beginning of its nth presentation, then with probability a the reinforcement will effect a move to state L, whereas an item which has been forgotten between the nth and $(n + 1)$st presentations will enter long-term storage with probability b. These changes are represented by modifying the matrix in Eq. 4 as follows:

$$\mathbf{T}_n = \begin{array}{c} \\ L_n \\ S_n \\ U_n \end{array} \begin{array}{c} \begin{array}{ccc} L_{n+1} & \qquad S_{n+1} & \qquad U_{n+1} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ a & (1-a)(1-F_n) & (1-a)F_n \\ b & (1-b)(1-F_n) & (1-b)F_n \end{bmatrix} \end{array} . \tag{11}$$

The revised model remains otherwise unchanged from the previous formulation. As before, in computing results for this model, the expected value of ξ_n will be used to approximate F_n.

The minimum χ^2 procedure was used to obtain estimates of the parameters a, b, and f for the three experimental groups, with the minimization being carried out simultaneously over all three groups. The results for this revised version of the TDF model are presented in Tables 1, 2, and 3. As may be seen, the modification resulted in a considerable improvement in the model. At the expense of one additional parameter, the χ^2 value was reduced by more than 25%, from 169.1 to 115.5. Figure 1 presents predicted mean learning curves from this model, and as might be expected from the χ^2 value, the agreement between theory and data is quite good. The predictions for the conditional error curves, $\Pr(e_{n+1} \mid e_n)$, are presented in Fig. 2, and although the data are somewhat variable, the theoretical curves yield a reasonably good fit.

It is of interest to note that in the revised TDF model, the estimated value of a is almost four times the estimated value of b. In the model, this means that the probability of transition by an unlearned item to state L on any trial is four times greater if the item has *not* been forgotten since its last reinforced presentation (i.e., if it has remained in short-term storage) than if it has been forgotten. This relation between the parameters a and b implies that learning should proceed more efficiently as successive presentations of a given stimulus response pair are separated by fewer interpolated items. In particular, if certain items in a list were to be presented twice

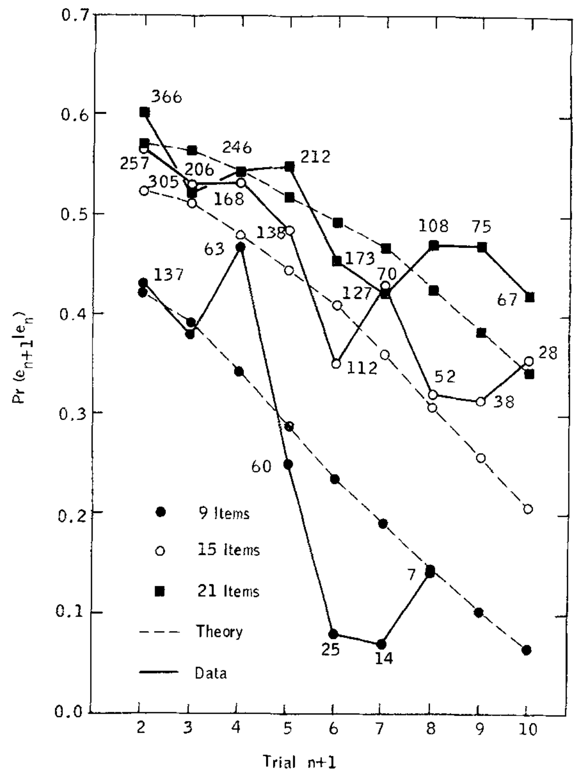

FIG. 2. Average probability of an error on trial $n + 1$, given an error on trial n for three groups with different list lengths. Numerals by data points indicate number of observations on which the data point is based.

during a trial, then fewest errors and trials to criterion should be observed on such items if no other pairs were interpolated between successive presentations within a trial. Interpolated items produce transitions to state U, where learning is less likely to take place. A paired-associate experiment reported by Greeno (1964) yielded results contradicting this prediction. Experimental items presented twice in succession on each trial took the same number of trials to reach criterion (i.e., twice the number of stimulus presentations) as control items presented once per trial, indicating that little or no learning took place during the second presentation on each trial, when an item would almost certainly be in state S. These differing results may be clarified by determining in more detail the function relating learning and performance to number of interpolated items intervening between successive presentations of a given stimulus-response pair.

REFERENCES

ATKINSON, R. C., BOWER, G. H., AND CROTHERS, E. J. *An introduction to mathematical learning theory*. New York: Wiley, 1965.

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

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

RECEIVED: August 19, 1964