

A Forgetting Model for Paired-Associate Learning

HARLEY A. BERNBACH

University of Michigan, Ann Arbor, Michigan

This paper presents a Markov model for paired-associate learning which is based on conclusion drawn from experimental studies of short-term memory. Learning of paired-associates is considered as a decrease in the probability of forgetting an association between trials. The ability of the model to account for experimental results is demonstrated by considering data from two experiments involving paired-associate lists. Observed values of the mean learning curve and response probabilities conditional on the outcome of the preceding trial are compared with the values predicted by the forgetting model and the one-element model. Comparison of parameter values for several experimental situations provides interpretations for the forgetting model's parameters. Finally, the model is applied to an experiment using an RRTTTT design to demonstrate its potential generality.

In recent years, there has been considerable impact by studies of short-term memory on the thinking of theorists in the area of human verbal learning. Peterson and Peterson (1959) showed immediate recall for trigrams to be perfect, with retention falling off with time before recall, and this result has been generalized to recall of paired-associates (Murdock, 1961a, 1961b; Peterson and Peterson, 1962). Evidently, one-trial acquisition of these simple items is assured, although retention is a function of the interval between presentation and recall. Further experimental results (Hellyer, 1962; Peterson, Saltzman, Hillner, and Land, 1962) have indicated that the effect of repetition on short-term memory is to decrease the slope of the forgetting curve which relates recall probability to time before recall. This evidence suggests a view of paired-associate learning in which "learning" consists of a decrease with repetition in the probability of forgetting.

The model that is presented here is a Markov model based on conclusions drawn mainly from the results of these studies, and it is intended principally to serve as a quantification of the theory of paired-associate learning which is suggested by these findings as interpreted by interference theory. The purpose of this paper is to present the "forgetting model" and to demonstrate the feasibility of its approach. This will be accomplished by considering the model in terms of three important criteria for such a quantified theory: its ability to account for or "fit" data from standard paired-associate

learning experiments, the consistency of the values assigned its parameters with the interpretations given them, and the generality of its application.

The model as formulated here has been restricted to the simplest paired-associate task, one in which neither stimulus learning nor response learning is involved. Due to the nature of the complications introduced by these factors, it seemed appropriate to make this limitation at present. Even at this level, however, the forgetting model has more general application than to the standard anticipation design, and this will be demonstrated in the final section of the paper.

PRESENTATION OF THE MODEL

Before the model is presented in detail, it seems wise to trace the development of its principal characteristics from the short-term memory results. First, consider a simple test for short-term retention of paired-associates (e.g., Peterson and Peterson, 1962). Let us define a state C in which the response will be correct with probability 1, and a state G in which the response will be correct with a chance probability, g . Before its first presentation, an item is in state G , since the correct response is not known to the subject. Upon presentation, it moves to state C with probability 1; then, during the interval before recall, it may be "forgotten", i.e. return to state G , with some probability δ_i , where δ_i is a function of the length and content of the retention interval. This process is represented diagrammatically in Fig. 1.

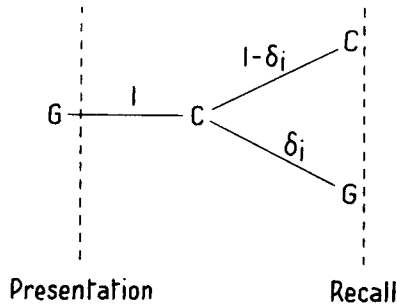


FIG. 1. Tree for state G .

Now consider the situation in which an item is presented repeatedly, with the retention interval and intervening material held constant. We have noted that repetition decreases the slope of the forgetting curve; i.e., the proportion of items forgotten during a fixed retention interval is inversely related to the number of repetitions. In the model, this indicates that the value of δ_i will decrease with repetition of the item. For

perfect learning to take place, δ_i will have to decrease to an asymptote of 0. In the model, the relationship between δ_i and the number of trials is approximated by a step-function which takes on the values δ and 0. Further, it is assumed that the step from δ to 0 can only take place when the item is in state C at the start of a trial, and that after each such trial there is a probability θ that the step will occur. The process is described by establishing a state C' , in which the probability of a correct response is 1 and there is 0 probability of leaving the state, while there remains a probability δ of regression from state C to state G .

Now consider a situation in which the subject has made an error and then been presented with the correct response. It might be inferred from proactive inhibition data (Postman, 1961) that the incorrect response will provide PI to the recall of the correct response, since this situation is an example of the familiar $A-B, A-C$ paradigm. For the model to handle this let us establish an error state, E , in which there is a probability of 1 that the preceding error will be recalled in place of the correct response, and define a parameter β as the probability that this will occur on a given trial.

We can now proceed to a detailed statement of the model, as applied to an ideal paired-associate learning experiment in which the time between presentations of item i is constant, and there are no effects of serial position. The n stimulus items and r responses are completely familiar to the subject, so that only the association must be learned. Each stimulus is presented alone, the response anticipated, and the correct response presented as the information event.

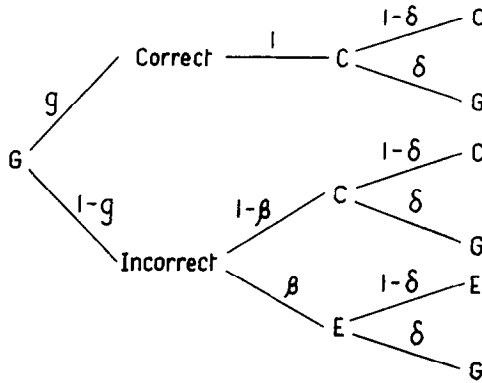


FIG. 2. Transition tree for state G .

Every item is assumed to be in state G at the commencement of the experiment, so the first response will be by chance. If this response is correct, the item will move into state C with probability 1. Before the start of the next trial, the item may return to state G with probability δ or remain in state C with probability $1 - \delta$. If the first response is incorrect, however, the recalled response on the next trial may be the error,

and here β is introduced as the probability that after an incorrect response and reinforcement the item will move into state E rather than into state C . There remains a probability δ of return to state G from either state E or C before the next trial. The branching process starting in state G is diagrammed in Fig. 2.

If an item is in state E at the start of a trial, the response will be incorrect, and the transition probabilities will be identical with those in the case where a guess in state G is incorrect. If the item is in state C , the response will be correct, and the item may move to state C' with probability θ . If it remains in state C , then it may be forgotten with probability δ and move to state G before the next trial. Figure 3 shows the branching process for an item starting in state E and C .

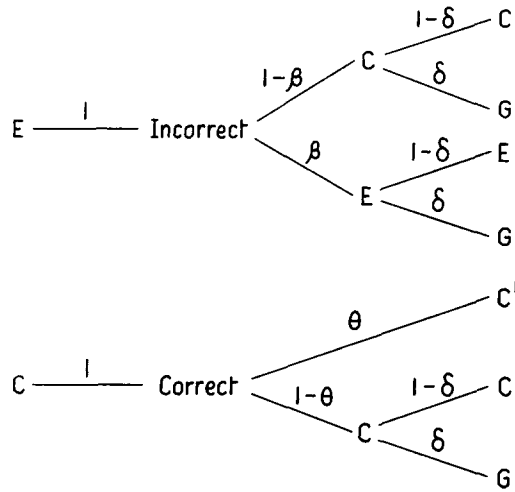


FIG. 3. Transition trees for states E and C .

Since response probabilities depend only on the state, and the transition probabilities are independent of the trial number, the process may be represented by an absorbing Markov chain with the following transition matrix and response probability vector:

$$\begin{array}{c}
 C' \quad C \quad G \quad E \quad \text{Pr (correct)} \\
 \begin{bmatrix}
 1 & 0 & 0 & 0 \\
 \theta & (1-\theta)(1-\delta) & (1-\theta)\delta & 0 \\
 0 & [1-\beta(1-g)](1-\delta) & \delta & \beta(1-g)(1-\delta) \\
 0 & (1-\beta)(1-\delta) & \delta & \beta(1-\delta)
 \end{bmatrix}
 \begin{bmatrix}
 1 \\
 1 \\
 g \\
 0
 \end{bmatrix}
 \end{array} \quad (1)$$

in matrix form, the probability of transition to state C' from state C during an entire trial. We have:

$$\mathbf{Q} = \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & g & 0 & 1-g \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (9)$$

$$\mathbf{R} = \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1-\beta & 0 & \beta \end{bmatrix}, \quad (10)$$

$$\mathbf{F} = \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-\delta & \delta & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \delta & 1-\delta \end{bmatrix}, \quad (11)$$

$$\mathbf{A} = \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ \theta & 1-\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (12)$$

These basic component matrices provide a more formal statement of the model than was given previously, since they do not depend on any particular experimental procedure but can be applied in general to any test, study period, or interval. Likewise, the system parameters can be more formally defined as the appropriate transition probabilities in these component matrices.

From the components, we can construct the transition matrix, \mathbf{T} , for any paired-associate learning procedure. Thus, the matrix for the anticipation method will be the product $\mathbf{AQR}\mathbf{F}$; i.e., \mathbf{A} operates on a trial consisting of test-reinforcement-interval, in that order. This can be seen to yield precisely the transition matrix given in Eq. 1. In similar fashion, matrices can be constructed to handle other procedures, such as prompting or delayed reinforcement.

As an example, we will apply the model to an experiment reported by Atkinson and Calfee (1964). They used the experimental paradigm which has been called $R_1R_2T_1T_2T_3T_4$ (Jones, 1962); i.e., the list of 18 paired-associates was presented twice

and \mathbf{P}' is the error probability vector, $(0, 0, 1 - g, 1)$. It is immediately apparent the \mathbf{U} is simply the product of a scalar and the vector \mathbf{P}' , so that:

$$\Pr(c_{n+1} | e_n) = \delta g + (1 - \beta)(1 - \delta), \quad (5)$$

and is independent of the trial number.

The forgetting model also makes possible the prediction of the probability that an error will be repeated, as opposed to the probability of any error following an error, as follows:

$$\begin{aligned} \Pr(\text{repeat of error}) &= \frac{\Pr(e_n, \text{repeat on } n + 1)}{\Pr(e_n)} \\ &= \frac{\mathbf{W}_n \mathbf{U}'}{\mathbf{W}_n \mathbf{P}'} \end{aligned}$$

where \mathbf{U}' consists of:

$$\begin{aligned} u'_{c'} &= 0, \\ u'_c &= 0, \\ u'_g &= (1 - g) [\delta g + \beta(1 - \delta)], \\ u'_e &= \delta g + \beta(1 - \delta). \end{aligned}$$

This can be simplified in the same manner as Eq. 5, yielding

$$\Pr(\text{repeat of error}) = \delta g + \beta(1 - \delta). \quad (6)$$

EXPERIMENTAL RESULTS

We will now turn to consideration of the model's performance in accounting for experimental data. Experiment I consisted of the learning of a twelve-item list of paired-associates by the anticipation method. Stimuli were single consonants and the response set consisted of the numbers 4 and 6. Subjects were members of elementary psychology classes who participated as part of class demonstrations. In all, 63 subjects were run, in eight groups ranging in size from five to fourteen; each group had a different pairing of stimuli and responses. The stimulus was projected onto a screen for 2.7 sec while the subjects wrote their anticipated responses; the correct response was then shown for 0.3 sec. The complete list was presented, in random orders, ten times to each group.

Parameters were estimated by a minimum χ^2 gridding procedure on the complete response protocols for the first four trials. That is, equations were derived for the

probabilities associated with the sixteen response sequences, $c_1c_2c_3c_4$, $c_1c_2c_3e_4$, etc., and values of the parameters were chosen which jointly minimized the function:

$$\chi^2(\delta, \theta, \beta) = \sum_{i=1}^{16} \frac{[Tp_i(\delta, \theta, \beta) - N_i]^2}{Tp_i(\delta, \theta, \beta)} \quad (7)$$

where $p_i(\delta, \theta, \beta)$ is the predicted probability, for specific values of the parameters, of the occurrence of response sequence i , N_i is the observed number of occurrences of sequence i , and T is the total number of observations. This procedure has been discussed in detail by Atkinson and Crothers (1964).

With a minimum χ^2 value of 12.69 ($df = 12$), the parameters estimated were:

$$\delta = 0.720,$$

$$\theta = 0.448,$$

$$\beta = 0.343.$$

For purposes of comparison, the one-element model (Bower, 1962) was applied to this data also, and the estimated parameter was $c = 0.147$, with $\chi^2 = 26.27$. Table 1

TABLE 1
OBSERVED PROPORTIONS AND PREDICTIONS OF RESPONSE SEQUENCES FOR THE FORGETTING MODEL
AND THE ONE-ELEMENT MODEL, EXPERIMENT I

Response sequence	Observed	Forgetting model predictions	One-element model predictions
cccc	0.196	0.180	0.157
ccce	0.045	0.048	0.039
ccec	0.062	0.050	0.052
ccee	0.044	0.042	0.039
cecc	0.086	0.068	0.084
cece	0.025	0.030	0.039
ceec	0.041	0.045	0.052
ceee	0.034	0.037	0.039
eccc	0.126	0.146	0.157
ecce	0.041	0.043	0.039
ecec	0.046	0.045	0.052
ecce	0.037	0.038	0.039
eecc	0.075	0.086	0.084
eece	0.038	0.038	0.039
eeec	0.049	0.057	0.052
eeee	0.054	0.047	0.039
χ^2		12.69	26.67

shows the observed frequencies of the sixteen response protocols as well as the values predicted by both the forgetting model and the one-element model. Inspection of the table will show that the greatest discrepancy between the data and the predictions of the forgetting model is that the observed proportion of correct guesses on the first trial is higher than the chance rate of $\frac{1}{2}$. It seems reasonable to suppose that on the first trial, when all responses must be guesses, some strategy other than random responding might be used, even though chance guessing is an appropriate assumption for items not recalled on later trials. In fact, by correcting only the first trial prediction for the observed frequency of correct guesses on that trial, which was 0.533, a χ^2 of 9.20 ($df = 11$) was obtained in the estimation procedure with the identical parameters estimated. For reasons of mathematical simplicity, however, it was decided to use a constant guessing rate of 0.5 in the remainder of this discussion.

Predicted and observed response probabilities for the entire series of ten trials are shown in Fig. 4. Since parameters were estimated from the results of the first four

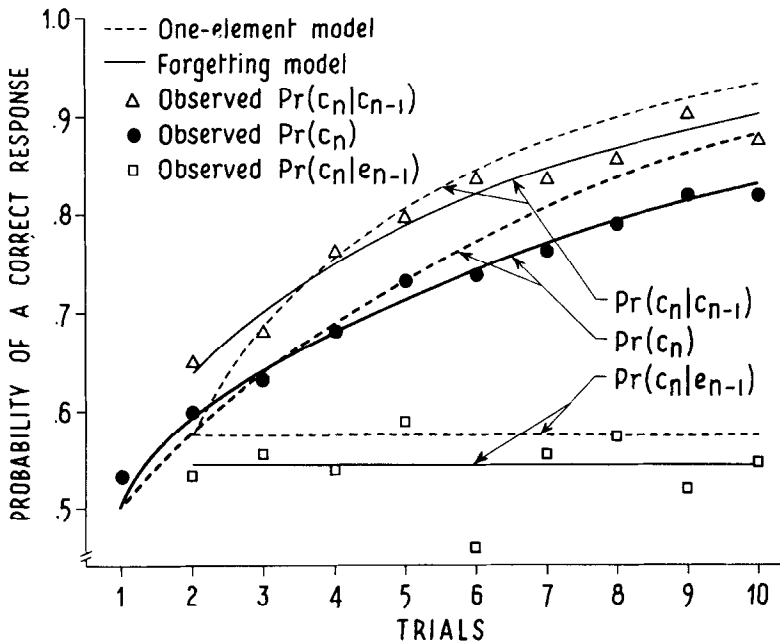


FIG. 4. Predicted and observed response probabilities for the entire series of ten trials.

trials only, the predictions for trials 5 through 10 are extrapolations independent of the results on these trials. It is clear that the forgetting model does a better job in accounting for these data than does the one-element model, which tends to overestimate responses probabilities after several trials. Comparison of the predictions by the two

models of the conditional probabilities on trial 2 is particularly interesting. Since the one-element model does not allow any effect of the response which is made, conditional probabilities depend only on the state on the preceding trial and not on the response. Thus, the one-element model makes the same prediction for the conditional probabilities as for the mean learning curve on trial 2, since all items are in state *G* on trial 1. The forgetting model, on the other hand, makes predictions which do depend on the response made, so these predictions can be different. It is quite clear that the data support the position of response dependence in this case, and this is a strong feature of the model.

The response dependence is a result of the error interference feature of the forgetting model. This feature can be tested more directly by considering the prediction for the probability of repeating an error. In order to do this, we will turn to what we shall refer to as Experiment II; this experiment is actually a part of a larger-scale investigation by J. Baerwaldt, who has kindly allowed the use of this data in this report.

In the experiment, 44 subjects, all undergraduate students who participated as a course requirement, learned a list of twelve nonsense syllable-digit pairs with a three-item set of responses. The anticipation method was used, with a two-second period each for the test and reinforcement. Parameters were estimated by the same procedure used in Experiment I and, with a χ^2 of 9.88 ($df = 12$), the values were:

$$\begin{aligned}\delta &= 0.878, \\ \theta &= 0.379, \\ \beta &= 0.131.\end{aligned}$$

As before, the one-element model was applied also, and a value of $c = 0.060$ was obtained with a minimum χ^2 of 14.58.

The statistics of interest here are shown in Table 2, which gives the response

TABLE 2

RESPONSE PROBABILITIES CONDITIONAL ON A PRECEDING ERROR, EXPERIMENT II

	Observed			Forgetting model predictions			One-element model predictions		
	Correct	Same error	Diff. error	Correct	Same error	Diff. error	Correct	Same error	Diff. error
2	0.412	0.311	0.277	0.399	0.309	0.292	0.373	0.314	0.314
3	0.394	0.351	0.255	0.399	0.309	0.292	0.373	0.314	0.314
4	0.298	0.320	0.282	0.399	0.309	0.292	0.373	0.314	0.314

probabilities conditional on a preceding error. While the one-element model predicts equal probabilities of repeating or making a different error, the forgetting model can be seen to be successful in accounting for the observation that a repeat of an error is more probable than another error.

PARAMETER VALUES

If it is to prove possible to relate the model's parameters directly to experimental variables, the parameter values obtained from various experiments should be consistent with the differences in experimental conditions. In particular, δ should be directly related to the average intertrial interval or the length of the paired-associate list, θ should depend on the opportunity for practice or repetition during and between trials, and β should be a function of factors influencing the amount of interference provided by an error response.

In order to test the performance of the forgetting model to these criteria, we will consider two experiments in addition to Experiment I and II of this report. These were reported by Atkinson and Crothers (1964) as Experiments Ia and Ib; for clarity, we shall refer to them here as simply Experiments A and B. Identical conditions and materials were used in both, Greek letters as stimuli and a three-item response set; they differed only in that A used a nine-item stimulus list and B an eighteen-item list. A χ^2 minimizing procedure on the complete response protocols for Trials 2-5 was used, and simultaneously estimated parameter values for all four experiments are given in Table 3.

TABLE 3
PARAMETER ESTIMATES FOR THE FORGETTING MODEL

	Experiment			
	I	II	A	B
δ	0.720	0.878	0.468	0.562
θ	0.448	0.379	0.667	0.700
β	0.343	0.131	0	0

Experiments A and B provide an excellent test of the effect of increasing list length while other conditions are held constant. As expected, the major difference is in the value of δ ; i.e., there is a greater probability of forgetting during an average intertrial interval of seventeen items than during one of eight items. The value of θ , however,

changes very slightly, evidencing only a small additional opportunity for extra rehearsal of items which are recalled during the longer interval. The value of β can be seen to be very strongly effected by the opportunity for "correction" provided by the reinforcement. Thus, in Experiment I, the 0.3-sec. presentation of the correct response left a very high probability of interference from an error response; this interference was considerably less in Experiment II where the reinforcement was presented for two seconds, and became completely ineffective with the longer presentation times of Experiments A and B.

Comparison of the values of δ and θ for Experiments I and II is not so clear as that between Experiments A and B, since the conditions of Experiments I and II were quite different, particularly in that Experiment II used materials that were considerably more difficult to learn than those used in any of the others. Nevertheless, it is clear that the slower learning of the more difficult material can be mainly accounted for by the greater probability of forgetting between trials, and that where list lengths are the same, as in Experiments I and II, θ will be smaller where greater forgetting provides less opportunity for between-trial rehearsal.

While a great deal more systematic study of the variables influencing parameter values is in order, the present results would seem to indicate that the forgetting model has a great potential for relating its parameters directly to experimental variables.

ALL-OR-NONE PROPERTIES OF THE MODEL

In the special case where $\beta = 0$, as it did in Experiments A and B, the forgetting model reduces to the following three-state model:

$$\begin{bmatrix} 1 & 0 & 0 \\ \theta & (1 - \theta)(1 - \delta) & (1 - \theta)\delta \\ 0 & 1 - \delta & \delta \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ g \end{bmatrix}. \quad (8)$$

This model is identical in form to the three-parameter model identified as T_2 by Atkinson and Crothers (1964), and as they point out, it can be shown to be mathematically equivalent to their LS-2 model. As such, it shares with the one-element model the fundamental properties of all-or-none models in generating predictions for trials after the first guessing trial. Thus, this three-state forgetting model predicts stationarity of response probabilities from the second response to the last error, as well as a constant probability of an error conditional on the occurrence of an error on the preceding trial. When β is greater than 0, the predicted probability of an error prior to the last error increases with trials, since the probability that the item was in state E is greater for trials just prior to the occurrence of an error than for trials further back. The probability of an error following an error is still constant in this case, however, as can be seen from Eq. 5.

While the data reported in this paper as Experiments I and II bear out these pre-

dictions of stationarity quite well, with the predicted increase in the probability of an error prior to the last error being quite clear in Experiment I, where $\beta = 0.343$, other investigators have found results to contradict this. In particular, Atkinson and Crothers (1964) noted that the probability of an error following an error decreased with trials in six of the eight experiments that they reported, and other investigators (e.g., Calfee, Atkinson, and Shelton, 1965) have obtained similar results. Nonstationarity of response probabilities prior to the last error has also been found (Suppes and Ginsberg, 1963), with the probability of an error decreasing with trials. The model, as we have mentioned, predicts stationarity only between the second response and the last error, and Atkinson and Crothers (1964) noted that nonstationarity was not so pronounced if the first guessing trial was not included. Nevertheless, response probabilities were clearly not stationary as the model predicts.

These results do not necessarily reflect a basic inadequacy of the forgetting model, however. In the first place, nonspecific practice effects may have been present in many of these experiments, and this would have caused a decrease with trials in the value of δ which is not specified by the model. It is also possible that the approximation of the relationship between δ and the number of trials with a step function is not altogether appropriate to some experiments. A model incorporating trial-dependent forgetting has been proposed by Atkinson and Crothers (1964), and this formulation has proved quite successful in accounting for some data exhibiting nonstationarity (Calfee *et al.*, 1965). Secondly, though the forgetting model has been designed for situation in which only association learning is taking place, there may well have been some stimulus learning present in these experiments. The success of the LS-3 model of Atkinson and Crothers (1964), which includes a specific stimulus encoding process, in accounting for the nonstationarity aspects of the data would seem to lend support to this hypothesis.

Processes such as these could, of course, be incorporated in the forgetting model. The model has, however, proved able to account for most of the significant aspects of data from the kind of paired-associate learning experiment for which it was designed, and we will choose not to complicate it at this time.

EXTENSION OF THE MODEL

So far in our discussion, we have applied the forgetting model to one specific experimental situation: paired-associate learning by the anticipation method. The model, however, has considerably more general application. To show this, let us break down the experiment into its three component parts: the test, the reinforcement, and the forgetting interval. The model's transition matrix can be likewise broken down into these components, as follows. We let **Q** represent the transition probabilities during test, **R** during reinforcement, and **F** during the intertrial interval, while **A** represents,

in matrix form, the probability of transition to state C' from state C during an entire trial. We have:

$$\mathbf{Q} = \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & g & 0 & 1-g \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (9)$$

$$\mathbf{R} = \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1-\beta & 0 & \beta \end{bmatrix}, \quad (10)$$

$$\mathbf{F} = \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-\delta & \delta & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \delta & 1-\delta \end{bmatrix}, \quad (11)$$

$$\mathbf{A} = \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ \theta & 1-\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (12)$$

These basic component matrices provide a more formal statement of the model than was given previously, since they do not depend on any particular experimental procedure but can be applied in general to any test, study period, or interval. Likewise, the system parameters can be more formally defined as the appropriate transition probabilities in these component matrices.

From the components, we can construct the transition matrix, \mathbf{T} , for any paired-associate learning procedure. Thus, the matrix for the anticipation method will be the product $\mathbf{AQR F}$; i.e., \mathbf{A} operates on a trial consisting of test-reinforcement-interval, in that order. This can be seen to yield precisely the transition matrix given in Eq. 1. In similar fashion, matrices can be constructed to handle other procedures, such as prompting or delayed reinforcement.

As an example, we will apply the model to an experiment reported by Atkinson and Calfee (1964). They used the experimental paradigm which has been called $R_1R_2T_1T_2T_3T_4$ (Jones, 1962); i.e., the list of 18 paired-associates was presented twice

for study, and then the stimuli alone were presented for test four times, with no reinforcement following the tests. The stimulus member of each pair was a single Greek letter, the response was the number 1 or 2, and the subjects for the experiment which we will consider were 60 college students.

Since there were two types of trials, we must construct two transition matrices to describe this experiment. For a study trial, this is simply the product \mathbf{ARF} , with the condition that, since there is no response made which can provide interference, $\beta = 0$. Thus we have:

$$\mathbf{T}_r = \begin{array}{c} C' \\ C \\ G \\ E \end{array} \begin{array}{cccc} C' & C & G & E \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ \theta & (1-\theta)(1-\delta) & (1-\theta)\delta & 0 \\ 0 & 1-\delta & \delta & 0 \\ 0 & 1-\delta & \delta & 0 \end{array} \right] \end{array}. \quad (13)$$

On the test trials, we will assume that a subject not recalling the correct response will repeat his previous response, *if he recalls that*. With this assumption, we can apply the forgetting model simply by augmenting the matrix \mathbf{Q} to include a state E' , from which there is a probability of 0 of forgetting a previously made incorrect response. That is,

$$\mathbf{Q}' = \begin{array}{c} C' \\ C \\ G \\ E \\ E' \end{array} \begin{array}{ccccc} C' & C & G & E & E' \\ \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & g & 0 & 1-g & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right] \end{array}. \quad (14)$$

The matrix for the process is then the product $\mathbf{A}'\mathbf{Q}'\mathbf{F}'$, with each component matrix augmented to include state E' . Thus:

$$\mathbf{T}_t = \begin{array}{c} C' \\ C \\ G \\ E \\ E' \end{array} \begin{array}{ccccc} C' & C & G & E & E' \\ \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ \theta & (1-\theta)(1-\delta) & (1-\theta)\delta & 0 & 0 \\ 0 & g(1-\delta) & \delta & (1-g)(1-\delta) & 0 \\ 0 & 0 & (1-\theta)\delta & (1-\theta)(1-\delta) & \theta \\ 0 & 0 & 0 & 0 & 1 \end{array} \right] \end{array}. \quad (15)$$

If we do not assume that the parameters θ and δ will be the same for both the test and study processes, the entire system has four parameters, θ_t , δ_t , θ_r , and δ_r , while g , since there are two responses used, is simply $\frac{1}{2}$.

The best of the models proposed by Atkinson and Calfee for this experiment is an all-or-none model, also using four parameters. Their transition matrix for a test trial is:

$$\begin{matrix} & C & G & E \\ \begin{matrix} C \\ G \\ E \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ \mu/2 & 1 - (\mu + \delta)/2 & \delta/2 \\ 0 & 0 & 1 \end{bmatrix} & &
 \end{matrix} \tag{16}$$

the starting vector is $(x', 1 - x' - y', y')$ and μ, δ, x' and y' are the system parameters.

Parameters were estimated by deriving equations for the probabilities of the sixteen possible response sequences on the four test trials, and then using a minimum χ^2 gridding procedure such as was used for Experiments I and II. The values of the parameter estimates for both models were as follows:

Forgetting Model		All-or-None Model	
δ_r	0.573	μ	0.207
θ_r	0.393	δ	0.289
δ_t	0.656	x'	0.478
θ_t	0.731	y'	0

TABLE 4

OBSERVED PROPORTIONS AND PREDICTIONS OF RESPONSE SEQUENCES FOR THE ALL-OR-NONE MODEL AND THE FORGETTING MODEL, RRTTTT EXPERIMENT

Response sequence	Observed	All-or-none model predictions	Forgetting model predictions
cccc	0.586	0.578	0.579
ccce	0.020	0.016	0.018
ccec	0.018	0.015	0.014
ccee	0.026	0.026	0.031
cecc	0.015	0.022	0.023
cece	0.018	0.015	0.012
ceec	0.021	0.013	0.015
ceee	0.050	0.054	0.060
eccc	0.040	0.041	0.041
ecce	0.006	0.015	0.011
ecec	0.010	0.013	0.008
ecce	0.009	0.024	0.018
eecc	0.024	0.020	0.022
eece	0.010	0.013	0.011
eeec	0.013	0.012	0.015
eeee	0.134	0.123	0.121
χ^2		30.2	23.0

Observed response frequencies and predictions for both models are given in Table 4. Clearly, the forgetting model does an adequate job in fitting the data from this experiment; the χ^2 value obtained is even less than that for the best model presented by Atkinson and Calfee, providing a good demonstration of the generality of the model.

Of greater interest here, however, is the opportunity for analysis of the experiment provided by the valuation of the parameters. Comparison of δ_r with δ_t indicates that there is a smaller probability of forgetting when the interval is filled with studying other items than when the same interval is filled with active recall. While hardly conclusive, this result would seem to bear quite directly on the question of input versus output interference (Tulving and Arbuckle, 1963). Comparison of θ_r with θ_t is even more striking, the latter being almost double the former. Since we are considering learning to be a change in the probability of forgetting, the difference in the values of the parameter θ under the test and study conditions represents considerable difference in learning rate. One possible interpretation is that active responding produces much greater learning than does simply studying the material. Whatever the interpretation, it is clear that the forgetting model, through application to this kind of experiment, is starting to touch on issues which are of more general interest than simply accounting for data in an experiment of one particular design.

ACKNOWLEDGMENTS

I am indebted to Richard C. Atkinson for the early encouragement of this work, and to Arthur W. Melton for providing the theoretical direction which underlies this work. Appreciation is also expressed to David Birch for his continuing advice and support during the course of this study, as well as to Frank Goode for his assistance with the computer programming and James Baerwaldt for providing some of the data analyzed here.

REFERENCES

- ATKINSON, R. C., & CALFEE, R. C. Mathematical learning theory. In E. Nagel and B. B. Wolman (Eds.) *Psychology and theory of science*. New Yorks: Basic Books, 1964, in press.
- ATKINSON, R. C., & CROTHERS, E. J. A comparison of paired-associate learning models having different acquisition and retention axioms. *J. math. Psychol.*, 1964, 1, 285-315.
- BOWER, G. H. An association model for response and training variables in paired associate learning. *Psychol. Rev.*, 1962, 69, 34-35.
- CALFEE, R. C., ATKINSON, R. C., & SHELTON, T., Jr. Mathematical models for verbal learning. In J. Schade (Ed.) *A Festschrift for Norbert Weiner*, in press, 1965.
- HELLYER, S. Supplementary report: Frequency of stimulus presentation and short-term decrement in recall. *J. exp. Psychol.*, 1962, 64, 650.
- JONES, J. E. All-or-none versus incremental learning. *Psychol. Rev.*, 1962, 69, 156-160.
- MURDOCK, B. B. Short-term retention of single paired-associates. *Psychol. Rep.*, 1961, 8, 280. (a)
- MURDOCK, B. B. The retention of individual items. *J. exp. Psychol.*, 1961, 62, 618-625. (b)

- PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193-198.
- PETERSON, L. R., & PETERSON, M. J. Minimal paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 521-527.
- PETERSON, L. R., SALTZMAN, D., HILLNER, K., & LAND, V. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396-403.
- POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 152-179.
- SUPPES, P., GINSBERG, R. A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning with application to concept formation in children. *Psychol. Rev.*, 1963, **70**, 139-161.
- TULVING, E., & ARBUCKLE, T. Y. Sources of intertrial interference in immediate recall of paired-associates. *J. verb. Learn. verb. Behav.*, 1963, **1**, 321-334.

RECEIVED: March 13, 1964.