PSYCHOMETRIKA—VOL. 22, NO. 2
JUNE, 1957

THEORY OF LEARNING WITH CONSTANT, VARIABLE, OR CONTINGENT PROBABILITIES OF REINFORCEMENT*

W. K. ESTES

INDIANA UNIVERSITY

The methods used in recent probabilistic learning models to generate mean curves of learning under random reinforcement are extended to the general case in which probability of reinforcement may vary in any specified manner as a function of trials and to cases in which probability of reinforcement on a given trial is contingent upon responses or outcomes of preceding trials.

Our purpose is to develop a general model for mean curves of learning under random reinforcement in "determinate" situations. By "determinate" we signify the following restrictions. In these situations the subject is confronted with the same stimulating situation, e.g., a ready signal, at the beginning of each trial. The subject responds with one of a specified set of alternative responses, (A_1 , A_2 , $\cdots$, A_r), and following his response is presented with one of a specified set of reinforcing events, (E_1 , E_2 , $\cdots$, E_r), exactly one reinforcing event E_i corresponding to each possible response A_i . In a T-maze experiment (with correction procedure), A_1 and A_2 correspond to left and right turns; E_1 and E_2 correspond to "food obtained on left" and "food obtained on right", respectively. In a simple prediction experiment with human subjects [3, 8, 9, 10, 11, 13], the responses (A_1 , A_2 , $\cdots$, A_r) correspond to the subject's predictions as to which of a set of "reinforcing lights" (E_1 , E_2 , $\cdots$, E_r) will appear on each trial; instructions are such that the subject interprets the appearance of E_i to mean that response A_i was correct. It is further assumed that one can specify in advance of any trial the probability that any given response will be followed by any given reinforcing event.

From the set-theoretical model of Estes and Burke [4, 6] plus an assumption of association of contiguity, it is possible (see [1, 8]) to derive the following quantitative law describing the change in the probability of response A_i on any trial:

If E_i occurs on trial n

$$p_{i,n+1} = (1 - \theta)p_{i,n} + \theta. \tag{1a}$$

*This paper was prepared while the writer was in residence at the Center for Advanced Study in the Behavioral Sciences, Stanford, California. The research on which it is based was supported by a faculty research grant from the Social Science Research Council.

If E_k $(k \neq j)$ occurs on trial n

$$p_{j,n+1} = (1 - \theta)p_{j,n} . \tag{1b}$$

The quantity $p_{j,n}$ represents the probability of response A_j on trial n, and θ is a parameter satisfying the restriction $0 \leqq \theta \leqq 1$. The parameter θ may vary in value from one organism to another, and for a given organism from one situation to another, but is assumed to remain constant during any given experiment. Functional equations of the form (1a) and (1b) may also be obtained from the stochastic learning model of Bush and Mosteller [2] by imposing suitable restrictions on the parameters.

Now if we can specify the probabilities with which each of the events $[E_j]$ will occur on each trial of a learning experiment, then, given the initial probability of A_j , it becomes a purely mathematical problem to deduce the expected value of $p_{j,n}$ on any trial and thus to generate a predicted learning curve which can be compared with experimental curves. For two special cases, the mathematical problem has already been solved and the desired theoretical curves have been computed and fitted to data [1, 2, 8, 13]. In the first of these, which we shall call the simple non-contingent case, the probability of E_j , hereafter designated π_j , has the same value on all trials of the series regardless of the subject's response. In the second of these, which we shall call the simple contingent case, the probability of E_j on any trial depends upon which response is made by the subject. Thus if the subject makes response A_1 , the probability of E_j is π_{1j} ; if the subject makes response A_2 , the probability of E_j is π_{2j} ; and so on; but the values of π_{ij} remain fixed throughout the series of trials. Now we wish to obtain a more general solution which will yield predicted curves for experiments in which the constancy requirement is removed and the π_j are permitted to vary over a series of trials.

General Solution and Asymptotic Matching Theorem

Let $\pi_{j,n}$ represent the probability that reinforcing event E_j will occur on trial n, with $\sum_j \pi_{j,n} = 1$ for all n. Then given that a subject's probability of making response A_j on trial n is $p_{j,n}$, the expected, or mean, value of the probability* on trial $n + 1$ must be

*Throughout the paper, the quantity p_j should be interpreted as follows. (*a*) In equations dealing with learning on a particular trial, e.g., (1a) and (1b), $p_{j,n+1}$ represents the new probability on trial $n + 1$ for a subject who had the value $p_{j,n}$ on trial n. (*b*) In equations dealing with the expected change on a trial, e.g., (2), (2a), $p_{j,n+1}$ represents the expected value of p_j on trial $n + 1$, where the average is taken over all possible values of $p_{j,n}$ and all possible outcomes of trial n; the term "all possible" is defined for any given situation by the initial values of p_j and the possible sequences of responses and reinforcing events over the first n trials. (c) In solutions giving p_j as a function of n, e.g., (3), (3a), $p_{j,n}$ is the expected value p_j on trial n, where the average is taken over all initial values of p_j and all possible sequences of responses and reinforcing events over the first $n - 1$ trials.

$$p_{i,n+1} = (1 - \theta)p_{i,n} + \theta\pi_{i,n} \,. \tag{2}$$

To obtain (2) average the right hand sides of (1a) and (1b), weighting them by the probabilities $\pi_{i,n}$ and $[1 - \pi_{i,n}]$, respectively, that E_i will and will not occur.

Some general asymptotic properties of the model can be clearly displayed if we consider, not simply $p_{i,n}$, the probability of a response on a particular trial, but the expected proportion of response occurrences over a series of trials. The latter quantity, which we shall designate $\bar{p}_i(n)$, must of course satisfy the relation

$$\bar{p}_i(n) = \frac{1}{n} \sum_{v=1}^{n} p_{i,v} \,.$$

Substituting into the right side of this expression from (2), we obtain

$$\begin{aligned}\bar{p}_i(n) &= \frac{1}{n}\left\{p_{i,1} + \sum_{v=1}^{n-1} [(1 - \theta)p_{i,v} + \theta\pi_{i,v}]\right\} \\ &= \frac{1}{n}\,[p_{i,1} + (n - 1)(1 - \theta)\bar{p}_i(n - 1) + \theta(n - 1)\bar{\pi}_i(n - 1)]\end{aligned}$$

where $\bar{\pi}_i\ (n - 1)$ represents the expected proportion of E_i reinforcing events over the first $n - 1$ trials. For large n, the right side of the last expression approaches the limit

$$(1 - \theta)\bar{p}_i(n - 1) + \theta\bar{\pi}_i(n - 1).$$

Further, since $\bar{p}_i(n - 1)$ always differs from $\bar{p}_i(n)$ by a term of the order of $1/n$, we can write, for sufficiently large n, the approximate equality

$$\bar{p}_i(n) \cong (1 - \theta)\bar{p}_i(n) + \theta\bar{\pi}_i(n - 1),$$

or

$$\bar{p}_i(n) \cong \bar{\pi}_i(n - 1).$$

Thus we find that no matter how π_i varies over a series of trials, the cumulative proportions of A_i and E_i occurrences tend to equality as n becomes large. It can be expected that this remarkably general "matching law" will play a central role in empirical tests of the theory.

To study the pre-asymptotic course of learning, we proceed as follows. Suppose that a subject begins an experiment with the probability $p_{i,1}$ of making an A_i; then his expected probability on trial 2 will be, applying (2),

$$p_{i,2} = (1 - \theta)p_{i,1} + \theta\pi_{i,1} \,;$$

on trial 3,

$$\begin{aligned}p_{i,3} &= (1 - \theta)p_{i,2} + \theta\pi_{i,2} \,; \\ &= (1 - \theta)^2 p_{i,1} + \theta(1 - \theta)\pi_{i,1} + \theta\pi_{i,2} \,;\end{aligned}$$

and, in general, on trial n

$$(3)\qquad p_{i,n} = (1-\theta)^{n-1}p_{i,1} + \theta[(1-\theta)^{n-2}\pi_{i,1} + (1-\theta)^{n-3}\pi_{i,2} + \cdots + (1-\theta)^{n-n}\pi_{i,n-1}]$$

$$= (1-\theta)^{n-1}p_{i,1} + \theta\sum_{v=1}^{n-1}(1-\theta)^{n-v-1}\pi_{i,v}\ .$$

A number of important features that will characterize the mean learning curve regardless of the nature of the function $\pi_{i,n}$ can be ascertained by inspection of (2) and (3). If the value of θ is zero, no learning will occur; in the remainder of the paper this case will be excluded from all derivations. If the value of θ is greater than zero then learning will occur. By rewriting (2) in the form

$$p_{i,n+1} = p_{i,n} + \theta(\pi_{i,n} - p_{i,n}),$$

we see that on the average, response probability on any trial changes in the direction of the current value of π_i . As n becomes large, the term $(1-\theta)^{n-1}p_{i,1}$ in (3) tends to zero. After n is large enough so that $(1-\theta)^{n-1}p_{i,1}$ is negligible, $p_{i,n}$ is essentially a weighted mean of the π_i values which obtained on preceding trials, with $\pi_{i,n-1}$ having most weight, $\pi_{i,n-2}$ less weight, and so on. If $\pi_{i,n}$ is some orderly function of n, as for example a straight line or a growth function, then the curve for $p_{i,n}$ tends to approach this function as n increases, but always "follows it with a lag." If rate of learning is maximal, i.e., θ is equal to one, then $p_{i,n}$ is simply equal to $\pi_{i,n-1}$ throughout the series of trials; the more θ deviates from one, the more the curve for $p_{i,n}$ lags behind that for $\pi_{i,n}$.

We may gain further insight into this learning process and at the same time develop functions that will be useful in experimental applications by considering some special cases in which $\pi_{i,n}$ can be represented by familiar functions with simple properties.

Non-Contingent Case

a. The special case of $\pi_{i,n}$ constant

If π_i is constant, then as one might expect, (2) and (3) reduce to the simple expressions

$$(2a)\qquad p_{i,n+1} = (1-\theta)p_{i,n} + \theta\pi_i\ ,$$

and

$$(3a)\qquad p_{i,n} = \pi_i - (\pi_i - p_{i,1})(1-\theta)^{n-1},$$

derived by Estes and Straughan [8] from the set-theoretical model [4, 6] and, with slightly different notation, by Bush and Mosteller [2] from their "linear operator" model. In this case the predicted learning curve is given by a negatively accelerated function tending to π_i asymptotically. Experimental applications of (3a) are described in references [2, 3, 5, 6, 13].

b. The special case of $\pi_{i,n}$ linear

We shall treat this case in some detail since it has a number of properties that will be especially convenient for experimental tests of the theory. The linear function

$$\pi_{i,n} = a_i + b_i n,$$

a_i and b_i being constants, is not in general bounded between zero and one for all n; for experimental purposes, however, one need only choose values of a_i and b_i which, for the number of trials to be given, keep the value of $\pi_{i,n}$ within the required range. Subject to this restriction, we may substitute into (2) and (3) to obtain the expected response probability on any trial,

$$\text{(2b)} \qquad p_{i,n+1} = (1 - \theta)p_{i,n} + \theta(a_i + b_i n),$$

and

$$\text{(3b)} \qquad p_{i,n} = a_i + b_i n - \frac{b_i}{\theta} - \left(a_i + b_i - \frac{b_i}{\theta} - p_{i,1}\right)(1 - \theta)^{n-1}.$$

In the interest of brevity we have omitted the detailed steps involved in summing the series in (3); the method of performing the summation in this case, and in others to be considered in following sections, is given in standard sources [12, 14]. The reader can verify that (3b) is the correct solution to (2b) by substituting the former into the latter. The main properties of (3b) are illustrated in Fig. 1. Regardless of the initial value $p_{i,1}$, after a sufficiently large number of trials the curve for $p_{i,n}$ approaches a straight line,

$$p_{i,n} = a_i - \frac{b_i}{\theta} + b_i n,$$

which has the same slope as the straight line representing $\pi_{i,n}$. If the initial value of $p_{i,n}$ is greater than $\pi_{i,1}$ and the slope of $\pi_{i,n}$ is positive, $p_{i,n}$ will decrease until its curve crosses the line $\pi_{i,n}$, following which it will increase; if b_i is small, the point of crossing will be approximately at the minimum value of $p_{i,n}$. To prove the last statement, we replace n by a continuous variable t, then set the derivative of $p_{i,t}$ with respect to t equal to zero and find that $p_{i,t}$ has as its minimum value

$$p_{i,t_m} = a_i - \frac{b_i}{\theta} + b_i t_m - \frac{b_i}{\log(1 - \theta)},$$

where

$$t_m = \frac{\log b_i - \log\log(1 - \theta)^K}{\log(1 - \theta)}$$

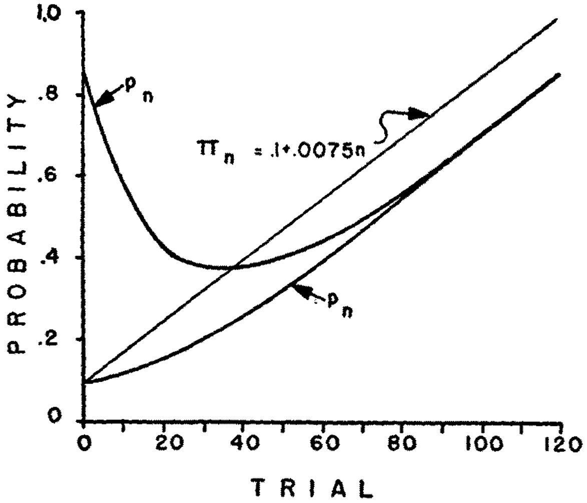

FIGURE 1

Curves describing changes in response probability when probability of reinforcement varies linearly with trials. The parameter θ has been taken equal to .05.

and

$$K = \frac{a_j - \frac{(1-\theta)b_j}{\theta} - p_{j,1}}{(1-\theta)}.$$

Subtracting π_{j,t_m} from the minimum value of p_j , we find that the difference is equal to

$$-\frac{b_j}{\theta} - \frac{b_j}{\log(1-\theta)} = b_j\left[-\frac{1}{\theta} + \frac{1}{\theta + \frac{\theta^2}{2} + \frac{\theta^3}{3} + \cdots}\right]$$

$$= -b_j\left[\frac{\frac{1}{2} + \frac{\theta}{3} + \frac{\theta^2}{4} + \cdots}{1 + \frac{\theta}{2} + \frac{\theta^2}{3} + \cdots}\right],$$

which is negative and does not exceed b_j in absolute value for any value of θ.

To obtain an expression for $R_j(n)$, the cumulative number of A$_j$ responses expected in n trials, we need only sum (3b):

$$(4) \qquad R_j(n) = \sum_{v=1}^{n} p_{j,v}$$

$$= \left[a_j - \frac{b_j}{\theta}\right]n + \frac{b_j n(n+1)}{2}$$

$$- \left[a_j + b_j - \frac{b_j}{\theta} - p_{j,1}\right]\frac{[1-(1-\theta)^n]}{\theta}.$$

Similarly, by summing (3b) over the mth block of k trials and dividing by k, we obtain the expected proportion of A_i responses in the block:

$$\begin{aligned} \bar{p}_i(k,\, m) &= a_i - \frac{b_i}{\theta} + \frac{b_i}{2}\,(2mk - k + 1) \\ &\quad - \frac{\left[a_i + b_i - \dfrac{b_i}{\theta} - p_{i,1}\right]}{k\theta}\,[1 - (1 - \theta)^k](1 - \theta)^{k(m-1)}. \end{aligned} \tag{5}$$

Equation (5), despite its cumbersome appearance, has essentially the same properties as (3b) and can readily be fitted to experimental data. For a block of k trials beginning with a value of n large enough so that $(1 - \theta)^{k(m-1)}$ is near zero, we have the approximation

$$\bar{p}_i(k,\, m) \cong a_i - \frac{b_i}{\theta} + \frac{b_i}{2}\,(2mk - k + 1).$$

By substituting the observed value of $p_i(k,\, m)$ from a set of experimental data and solving for θ, we obtain an estimate of this parameter which, although not unbiased, will be adequate for many experimental purposes.

c. The special case $\pi_{i,n} = a_i + c_i b_i^n$

Among the possible monotone relations between π_i and n, the second main type of interest is that in which $\pi_{i,n}$ approaches an asymptote. This type will be represented by the function $\pi_{i,n} = a_i + c_i b_i^n$, the values of the constants a_i, b_i, and c_i being so restricted that $\pi_{i,n}$ is properly bounded between zero and one for all n.

Equations (2) and (3) now take the forms

$$p_{i,n+1} = (1 - \theta)p_{i,n} + \theta(a_i + c_i b_i^n); \tag{2c}$$

and, if $b_i \neq 1 - \theta$,

$$p_{i,n} = a_i + \frac{\theta c_i b_i^n}{b_i - 1 + \theta} - \left(a_i + \frac{\theta c_i b_i}{b_i - 1 + \theta} - p_{i,1}\right)(1 - \theta)^{n-1};$$

or, if $b_i = 1 - \theta$,

$$p_{i,n} = a_i + c_i\theta(n - 1)(1 - \theta)^{n-1} - (a_i - p_{i,1})(1 - \theta)^{n-1}. \tag{3c}$$

Some properties of (3c) are illustrated in Fig. 2. In the upper panel, a_i has been taken equal to .50, c_i to 1.0, and b_i to .98 so that $\pi_{i,n}$ describes a negatively accelerated decreasing curve approaching .50 asymptotically. The effect of changing the sign of b_i from positive to negative can be seen by comparing the lower panel of Fig. 2, which has $b_i = -.98$, $a_i = .50$, and $c_i = 1.0$, with the upper panel. Now the values of π_i oscillate from trial to trial between a pair of curves, the upper envelope being identical with the $\pi_{i,n}$ curve in the upper panel and the lower envelope curve the mirror image of

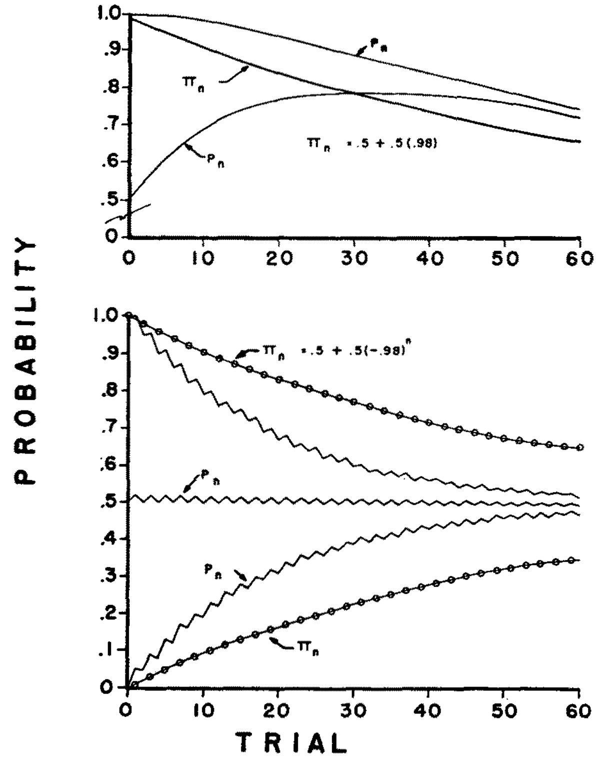

FIGURE 2

Curves describing changes in response probability when probability of reinforcement varies exponentially with trials. The parameter θ has been taken equal to .05.

it. The values of $p_{i,n}$ describe a damped oscillation around an exponential function; for any given set of parameter values, the values of $p_{i,n}$ will be alternately above and below those of the curve

$$p_{i,n} = a_i - \left(a_i + \frac{\theta c_i b_i}{b_i - 1 + \theta} - p_{i,1}\right)(1 - \theta)^{n-1},$$

with the deviation from the smooth curve decreasing progressively in magnitude toward zero as n increases.

A formula for the expected number of A_i responses in n trials can be obtained and utilized for estimation of θ as in case (c).

d. A periodic case

From an analysis of the general solution in section (a) above, we can predict that if π_i varies in accordance with a periodic function, then asymptotically the curve for $p_{i,n}$ will be described by a periodic function having the same period. A simple case with convenient properties for experimental purposes is the following: π_i is constant within any one block of k trials, but alternates between two values, say $a_i + b_i$ and $a_i - b_i$, on successive blocks so that the value of π_i on each trial of the mth block is given by

$$\pi_i = a_i + b_i(-1)^m.$$

The value of p_i at the end of the mth block can be taken directly from section (a) above:

$$p_{i,mk+1} = a_i + b_i(-1)^m - [a_i + b_i(-1)^m - p_{i,(m-1)k+1}](1 - \theta)^k.$$

Treating blocks of k trials as units, this expression may be viewed as a difference equation of the same form as (2). Substituting $a_i + b_i(-1)^m$, $(1 - \theta)^k$, and mk for the corresponding terms $\pi_{i,n}$, $(1 - \theta)$, and n of (2) and (3), we obtain the solution

$$\text{(3d)} \quad p_{i,mk+1} = a_i + b_i(-1)^m \frac{[1 - (1 - \theta)^k]}{[1 + (1 - \theta)^k]} - \left\{a_i + b_i \frac{[1 - (1 - \theta)^k]}{[1 + (1 - \theta)^k]} - p_{i,1}\right\}(1 - \theta)^{mk}.$$

Equation (3d) gives us the expected value of p_i at the end of the mth trial block. Using (3a) of section (a) again, we have for the expected value of p_i on the n'th trial of the $(m + 1)$st block

$$\text{(3e)} \quad p_{i,mk+n'} = a_i + b_i(-1)^{m+1} - [a_i + b_i(-1)^{m+1} - p_{i,mk+1}](1 - \theta)^{n'-1}.$$

Properties of this solution are illustrated in Fig. 3. It can be seen that regardless of its initial value, p_i settles down to a periodic function with period k.

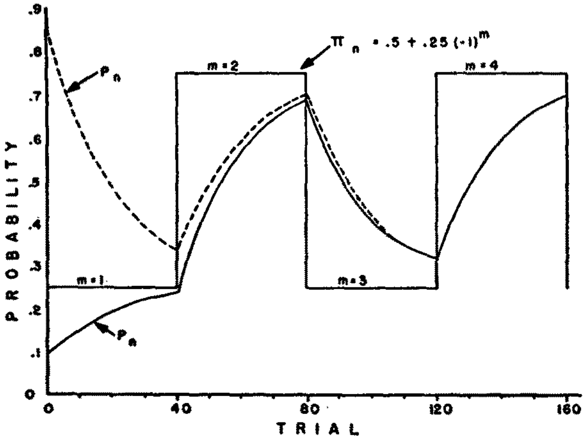

FIGURE 3

Curves describing changes in response probability when probability of reinforcement varies periodically with trials. The parameter θ has been taken equal to .05.

e. Outcome contingencies

Many cases in which the probability of a given reinforcing event on any trial depends on the outcome (reinforcing event) of some preceding trial can be reduced to cases already considered. Suppose, for example, that we set the probability of E_1 on any trial equal to π_{11} if an E_1 occurred on the vth preceding trial and to π_{21} if an E_2 occurred on the vth preceding trial. Then we can write the following difference equation for $\pi_{1,n}$, the expected probability of E_1 on trial n,

$$\begin{aligned}\pi_{1,n} &= \pi_{1,n-v}\pi_{11} + (1 - \pi_{1,n-v})\pi_{21}\\ &= (\pi_{11} - \pi_{21})\pi_{1,n-v} + \pi_{21} ,\end{aligned}$$

which has the general solution

$$\pi_{1,n} = \pi_1 + C_1r_1^n + C_2r_2^n + \cdots + C_vr_v^n .$$

The C_j are constants to be evaluated from the initial conditions of the experiment; the r_j are roots of the characteristic equation

$$r^v - \pi_{11} + \pi_{21} = 0;$$

and π_1 , the asymptotic value of $\pi_{1,n}$, is given by

$$\pi_1 = \frac{\pi_{21}}{1 - \pi_{11} + \pi_{21}}.$$

If $v = 1$, i.e., the probability of a given outcome depends on the outcome of the preceding trial, the formula for $\pi_{1,n}$ reduces to

$$\pi_{1,n} = \pi_1 - (\pi_1 - \pi_{1,1})(\pi_{11} - \pi_{21})^{n-1}.$$

Once a formula for $\pi_{j,n}$ has been deduced, it may be substituted into (2), and the machinery already developed for non-contingent cases with varying probabilities of reinforcement can be applied to generate predictions about the course of learning. In the case $v = 1$, the difference equation for $p_{1,n}$ and its solution will be given by (2c) and (3c), respectively, with $a = \pi_1$ and $b = \pi_{11} - \pi_{21}$; this case has been discussed in some detail by Bush and Mosteller [2].

It should be emphasized that functions derived from the present model for outcome contingencies with $v = 1$ will generally provide satisfactory descriptions of empirical relationships only if the experiments are conducted with well-spaced trials. According to this model, the asymptotic conditional probabilities of A_1 on trials following E_1 and E_2 occurrences, respectively, are given by

$$p_{11} = \pi_1 + \theta(1 - \pi_1)$$

and

$$p_{21} = \pi_1 - \theta\pi_1 .$$

When trials are adequately spaced, these relations may prove to be empirically confirmable, but if intertrial intervals are small enough so that the subject can form a discrimination based on the differential stimulus after-effects of E_1 and E_2 trials, then the asymptotic conditional probabilities will certainly approach π_{11} and π_{21}. A model for the massed-trial case can be derived from a set-theoretical model for discrimination learning [1, 7]. Although a detailed presentation of the discrimination model would be beyond the scope of this paper, it is interesting to note that the discrimination model yields the same asymptotic value for the over all mean value of p_1 as the present model, but yields asymptotic means for p_{11} and p_{21} which differ from π_{11} and π_{21}, respectively, only by terms which are smaller than θ.

Contingent Case

Let $\pi_{ij,n}$ represent the probability that reinforcing event E_j will occur on trial n of a series given that the subject makes response A_i on this trial, and assume that $\sum_j \pi_{ij,n} = 1$ for all i and n. Then to obtain the expected value of $p_{j,n+1}$ as a function of the value on trial n, we again average the right-hand sides of (1a) and (1b), weighting each of the possible outcomes by its probability of occurrence, viz.,

$$p_{j,n+1} = (1 - \theta)p_{j,n} + \theta \sum_i p_{i,n}\pi_{ij,n} \,. \tag{6}$$

a. General solution for the case of two response classes

If there are only two response classes, A_1 and A_2, with corresponding reinforcing events, E_1 and E_2, defined for a given situation, then we have for the expected probability of A_1 on the second trial of a series,

$$\begin{aligned} p_{1,2} &= (1 - \theta)p_{1,1} + \theta[p_{1,1}\pi_{11,1} + (1 - p_{1,1})\pi_{21,1}] \\ &= (1 - \theta + \theta\pi_{11,1} - \theta\pi_{21,1})p_{1,1} + \theta\pi_{21,1} \,, \end{aligned}$$

on the third trial

$$\begin{aligned} p_{1,3} &= (1 - \theta)p_{1,2} + \theta[p_{1,2}\pi_{11,2} + (1 - p_{1,2})\pi_{21,2}] \\ &= \alpha_2\alpha_1 p_{1,1} + \alpha_2\theta\pi_{21,1} + \theta\pi_{21,2} \,, \end{aligned}$$

when we have introduced the abbreviation

$$\alpha_v = 1 - \theta + \theta\pi_{11,v} - \theta\pi_{21,v} \,.$$

In general on the nth trial,

$$\begin{aligned} p_{1,n} &= p_{1,1}\alpha_1\alpha_2 \cdots \alpha_{n-1} + \theta\alpha_1\alpha_2 \cdots \alpha_{n-1} \sum_{u=1}^{n-1} \frac{\pi_{21,u}}{\alpha_1\alpha_2 \cdots \alpha_u} \\ &= p_{1,1} \prod_{v=1}^{n-1} \alpha_v + \theta \prod_{v=1}^{n-1} \alpha_v \sum_{u=1}^{n-1} \frac{\pi_{21,u}}{\prod_{v'=1}^{u} \alpha_{v'}} \,. \end{aligned} \tag{7}$$

Since each of the α_v is a fraction between zero and one, we can see by inspection of (7) that $p_{1,n}$ becomes independent of its initial value, $p_{1,1}$, as n becomes large; on later trials it is essentially equal to a weighted mean of the π_{21} values which obtained on preceding trials, with $\pi_{21,n-1}$ having most weight, $\pi_{21,n-2}$ less weight, and so on. [If $\pi_{11} = 1$ and $\pi_{21} = 0$, then $\alpha = 1$ and (6) reduces to

$$p_{1,n+1} = p_{1,n} ,$$

i.e., on the average no learning occurs. In all derivations presented, we shall assume this case to be excluded.] The smaller the average difference between $\pi_{11,v}$ and $\pi_{21,v}$, the more completely is the value of $p_{1,n}$ determined by the π_{ij} values of a few immediately preceding trials. As in the non-contingent case, the dependence of $p_{1,n}$ on the sequence of π_{ij} values, might be described as "tracking with a lag," but in this instance it will be necessary to study some special cases in order to see just what is being "tracked." For convenience in exposition we shall limit ourselves to situations involving two response classes while describing the special cases. In a later section we shall indicate how all of the results can be extended to situations involving more than two response classes.

b. The special case of π_{ij} constant

If $\pi_{11,n}$ and $\pi_{21,n}$ are both constant, then (6) and (7) reduce to the expressions

$$p_{j,n+1} = (1 - \theta)p_{j,n} + \theta \sum_{i=1}^{2} p_{i,n}\pi_{ij} , \tag{6a}$$

and

$$p_{1,n} = \frac{\pi_{21}}{1 - \pi_{11} + \pi_{21}} - \left(\frac{\pi_{21}}{1 - \pi_{11} + \pi_{21}} - p_{1,1}\right)(1 - \theta + \theta\pi_{11} - \theta\pi_{21})^{n-1}, \tag{7a}$$

previously derived by Estes [5] from the set-theoretical model and by Bush and Mosteller [2] from their "linear operator" model. Experimental applications of (7a) are described in references [2, 5, 13].

c. Special cases leading to linear difference equations with constant coefficients

Examination of (6) reveals that it will take the form of a linear difference equation with constant coefficients whenever $\pi_{11,n}$ and $\pi_{21,n}$ differ only by a constant. Thus, if

$$\pi_{11,n} = a_{11} + g_n$$

and

$$\pi_{21,n} = a_{21} + g_n ,$$

where g_n is any function that keeps $\pi_{ij,n}$ properly bounded for the range of n under consideration, then (7) has the form

$$\begin{aligned} p_{1,n} &= p_{1,1}\alpha^{n-1} + \theta\alpha^{n-1}\sum_{u=1}^{n-1}\frac{a_{21}+g_u}{\alpha^u} \\ &= p_{1,1}\alpha^{n-1} + \frac{a_{21}}{1-a_{11}+a_{21}}(1-\alpha^{n-1}) + \theta\alpha^{n-1}\sum_{u=1}^{n-1}\frac{g_u}{\alpha^u}, \end{aligned} \tag{7b}$$

where $\alpha = (1 - \theta + \theta a_{11} - \theta a_{21})$. For experimental purposes, it will usually be most convenient to make g_n a linear function of n, say $g_n = bn$, in which case we can perform the summation in (7b) and obtain a simple closed formula for $p_{1,n}$, viz.,

$$\begin{aligned} p_{1,n} = \frac{a_{21}+bn}{1-a_{11}+a_{21}} - \frac{b}{\theta(1-a_{11}+a_{21})^2} \\ - \left(\frac{a_{21}+b}{1-a_{11}+a_{21}} - \frac{b}{\theta(1-a_{11}+a_{21})^2} - p_{1,1}\right) \\ \cdot(1-\theta+\theta a_{11}-\theta a_{21})^{n-1}. \end{aligned} \tag{7c}$$

The properties of (7c) are very similar to those of (3b), the corresponding solution for the non-contingent case. Regardless of the initial value $p_{1,1}$, after a sufficiently large number of trials, the curve for $p_{1,n}$ approaches the straight line

$$p_{1,n} = \frac{a_{21}+bn}{1-a_{11}+a_{21}} - \frac{b}{\theta(1-a_{11}+a_{21})^2}.$$

Since θ is the only free parameter in the latter expression, its value can be estimated by fitting the straight line to data obtained from a block of trials relatively late in the learning series. It becomes apparent now, incidentally, what it is that the $p_{1,n}$ curve "tracks with a lag." The first term on the right-hand side of (7c) is simply $\pi_{21,n}/(1 - \pi_{11,n} + \pi_{21,n})$. Thus at any moment, the slope of the $p_{1,n}$ curve is such that it would approach $\pi_{21}/(1 - \pi_{11} + \pi_{21})$, the asymptote of the constant π_{ij} solution, (7a), if $\pi_{11,n}$ and $\pi_{21,n}$ were to remain constant from that moment on. Since the $\pi_{ij,n}$ do not remain constant, the subject's curve tracks the "moving asymptote" with a lag which depends inversely on θ. As in the corresponding non-contingent case, the slope of the terminal linear portion of the $p_{1,n}$ curve can be predicted in advance of an experiment since it depends only on the values of a_{11}, a_{21}, and b, which are assigned by the experimenter.

d. Contingent case with more than two response classes

The results of the preceding section can be extended without difficulty to situations involving more than two response classes. If $\pi_{ij,n} = a_{ij} + g_n$

for all i ($i = 1, 2, \cdots, r$), then for a situation involving r response classes, we obtain by application of (6) the system of r difference equations

$$
\begin{aligned}
&p_{1,n+1} = (1 - \theta + \theta a_{11})p_{1,n} + \theta a_{21}p_{2,n} + \cdots + \theta a_{r1}p_{r,n} + \theta g_n \\
(8)\qquad &p_{2,n+1} = \theta a_{12}p_{1,n} + (1 - \theta + \theta a_{22})p_{2,n} + \cdots + \theta a_{r2}p_{r,n} + \theta g_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&p_{r,n+1} = \theta a_{1r}p_{1,n} + \theta a_{2r}p_{2,n} + \cdots + (1 - \theta + \theta a_{rr})p_{r,n} + \theta g_n ,
\end{aligned}
$$

which must be solved simultaneously in order to obtain the desired formulas for $p_{j,n}$. To facilitate the solution, we define an operator $\mathbf{E}$ as follows:

$$\mathbf{E}p_{j,n} = p_{j,n+1} .$$

Then the system (8) can be rewritten in the form:

$$
\begin{aligned}
(\mathbf{E} - 1 + \theta - \theta a_{11})p_{1,n} - \theta a_{21}p_{2,n} - \cdots - \theta a_{r1}p_{r,n} &= \theta g_n \\
- \theta a_{12}p_{1,n} + (\mathbf{E} - 1 + \theta - \theta a_{22})p_{2,n} - \cdots - \theta a_{r2}p_{r,n} &= \theta g_n \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
- \theta a_{1r}p_{1,n} - \theta a_{2r}p_{2,n} - \cdots + (\mathbf{E} - 1 + \theta - \theta a_{rr})p_{r,n} &= \theta g_n .
\end{aligned}
$$

Now the symbol $\mathbf{E}$ may be treated as a number while we proceed to solve the system of equations by standard methods. The solution will express each of the $p_{j,n}$ as a polynomial in powers of $\mathbf{E}$. Then to obtain a formula expressing $p_{j,n}$ as an explicit function of n, we will have only to solve a linear difference equation with constant coefficients.

If the form of the function g_n is such that $a_{ij} + g_n$ approaches an asymptotic value, π_{ij}, as n increases, then the asymptotic values, call them λ_j, of the $p_{j,n}$ can be obtained by solving simultaneously the system of r linear equations in the r unknowns λ_j, ($j = 1, 2, \cdots, r$):

$$
\begin{aligned}
-(1 - \pi_{11})\lambda_1 + \pi_{21}\lambda_2 + \cdots + \pi_{r1}\lambda_r &= 0 \\
\pi_{12}\lambda_1 - (1 - \pi_{22})\lambda_2 + \cdots + \pi_{r2}\lambda_r &= 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\pi_{1r}\lambda_1 + \pi_{2r}\lambda_2 + \cdots - (1 - \pi_{rr})\lambda_r &= 0
\end{aligned}
$$

This system of equations has two properties of special interest, First, the asymptotic response probabilities λ_j are completely determined by the parameters π_{ij}. Second, the mean asymptotic probabilities of the reinforcing events are determined by the same system of equations. If we let π_j represent the mean asymptotic probability of E_j, then clearly

$$\pi_j = \lambda_1\pi_{1j} + \lambda_2\pi_{2j} + \cdots + \lambda_r\pi_{rj} .$$

But inspecting the jth row in the equation system above, we see that

$$\lambda_j = \lambda_1\pi_{1j} + \lambda_2\pi_{2j} + \cdots + \lambda_r\pi_{rj} .$$

Therefore, $\pi_j = \lambda_j$, i.e., asymptotically the mean probability of a response is equal to the mean probability of the corresponding reinforcing event. We have another example of the "probability matching" which has frequently been noted in studies of probability learning with simple, non-contingent reinforcement [3, 5, 8, 9, 13]. In the contingent case, there are no fixed environmental probabilities to be matched by the subject, but the matching property again obtains when the stimulus-response system arrives at a state of statistical equilibrium.

In the special case when $g_n = 0$ for all n and $a_{ij} = \pi_{ij}$, the value of $p_{j,n}$ will be given by an expression of the form

$$p_{j,n} = \lambda_j + C_1x_1^n + C_2x_2^n + \cdots + C_{r-1}x_{r-1}^n , \tag{9}$$

where the absolute value of each of the x_i is in the range $0 \leqq x_i \leqq 1$, and the C_i are constants whose values depend on the initial p_j values and on the π_{ij} . It may be noted that all of the C_i need not have the same sign, and consequently the curve of $p_{j,n}$ will not always be a monotone function of n. Some of the curve forms which arise are illustrated in Fig. 4; the curves in the upper and lower panels represent the same value of θ but different combinations of π_{ij} , viz.,

	Upper panel			*Lower panel*		
	E_1	E_2	E_3	E_1	E_2	E_3
A_1	.33	.33	.33	.33	.33	.33
A_2	.50	.50	.00	.50	.50	.00
A_3	.17	.00	.83	.83	.00	.17

It will be apparent from inspection of Fig. 4 that in this case, unlike the non-contingent case, not only the asymptotes of the learning curves but also the relative rates at which the curves approach their asymptotes depend upon the probabilities of reinforcement.

e. *Contingency with a lag*

The contingent cases discussed above cover the common types of experiments in which the probabilities of such reinforcing events as rewards or knowledge of results on any trial depend on the subject's response on that trial. Now we wish to extend the theory to include the more remote contingencies which arise in games or similar two-person situations. In this type of situation it is a common strategy to make one's choice of moves, or plays, on a given trial depend upon the choices made by one's adversary on preceding trials. Regarding the first player as the experimenter and the

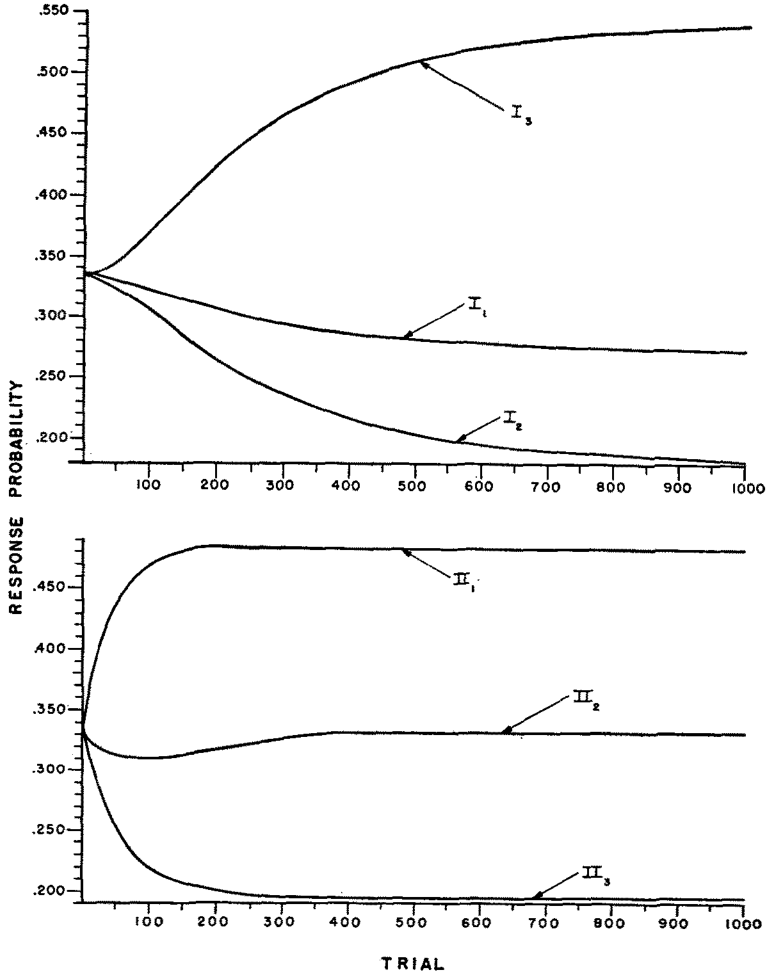

FIGURE 4

Curves describing changes in response probability under simple, contingent reinforcement. Probabilities of reinforcement following responses A_1 ane A_2 are the same under Schedules I and II, but probabilities following A_3 differ. The parameter θ has been taken equal to .015.

second as the subject, we can represent this kind of strategy in the present model by letting the probability of reinforcement of a given response on trial n depend upon the subject's response on some preceding trial, say $n - v$. By the same reasoning used in the case of (2) and (6), we can write a difference equation for mean probability of response A_j on any given trial:

$$p_{i,n+1} = (1 - \underline{\theta)p_{i,n}} + \theta \sum_{i} p_{i,n-v}\pi^{(v)}_{ij,n} , \tag{10}$$

where $\pi_{ij,n}^{(v)}$ represents the probability of reinforcement of A_j on trial $n + 1$ given that A_i occurred on trial $n - v$. (10) is difficult to handle unless the functions $\pi_{ij,n}^{(v)}$ differ only by constant terms [i.e., $\pi_{11} = a_{11} + g_n(v)$, $\pi_{21} = a_{21} + g_n(v)$, etc.]; for this case, (10) reduces to a linear difference equation with constant coefficients

$$\text{(10a)} \qquad p_{j,n+1} = (1 - \theta)p_{j,n} + \theta \sum_i p_{i,n-v}a_{ij} + \theta g_n^{(v)} ,$$

which can be solved explicitly. In order to exhibit some of the most readily testable implications of this model for experiments involving remote contingencies, let us consider the special case of two response classes and $\pi_{ij,n}^{(v)}$ independent of n. Then for a given contingency lag v, the $\pi_{ij,n}^{(v)}$ can be treated as constants, and (10) reduces to

$$\text{(10b)} \qquad \begin{aligned} p_{1,n+1} &= (1 - \theta)p_{1,n} + \theta[p_{1,n-v}\pi_{11} + (1 - p_{1,n-v})\pi_{21}] \\ &= (1 - \theta)p_{1,n} + \theta(\pi_{11} - \pi_{21})p_{1,n-v} + \theta\pi_{21} . \end{aligned}$$

Now [excluding, as before, the case ($\pi_{11} = 1$ and $\pi_{21} = 0$) for which $p_{1,\infty} = p_{1,1}$] we can obtain the asymptotic probability of response A_1 by setting $p_{1,n+1} = p_{1,n} = p_{1,n-v} = p_{1,\infty}$ in (10b) and solving, viz.,

$$p_{1,\infty} = \frac{\pi_{21}}{1 - \pi_{11} + \pi_{21}}.$$

We obtain the interesting prediction that asymptotic probability is independent of the contingency lag v. The complete solution of (10b) is (cf. [12] for the detailed method of derivation and for the treatment of cases in which the characteristic roots are not all distinct)

$$\text{(10c)} \qquad p_{1,n} = C_1x_1^n + C_2x_2^n + \cdots + C_{v+1}x_{v+1}^n + \frac{\pi_{21}}{1 - \pi_{11} + \pi_{21}} ,$$

where the C_i are constants which can be evaluated from the initial conditions of the experiment and the x_i are the roots of the characteristic equation

$$x^{v+1} - (1 - \theta)x^v - \theta(\pi_{11} - \pi_{21}) = 0.$$

Except for the degenerate case ($\pi_{11} = 1$ and $\pi_{21} = 0$), the characteristic roots will have absolute values in the range $0 \leqq x < 1$, and therefore x^n will tend to zero as n increases. If the lag v is zero, then the characteristic equation is simply

$$x - (1 - \theta) - \theta(\pi_{11} - \pi_{21}) = 0$$

which has the single root

$$x = 1 - \theta + \theta\pi_{11} - \theta\pi_{21} ,$$

and (10c) reduces to (7a) as it should.

If the lag v is 1, i.e., probability of reinforcement on a given trial depends on the response of the immediately preceding trial, the characteristic equation is

$$x^2 - (1 - \theta)x - \theta(\pi_{11} - \pi_{21}) = 0,$$

which has the two roots

$$x_1 = \frac{1 - \theta + \sqrt{(1 - \theta)^2 + 4\theta(\pi_{11} - \pi_{21})}}{2}$$

and

$$x_2 = \frac{1 - \theta - \sqrt{(1 - \theta)^2 + 4\theta(\pi_{11} - \pi_{21})}}{2}.$$

The properties of the solution will depend on the relative magnitudes of π_{11} and π_{21} as follows:

1. If $\pi_{11} = \pi_{21}$, then x_1 and x_2 are equal to $1 - \theta$ and 0, respectively, and (10c) reduces to the (3a) of the simple non-contingent case.

2. If $\pi_{11} > \pi_{21}$, then x_1 and x_2 are real numbers, positive and negative, respectively, with absolute values between 0 and 1. Comparing the larger root, x_1, with the characteristic root for the case of lag 0, we find that the difference between the former and the latter is always non-negative when $\pi_{11} > \pi_{21}$; i.e.,

$$\frac{1 - \theta + \sqrt{(1 - \theta)^2 + 4\theta(\pi_{11} - \pi_{21})}}{2} - (1 - \theta + \theta\pi_{11} - \theta\pi_{21}) \geqq 0,$$

and the equality holds only in the degenerate cases ($\theta = 0$; $\pi_{11} = 1$ and $\pi_{21} = 0$) for which (10c) is inapplicable. Thus it can be predicted that when $\pi_{11} > \pi_{21}$, the mean learning curve will approach its asymptote more slowly for the case of lag 1 than for the case of lag 0.

3. If $\pi_{11} < \pi_{21}$, then neither x_1 nor x_2 is negative. Both x_1 and x_2 are real numbers in the interval $0 < x < (1 - \theta)$ if the quantity

$$[(1 - \theta)^2 + 4\theta(\pi_{11} - \pi_{21})]$$

is positive; otherwise they are complex numbers with moduli in the interval $0 < | x | < 1$.

In general the estimation of parameters from data will be difficult when there is a contingency lag. Tests of this aspect of the theory can be achieved most conveniently by obtaining estimates of θ from data obtained under conditions of simple non-contingent or contingent reinforcement and then computing predicted relationships for experiments run under similar conditions except for the introduction of contingency lags. Predictions about asymptotic probabilities are, of course, independent of θ and thus can be made in advance of any experiment.

Interpretation of the Model

The theory of reinforcement developed here might be characterized as descriptive, rather than explanatory. The concept of reinforcing event represents an abstraction from a considerable body of experimental data on conditioning and simple motor and verbal learning. In a number of standard experimental situations used to study these elementary forms of learning, it is possible to identify experimentally defined events or operations whose effects upon response probability appear to satisfy the quantitative laws expressed by (1a) and (1b). The first task of our quantitative theory is simply to describe how learning should proceed under various experimental arrangements when these particular experimental operations are assigned the role of reinforcing events. A second task, which becomes important once the theory has survived preliminary tests, is to facilitate the identification of reinforcing operations in new empirical situations. We can test hypotheses concerning a class of events termed reinforcers only if we can state detailed testable consequences of class membership. To the extent that the model elaborated here acquires standing as a descriptive theory, it will serve also to specify the quantitative properties which define membership in the class of reinforcers. Although a quantitative theory of this kind does not contribute immediately to an intensive definition, or interpretive account, of reinforcement, it does provide an additional research tool which may contribute to the construction and testing of explanatory theories.

REFERENCES

[1] Burke, C. J. and Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika,* 1957, **22**, 133-145.

[2] Bush, R. R. and Mosteller, F. Stochastic models for learning. New York: Wiley, 1955.

[3] Detambel, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.,* 1955, **49**, 97-104.

[4] Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.,* 1950, **57**, 94-107.

[5] Estes, W. K. Individual behavior in uncertain situations. In R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.), Decision processes. New York: Wiley, 1954, pp. 127-137.

[6] Estes, W. K. and Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.,* 1953, **60**, 276-286.

[7] Estes, W. K. and Burke, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.,* 1955, **50**, 81-88.

[8] Estes, W. K. and Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.,* 1954, **47**, 225-234.

[9] Grant, D. A., Hake, H. W., and Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1-5.

[10] Hake, H. W. and Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64-74.

[11] Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294-301.

[12] Jordan, C. Calculus of finite differences. New York: Chelsea, 1950.

[13] Neimark, E. D. Effects of type of non-reinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209-220.

[14] Richardson, C. H. An introduction to the calculus of finite differences. New York: Van Nostrand, 1954.

Manuscript received 8/27/56