

# TOWARD A STATISTICAL THEORY OF LEARNING \*

BY WILLIAM K. ESTES

*Indiana University*

Improved experimental techniques for the study of conditioning and simple discrimination learning enable the present day investigator to obtain data which are sufficiently orderly and reproducible to support exact quantitative predictions of behavior. Analogy with other sciences suggests that full utilization of these techniques in the analysis of learning processes will depend to some extent upon a comparable refinement of theoretical concepts and methods. The necessary interplay between theory and experiment has been hindered, however, by the fact that none of the many current theories of learning commands general agreement among researchers. It seems likely that progress toward a common frame of reference will be slow so long as most theories are built around verbally defined hypothetical constructs which are not susceptible to unequivocal verification. While awaiting resolution of the many apparent disparities among competing theories, it may be advantageous to systematize well established empirical relationships at a peripheral, statistical level of analysis. The possibility of agreement on a theoretical framework, at least in certain intensively studied areas, may be maximized by defining concepts in terms of experimentally manipulable variables, and developing the consequences of assumptions by strict mathematical reasoning.

This essay will introduce a series of

\* For continual reinforcement of his efforts at theory construction, as well as for many specific criticisms and suggestions, the writer is indebted to his colleagues at Indiana University, especially Cletus J. Burke, Douglas G. Ellson, Norman Guttman, and William S. Verplanck.

studies developing a statistical theory of elementary learning processes. From the definitions and assumptions which appear necessary for this kind of formulation, we shall attempt to derive relations among commonly used measures of behavior and quantitative expressions describing various simple learning phenomena.

## PRELIMINARY CONSIDERATIONS

Since propositions concerning psychological events are verifiable only to the extent that they are reducible to predictions of behavior under specified environmental conditions, it appears likely that greatest economy and consistency in theoretical structure will result from the statement of all fundamental laws in the form

$$R = f(S),$$

where  $R$  and  $S$  represent behavioral and environmental variables respectively. Response-inferred laws, as for example those of differential psychology, should be derivable from relationships of this form. The reasoning underlying this position has been developed in a recent paper by Spence (8). Although developed within this general framework, the present formulation departs to some extent from traditional definitions of  $S$  and  $R$  variables.

Many apparent differences among contemporary learning theories seem to be due in part to an oversimplified definition of stimulus and response. The view of stimulus and response as elementary, reproducible units has always had considerable appeal because of its simplicity. This simplicity is deceptive, however, since it entails the postulation of various hypothetical processes to ac-

count for observed variability in behavior. In the present formulation, we shall follow the alternative approach of including the notion of variability in the definitions of stimulus and response, and investigating the theoretical consequences of these definitions.

It will also be necessary to modify the traditional practice of stating laws of learning in terms of relations between isolated stimuli and responses. Attempts at a quantitative description of learning and extinction of operant behavior have led the writer to believe that a self-consistent theory based upon the classical S-R model may be difficult, if not impossible, to extend over any very wide range of learning phenomena without the continual addition of *ad hoc* hypotheses to handle every new situation. A recurrent difficulty might be described as follows. In most formulations of simple learning, the organism is said originally to "do nothing" in the presence of some stimulus; during learning, the organism comes to make some predesignated response in the presence of the stimulus; then during extinction, the response gradually gives way to a state of "not responding" again. But this type of formulation does not define a closed or conservative system in any sense. In order to derive properties of conditioning and extinction from the same set of general laws, it is necessary to assign specific properties to the state of not responding which is the alternative to occurrence of the designated response. One solution is to assign properties as needed by special hypotheses, as has been done, for example, in the Pavlovian conception of inhibition. In the interest of simplicity of theoretical structure, we shall avoid this procedure so far as possible.

The role of competing reactions has been emphasized by some writers, but usually neglected in formal theorizing.

The point of view to be developed here will adopt as a standard conceptual model a closed system of behavioral and environmental variables. In any specific behavior-system, the environmental component may include either the entire population of stimuli available in the situation or some specified portion of that population. The behavioral component will consist in mutually exclusive classes of responses, defined in terms of objective criteria; these classes will be exhaustive in the sense that they will include all behaviors which may be evoked by that stimulus situation. Given the initial probabilities of the various responses available to an organism in a given situation, we shall expect the laws of the theory to enable predictions of changes in those probabilities as a function of changes in values of independent variables.

#### DEFINITIONS AND ASSUMPTIONS

1. *R-variables*. It will be assumed that any movement or sequence of movements may be analyzed out of an organism's repertory of behavior and treated as a "response," various properties of which can be treated as dependent variables subject to all the laws of the theory. (Hereafter we shall abbreviate the word response as *R*, with appropriate subscripts where necessary.) In order to avoid a common source of confusion, it will be necessary to make a clear distinction between the terms *R*-class and *R*-occurrence.

The term *R*-class will always refer to a class of behaviors which produce environmental effects within a specified range of values. This definition is not without objection (*cf.* 4) but has the advantage of following the actual practice of most experimenters. It may be possible eventually to coordinate *R*-classes defined in terms of environmental effects with *R*-classes defined in terms of effector activities.

By *R*-occurrence we shall mean a particular, unrepeatable, behavioral event. All occurrences which meet the defining criteria of an *R*-class are counted as instances of that class, and as such are experimentally interchangeable. In fact, various instances of an *R*-class are ordinarily indistinguishable in the record of an experiment even though they may actually vary with respect to properties which are not picked up by the recording mechanism.

Indices of tendency to respond, *e.g.*, probability as defined below, always refer to *R*-classes.

These distinctions may be clarified by an illustration. In the Skinner-type conditioning apparatus, bar-pressing is usually treated as an *R*-class. Any movement of the organism which results in sufficient depression of the bar to actuate the recording mechanism is counted as an instance of the class. The *R*-class may be subdivided into finer classes by the same kind of criteria. We could, if desired, treat depression of a bar by the rat's right forepaw and depression of the bar by the left forepaw as instances of two different classes provided that we have a recording mechanism which will be affected differently by the two kinds of movements and mediate different relations to stimulus input (as for example the presentation of discriminative stimuli or reinforcing stimuli). If probability is increased by reinforcement, then reinforcement of a right-forepaw-bar-depression will increase the probability that instances of that subclass will occur, and will also increase the probability that instances of the broader class, bar-pressing, will occur.

2. *S*-variables. For analytic purposes it is assumed that all behavior is conditional upon appropriate stimulation. It is not implied, however, that responses can be predicted only when eliciting stimuli can be identified. Ac-

ording to the present point of view, laws of learning enable predictions of changes in probability of response as a function of time under given environmental conditions.

A stimulus, or stimulating situation, will be regarded as a finite population of relatively small, independent, environmental events, of which only a sample is effective at any given time. In the following sections we shall designate the total number of elements associated with a given source of stimulation as *S* (with appropriate subscripts where more than one source of stimulation must be considered in an experiment), and the number of elements effective at any given time as *s*. It is assumed that when experimental conditions involve the repeated stimulation of an organism by the "same stimulus," that is by successive samples of elements from an *S*-population, each sample may be treated as an independent random sample from *S*. It is to be expected that sample size will fluctuate somewhat from one moment to the next, in which case *s* will be treated as the average number of elements per sample over a given period.

In applying the theory, any portion of the environment to which the organism is exposed under uniform conditions may be considered an *S*-population. The number of different *S*'s said to be present in a situation will depend upon the number of independent experimental operations, and the degree of specificity with which predictions of behavior are to be made. If the experimenter attempts to hold the stimulating situation constant during the course of an experiment, then the entire situation will be treated as a single *S*. If in a conditioning experiment, a light and shock are to be independently manipulated as the *CS* and *US*, then each of these sources of stimulation will be

treated as a separate  $S$ -population, and so on.

It should be emphasized that the division of environment and behavior into elements is merely an analytic device adopted to enable the application of the finite-frequency theory of probability to behavioral phenomena. In applying the theory to learning experiments we shall expect to evaluate the ratio  $s/S$  for any specific situation from experimental evidence, but for the present at least no operational meaning can be given to a numerical value for either  $S$  or  $s$  taken separately.

3. *Probability of response.* Probability will be operationally defined as the average frequency of occurrence of instances of an  $R$ -class relative to the maximum possible frequency, under a specified set of experimental conditions, over a period of time during which the conditions remain constant. In accordance with customary usage the term probability, although defined as a relative frequency, will also be used to express the likelihood that a response will occur at a given time.

4. *Conditional relation.* This relation may obtain between an  $R$ -class and any number of the elements in an  $S$ -population, and has the following implications.

(a) If a set of  $x$  elements from an  $S$  are conditioned to (*i.e.*, have the conditional relation to) some  $R$ -class,  $R_1$ , at a given time, the probability that the next response to occur will be an instance of  $R_1$  is  $x/S$ .

(b) If at a given time in an  $S$ -population,  $x_1$  elements are conditioned to some  $R$ -class,  $R_1$ , and  $x_2$  elements are conditioned to another class,  $R_2$ , then  $x_1$  and  $x_2$  have no common elements.

(c) If all behaviors which may be evoked from an organism in a given situation have been categorized into mutually exclusive classes, then the probabilities attaching to the various classes must sum to unity at all times.

We consider the organism to be always "doing something." If any arbitrarily defined class of activities may be selected as the dependent variable of a given experiment, it follows that the activity of the organism at any time must be considered as subject to the same laws as the class under consideration. Any increase in probability of one  $R$ -class during learning will, then, necessarily involve the reduction in probability of other classes; similarly, while the probability of one  $R$  decreases during extinction, the probabilities of others must increase. In other words, learning and unlearning will be considered as transfers of probability relations between  $R$ -classes.

5. *Conditioning.* It is assumed that on each occurrence of a response,  $R_1$ , all new elements (*i.e.*, elements not already conditioned to  $R_1$ ) in the momentarily effective sample of stimulus elements,  $s$ , become conditioned to  $R_1$ .

An important implication of these definitions is that the conditioning of a stimulus element to one  $R$  automatically involves the breaking of any pre-existing conditional relations with other  $R$ 's.

6. *Motivation.* Experimental operations which in the usual terminology are said to produce motives (*e.g.*, food-deprivation) may affect either the composition of an  $S$  or the magnitude of the  $s/S$  ratio. Detailed discussion of these relations is beyond the scope of the present paper. In all derivations presented here we shall assume motivating conditions constant throughout an experiment.

7. *Reinforcement.* This term will be applied to any experimental condition which ensures that successive occurrences of a given  $R$  will each be contiguous with a new random sample of elements from some specified  $S$ -population. Various ways of realizing this definition experimentally will be discussed in the following sections.

## SIMPLE CONDITIONING: REINFORCEMENT BY CONTROLLED ELICITATION

Let us consider first the simplest type of conditioning experiment. The system to be described consists of a subpopulation of stimulus elements,  $S_o$ , which may be manipulated independently of the remainder of the situation,  $S$ , and a class,  $R$ , of behaviors defined by certain measurable properties. By means of a controlled original stimulus, that is, one which has initially a high probability of evoking  $R$ , it is ensured that an instance of  $R$  will occur on every trial contiguously with the sample of stimulus elements which is present. In the familiar buzz-shock conditioning experiment, for example,  $S_o$  would represent the population of stimulus elements emanating from the sound source and  $R$  would include all movements of a limb meeting certain specifications of direction and amplitude; typically, the  $R$  to be conditioned is a flexion response which may be evoked on each training trial by administration of an electric shock.

Designating the mean number of elements from  $S_o$  effective on any one trial as  $s_o$ , and the number of elements from  $S_o$  which are conditioned to  $R$  at any time as  $x$ , the expected number of new elements conditioned on any trial will be

$$\Delta x = s_o \frac{(S_o - x)}{S_o}. \quad (1)$$

If the change in  $x$  per trial is relatively small, and the process is assumed continuous, the right hand portion of (1) may be taken as the average rate of change of  $x$  with respect to number of trials,  $T$ , at any moment, giving

$$\frac{dx}{dT} = s_o \frac{(S_o - x)}{S_o}. \quad (2)$$

This differential equation may be in-

tegrated to yield

$$x = S_o - (S_o - x_0)e^{-qT}, \quad (3)$$

where  $x_0$  is the initial value of  $x$ , and  $q$  represents the ratio  $s_o/S_o$ . Thus  $x$  will increase from its initial value to approach the limiting value,  $S_o$ , in a negatively accelerated curve. A method of evaluating  $x$  in these equations from empirical measures of response latency, or reaction time, will be developed in a later section.

If the remainder of the situation has been experimentally neutralized, the probability of  $R$  in the presence of a sample from  $S_o$  will be given by the ratio  $x/S_o$ . Representing this ratio by the single letter  $p$ , and making appropriate substitutions in (3), we have the following expression for probability of  $R$  as a function of the number of reinforced trials.

$$p = 1 - (1 - p_0)e^{-qT}. \quad (3')$$

Since we have not assumed any special properties for the original (or unconditioned) stimulus other than that of regularly evoking the response to be conditioned, it is to be expected that the equations developed in this section will describe the accumulation of conditional relations in other situations than classical conditioning, provided that other experimental operations function to ensure that the response to be learned will occur in the presence of every sample drawn from the  $S$ -population.

## OPERANT CONDITIONING: REINFORCEMENT BY CONTINGENT STIMULATION

In the more common type of experimental arrangement, various termed operant, instrumental, trial and error, etc. by different investigators, the response to be learned is not elicited by a controlled original stimulus, but has some initial strength in the experimental

situation and occurs originally as part of so-called "random activity." Here the response cannot be evoked concurrently with the presentation of each new stimulus sample, but some of the same effects can be secured by making changes in the stimulating situation contingent upon occurrences of the response. Let us consider a situation of this sort, assuming that the activities of the organism have been catalogued and classified into two categories, all movement sequences characterized by a certain set of properties being assigned to class  $R$  and all others to the class  $R_0$ , and that members of class  $R$  are to be learned.

If changes in the stimulus sample are independent of the organism's behavior, we should expect instances of the two response classes to occur, on the average, at rates proportional to their initial probabilities. For if  $x$  elements from the  $S$ -population are originally conditioned to  $R$ , then the probability of  $R$  will be  $x/S$ ; the number of new elements conditioned to  $R$  if an instance occurs will be  $s[(S-x)/S]$ ,  $s$  again representing the number of stimulus elements in a sample; and the mathematically expected increase in  $x$  will be the product of these quantities,  $sx[(S-x)/S^2]$ . At the same time, the probability of  $R_0$  will be  $(S-x)/S$ , and the number of new elements conditioned to  $R_0$  if an instance occurs will be  $sx/S$ ; multiplying these quantities, we have  $sx[(S-x)/S^2]$  as the mathematically expected decrease in  $x$ . Thus we should predict no average change in  $x$  under these conditions.

In the acquisition phase of a learning experiment two important restrictions imposed by the experimenter tend to force a correlation between changes in the stimulus sample and occurrences of  $R$ . The organism is usually introduced into the experimental situation at the beginning of a trial, and the

trial lasts until the pre-designated response,  $R$ , occurs. For example, in a common discrimination apparatus the animal is placed on a jumping stand at the beginning of each trial and the trial continues until the animal leaves the stand; a trial in a runway experiment lasts until the animal reaches the end box, and so on. Typically the stimulating situation present at the beginning of a trial is radically changed, if not completely terminated, by the occurrence of the response in question; and a new trial begins under the same conditions, except for sampling variations, after some pre-designated interval. The pattern of movement-produced stimuli present during a trial may be changed after occurrences of  $R$  by the evocation of some uniform bit of behavior such as eating or drinking; in some cases the behavior utilized for this purpose must be established by special training prior to a learning experiment. In the Skinner box, for example, the animal is trained to respond to the sound of the magazine by approaching it and eating or drinking. Then when operation of the magazine follows the occurrence of a bar-pressing response during conditioning of the latter, the animal's response to the magazine will remove it from the stimuli in the vicinity of the bar and ensure that for an interval of time thereafter the animal will not be exposed to most of the  $S$ -population; therefore the sample of elements to which the animal will next respond may be considered very nearly a new random sample from  $S$ .

In the simplest operant conditioning experiments it may be possible to change almost the entire stimulus sample after each occurrence of  $R$  (complete reinforcement), while in other cases the sampling of only some restricted portion of the  $S$ -population is correlated with  $R$  (partial reinforcement). We shall consider the former

case in some detail in the remainder of this section.

By our definition of the conditional relation, we shall expect all  $R$ -classes from which instances actually occur on any trial to be conditioned to stimulus elements present on that trial. The first movement to occur will be conditioned to the environmental cues present at the beginning of the trial; the next movement will be conditioned to some external cues, if the situation is not completely constant during a trial, and to proprioceptive cues from the first movement, and so on, until the predesignated response,  $R$ , occurs and terminates the trial. If complete constancy of the stimulating situation could be maintained, the most probable course of events on the next trial would be the recurrence of the same sequence of movements. In practice, however, the sample of effective stimulus elements will change somewhat in composition, and some responses which occur on one trial may fail to occur on the next. The only response which may never be omitted is  $R$ , since the trial continues until  $R$  occurs. This argument has been developed in greater detail by Guthrie (4). In order to verify the line of reasoning involved, we need now to set these ideas down in mathematical form and investigate the possibility of deriving functions which will describe empirical curves of learning.

Since each trial lasts until  $R$  occurs, we need an expression for the probable duration of a trial in terms of the strength of  $R$ . Suppose that we have categorized all movement sequences which are to be counted as "responses" in a given situation, and that the minimum time needed for completion of a response-occurrence is, on the average,  $h$ . For convenience in the following development, we shall assume that the mean duration of instances of class  $R$  is ap-

proximately equal to that of class  $R_0$ . Let the total number of stimulus elements available in the experimental situation be represented by  $S$ , the sample effective on any one trial by  $s$ , and the ratio  $s/S$  by  $q$ . The probability,  $p$ , of class  $R$  at the beginning of any trial will have the value  $x/S$ ; if this value varies little within a trial, we can readily compute the probable number of responses (of all classes) that will occur before the trial is terminated. The probability that an instance of  $R$  will be the first response to occur on the trial in question is  $p$ ; the probability that it will be the second is  $p(1-p)$ ; the probability that it will be the third is  $p(1-p)^2$ ; etc. If we imagine an indefinitely large number of trials run under identical conditions, and represent the number of response occurrences on any trial by  $n$ , we may weight each possible value of  $n$  by its probability (*i.e.*, expected relative frequency) and obtain a mean expected value of  $n$ . In symbolic notation we have

$$\bar{n} = \sum np(1-p)^{n-1} = p \sum n(1-p)^{n-1}.$$

The expression inside the summation sign will be recognized as the general term of a well-known infinite series with the sum  $1/(1-(1-p))^2$ . Then we have, by substitution,

$$\bar{n} = p/(1-(1-p))^2 = 1/p.$$

Then  $\bar{L}$ , the average time per trial, will be the product of the expected number of responses and the mean time per response.

$$\bar{L} = \bar{n}h = h/p = Sh/x.$$

Since  $R$  will be conditioned to all new stimulus elements present on each trial, we may substitute for  $x$  its equivalent from equation (3), dropping the sub-

scripts from  $S_0$  and  $s_0$ , and obtaining

$$\begin{aligned} L &= \frac{Sh}{S - (S - x_0)e^{-qT}} \\ &= \frac{h}{1 - \frac{(\bar{L}_0 - h)e^{-qT}}{\bar{L}_0}} \quad (4) \end{aligned}$$

Thus,  $\bar{L}$  will decline from an initial value of  $\bar{L}_0$  (equal to  $Sh/x_0$ ) and approach the asymptotic minimum value  $h$  over a series of trials.

A preliminary test of the validity of this development may be obtained by applying equation (4) to learning data from a runway experiment in which the conditions assumed in the derivation are realized to a fair degree of approximation. In Fig. 1 we have plotted acquisition data reported by Graham and Gagné (3). Each empirical point represents the geometric mean latency for a group of 21 rats which were reinforced with food for traversing a simple elevated runway. The theoretical curve

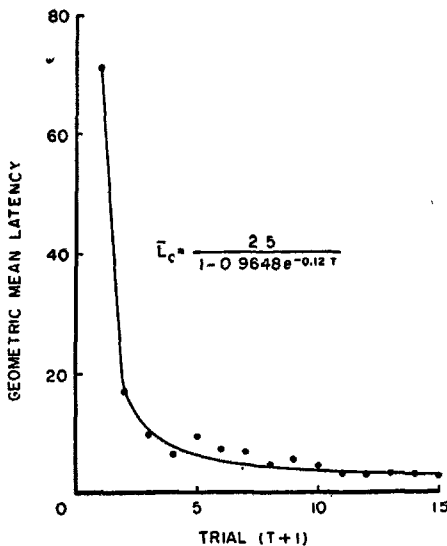


FIG. 1. Latencies of a runway response during conditioning, obtained from published data of Graham and Gagné (3), are fitted by a theoretical curve derived in the text.

in the figure represents the equation

$$\bar{L} = \frac{2.5}{1 - .9648e^{-.12T}},$$

where values of  $\bar{L}_0$ ,  $h$ , and  $q$  have been estimated from the data. This curve appears to give a satisfactory graduation of the obtained points and, it might be noted, is very similar in form to the theoretical acquisition curve developed by Graham and Gagné. The present formulation differs from theirs chiefly in including the time of the first response as an integral part of the learning process. The quantitative description of extinction in this situation will be presented in a forthcoming paper.

In order to apply the present theory to experimental situations such as the Skinner box, in which the learning period is not divided into discrete trials, we shall have to assume that the intervals between reinforcements in those situations may be treated as "trials" for analytical purposes. Making this assumption, we may derive an expression for rate of change of conditioned response strength as a function of time in the experimental situation, during a period in which all responses of class  $R$  are reinforced.

$\bar{L}$ , as defined above, will represent the time between any two occurrences of  $R$ . Then if we let  $t$  represent time elapsed from the beginning of the learning period to a given occurrence of  $R$ , and  $T$  the number of occurrences (and therefore reinforcements) of  $R$ , we have from the preceding development

$$\bar{L} = Sh/x.$$

Since  $\bar{L}$  may be considered as the increment in time during a trial, we can write the identity

$$\frac{\Delta x}{\Delta t} = \frac{\Delta x}{\Delta T} \cdot \frac{\Delta T}{\Delta t}.$$

Substituting for  $\Delta x/\Delta T$  its equivalent from (1), without subscripts, and for



$\Delta T/\Delta t$  its equivalent from the preceding equation, we have

$$\frac{\Delta x}{\Delta t} = \frac{s(S-x)}{S} \cdot \frac{x}{hS} = \frac{s(S-x)x}{hS^2}. \quad (5)$$

If the change in  $x$  per reinforcement is small and the process is assumed continuous, the right hand portion of equation (5) may be taken as the value of the derivative  $dx/dt$  and integrated with respect to time—

$$x = \frac{S}{1 + \frac{(S-x_0)}{x_0}e^{-Bt}}, \quad (6)$$

where  $B = s/Sh$ . In general, this equation defines a logistic curve with the amount of initial acceleration depending upon the value of  $x_0$ . Curves of probability ( $x/S$ ) vs. time for  $S = 100$ ,  $B = 0.25$ , and several different values of  $x_0$  are illustrated in Fig. 2.

Since we are considering a situation in which a reinforcement is administered (or a new "trial" is begun) after each occurrence of  $R$ , we are now in a position to express the expected rate of occurrence of  $R$  as a function of time. Representing rate of occurrence of  $R$  by  $r = dR/dt$ , and the ratio  $1/h$  by  $w$ , we have

$$r = \frac{dR}{dt} = \frac{dT}{dt} = \frac{wx}{S} = \frac{w}{1 + \frac{(S-x_0)}{x_0}e^{-Bt}}$$

and if we take the rate of  $R$  at the beginning of the experimental period as  $r_0 = wx_0/S$  this relation becomes

$$r = \frac{w}{1 + \frac{(w-r_0)}{r_0}e^{-Bt}}. \quad (7)$$

To illustrate this function, we have plotted in Fig. 3 measures of rate of

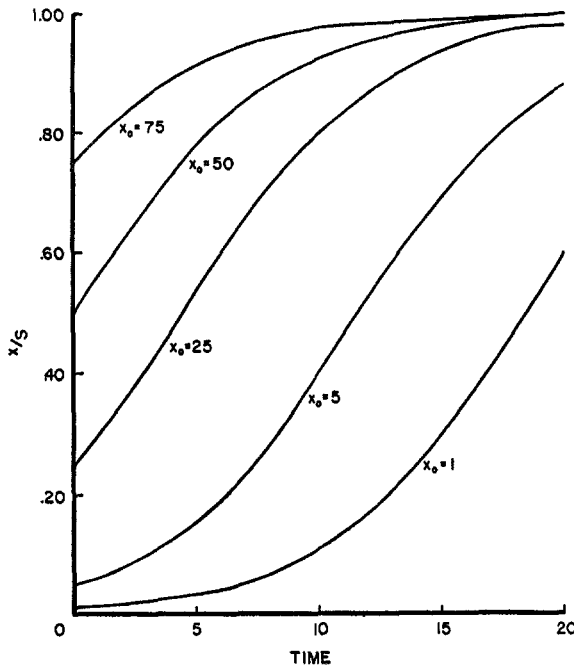


FIG. 2. Illustrative curves of probability vs. time during conditioning; parameters of the curves are the same except for the initial  $x$ -values.

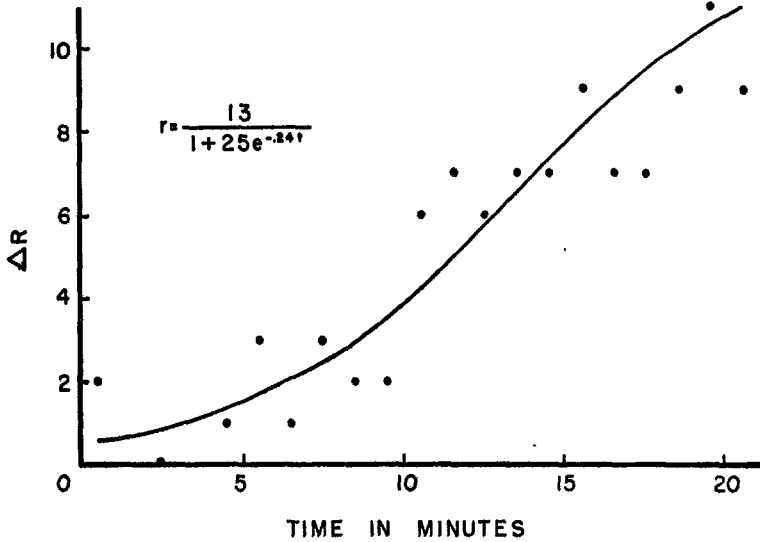


FIG. 3. Number of responses per minute during conditioning of a bar-pressing habit in a single rat; the theoretical curve is derived in the text.

responding during conditioning of a bar-pressing response by a single rat. The apparatus was a Skinner box; motivation was 24 hours thirst; the animal had previously been trained to drink out of the magazine, and during the period illustrated was reinforced with water for all bar-pressing responses. Measures of rate at various times were obtained by counting the number of responses made during the half-minute before and the half-minute after the point being considered, and taking that value as an estimate of the rate in terms of responses per minute at the midpoint. The theoretical curve in the figure represents the equation

$$r = \frac{13}{1 + 25e^{-.24t}}$$

A considerable part of the variability of the empirical points in the figure is due to the inaccuracy of the method of estimating rates. In order to avoid this loss of precision, the writer has adopted the practice of using cumulative curves of responses *vs.* time for

most purposes, and fitting the cumulative records with the integral of equation (7):

$$R = wt + \frac{w}{B} \log \left( \frac{r_0}{w} + \frac{(w-r_0)}{w} e^{-Bt} \right), \quad (8)$$

where  $R$  represents the number of responses made after any interval of time,  $t$ , from the beginning of the learning period. The original record of responses *vs.* time, from which the data of Fig. 3 were obtained, is reproduced in Fig. 4. Integration of the rate equation for this animal yields

$$R = 13t + 125 \log_{10} (.038 + .962e^{-.24t}).$$

Magnitudes of  $R$  computed from this equation for several values of  $t$  have been plotted in Fig. 4 to indicate the goodness of fit; the theoretical curve has not been drawn in the figure since it would completely obscure most of the empirical record. In an experimental report now in press (2), equation (8) is fitted to several mean conditioning curves for groups of four rats; in all cases, the theoretical curve accounts for

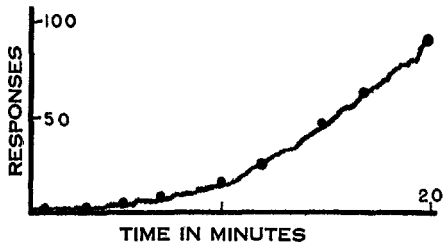


FIG. 4. Reproduction of the original cumulative record from which the points of Fig. 3 were obtained. Solid circles are computed from an equation given in the text.

more than 99 per cent of the variance of the observed  $R$  values. Further verification of the present formulation has been derived from that study by comparing the acquisition curves of successively learned bar-pressing habits, obtained in a Skinner-type conditioning apparatus which included two bars differing only in position. It has been found that the parameters  $w$  and  $s/S$  can be evaluated from the conditioning curve of one bar response, and then used to predict the detailed course of conditioning of a second learned response.

The overall accuracy of these equations in describing the rate of conditioning of bar-pressing and runway responses should not be allowed to obscure the fact that a small but systematic error is present in the initial portion of most of the curves. It is believed that these disparities are due to the fact that experimental conditions do not usually fully realize the assumption that only one  $R$ -class receives any reinforcement during the learning period. A more general formulation of the theory, which does not require this assumption, will be discussed in the next section.

#### PARTIAL REINFORCEMENT

It can be shown that a given response may be "learned" in a trial and error situation provided that some sub-popu-

lation of stimulus elements is so controlled by experimental conditions that each sample of elements drawn from it is contiguous with an occurrence of the response. The sort of derivation needed to handle this kind of partial reinforcement will be sketched briefly in this section. A more detailed treatment will be given, together with relevant experimental evidence, in a paper now in preparation. It should be emphasized that we are using the term "partial" to refer to incomplete change of the stimulus sample on each occurrence of a given response, and not to periodic, or intermittent reinforcement.

Consider a behavior system involving two classes of competing behaviors,  $R$  and  $R_0$ , which may occur in a situation,  $S$ , composed of two independently manipulable sub-populations,  $S_r$  and  $S_0$ . Experimental conditions are to ensure that of the sample,  $s$ , of elements stimulating the organism at any time, elements from  $S_r$  remain effective until terminated by the occurrence of  $R$ , while elements from  $S_0$  remain effective until terminated by the occurrence of  $R_0$ . This kind of system might be illustrated by a Skinner box in which the entire stimulus sample is not terminated by occurrence of the bar-pressing response; for example, if the box is illuminated, the visual stimulation will be relatively unaffected by bar-pressing but will be terminated if the animal closes its eyes (the latter behavior being, then, an instance of  $R_0$ ).

Let  $x$  represent the total number of elements from  $S$  conditioned to  $R$  at a given time,  $x_r$  the number of elements from  $S_r$  conditioned to  $R$ ,  $T_r$  and  $T_0$  the numbers of occurrences of  $R$  and  $R_0$  prior to the time in question, and  $q$  the ratio  $s/S$ . By reasoning similar to that utilized in deriving equations (2) and (5), we may obtain for the average rate of change of  $x_r$  with respect to

$T_r$  at any time

$$\begin{aligned}\frac{dx_r}{dT_r} &= s \cdot \frac{S_r}{S} \cdot \frac{(S_r - x_r)}{S_r} \\ &= q(S_r - x_r).\end{aligned}\quad (9)$$

This may be integrated to yield

$$x_r = S_r - (S_r - x_{r0})e^{-qT_r}, \quad (10)$$

which is identical in form with equation (2).

The other component of  $x$ ,  $(x - x_r)$ , will decrease as these elements become conditioned to the competing response class,  $R_e$ , according to the following relations.

$$\begin{aligned}\frac{d(x - x_r)}{dT_e} &= \frac{-s(S - S_r)}{S} \cdot \frac{(x - x_r)}{(S - S_r)} \\ &= -q(x - x_r)\end{aligned}\quad (11)$$

and the integral,

$$x - x_r = (x_0 - x_{r0})e^{-qT_e}. \quad (12)$$

It will be observed that an analogous set of equations could be written for changes in the number of elements conditioned to  $R_e$ , and that the argument could be extended to any number of mutually exclusive classes of responses. From these relations it is not difficult to deduce differential equations which may be at least numerically integrated to yield curves giving probability of occurrence of each response class as a function of number of reinforcements. We shall not carry out the derivations here, but shall point out a number of properties of the curves obtained which will be evident from inspection of equations (10) and (12).

1. Regardless of the initial probabilities, the behavior system will tend to a state of equilibrium in which the final mean probability of  $R$  will be  $S_r/S$  and the final mean probability of  $R_e$  will be  $(S - S_r)/S$ .

2. If the number of elements from  $S$  conditioned to  $R$  at the start of an

experiment is greater than  $S_r$ , the probability of  $R$  will decrease until the equilibrium is reached. (Of course all statements made here about  $R$  have analogues for  $R_e$ .)

3. If the number of elements from  $S$  conditioned to  $R$  at the start of an experiment is less than  $S_r$ , the probability of  $R$  will increase until the equilibrium value is reached.

4. If all elements originally conditioned to  $R$  belong to the sub-population  $S_r$ , then the curve relating probability to number of reinforcements will be identical with equation (3') except for the asymptote, which will be  $S_r/S$  rather than unity.

5. If some of the elements originally conditioned to  $R$  do not belong to  $S_r$ , but  $x_0$  is less than  $S_r$ , then the curve relating probability to number of reinforcements will rise less steeply at first than equation (3'), and may even have an initial positively accelerated limb.

It will be noted that from the present point of view, conditioning and extinction are regarded simply as two aspects of a single process. In practice we categorize a given experiment as a study of conditioning or a study of extinction depending upon which behaviors are being recorded. It seems quite possible that both conditioning and extinction always occur concurrently in any behavior system, and that the common practice of regarding them as separate processes is based more on tradition and the limitations of recording apparatus than upon rational considerations. In the present formulation, reinforcement is treated as a quantitatively graded variable with "pure extinction" at one end of a continuum. Any portion of an  $S$ -population may be related to an  $R$ -class by experimental conditions which produce a correlation between stimulus sampling and  $R$ -occurrences. Under given conditions of reinforcement an  $R$ -class may in-

crease or decrease in probability of occurrence over a series of trials depending upon whether the momentary probability is less than or greater than the equilibrium value for those conditions.

#### DISCUSSION

The foregoing sections will suffice to illustrate the manner in which problems of learning may be handled within the framework of a statistical theory. The extent to which the formal system developed here may be fruitfully applied to interpret experimental phenomena can only be answered by a considerable program of research. A study of concurrent conditioning and extinction of simple skeletal responses which realizes quite closely the simplified conditions assumed in the derivations of the present paper has been completed, and a report is now in press. Other papers in preparation will apply this formulation to extinction, spontaneous recovery, discrimination, and related phenomena.

The relation of this program to contemporary theories of learning requires little comment. No attempt has been made to present a "new" theory. It is the purpose of our investigation to clarify some of the conceptions of learning and discrimination by stating important concepts in quantitative form and investigating their interrelationships by mathematical analysis. Many similarities will be noted between functions developed here and "homologous" expressions in the quantitative formulations of Graham and Gagné (3) and of Hull (6). A thorough study of those theories has influenced the writer's thinking in many respects. Rather than build directly on either of those formulations, I have felt it desirable to explore an alternative point of view based on a statistical definition of environment and behavior and doing greater justice to the theoretical views of Skinner and Guthrie. A statistical

theory seems to be an inevitable development at the present stage of the science of behavior; agreement on this point may be found among writers of otherwise widely diverse viewpoints, e.g., Brunswik (1), Hoagland (5), Skinner (7), and Wiener (9). It is to be expected that with increasing rigor of definition and continued interplay between theory and experiment, the various formulations of learning will tend to converge upon a common set of concepts.

It may be helpful to outline briefly the point of view on certain controversial issues implied by the present analysis.

*Stimulus-response terminology.* An attempt has been made to overcome some of the rigidity and oversimplification of traditional stimulus-response theory without abandoning its principal advantages. We have adopted a definition of stimulus and response similar to Skinner's (7) concept of generic classes, and have given it a statistical interpretation. Laws of learning developed within this framework refer to behavior systems (as defined in the introductory section of this paper) rather than to relations between isolated stimulus-response correlations.

*The learning curve.* This investigation is not intended to be another search for "the learning function." The writer does not believe that any simple function will be found to account for learning independently of particular experimental conditions. On the other hand, it does seem quite possible that from a relatively small set of definitions and assumptions we may be able to derive expressions describing learning under various specific experimental arrangements.

*Measures of behavior.* Likelihood of responding has been taken as the primary dependent variable. Analyses presented above indicate that simple rela-

tions can be derived between probability and such common experimentally obtained measures as rate of responding and latency.

*Laws of contiguity and effect.* Available experimental evidence on simple learning has seemed to the writer to require the assumption that temporal contiguity of stimuli and behavior is a necessary condition for the formation of conditional relations. At the level of differential analysis, that is of laws relating momentary changes in behavior to changes in independent variables, no other assumption has proved necessary at the present stage of the investigation. In order to account for the accumulation of conditional relations in favor of one *R*-class at the expense of others in any situation, we have appealed to a group of experimental operations which are usually subsumed under the term "reinforcement" in current experimental literature. Both Guthrie's (4) verbal analyses and the writer's mathematical investigations indicate that an essential property of reinforcement is that it ensures that successive occurrences of a given *R* will be contiguous with different samples from the available population of stimuli. We have made no assumptions concerning the role of special properties of certain after-effects of responses, such as drive-reduction, changes in affective tone, etc. Thus the quantitative relations developed here may prove useful to investigators of learning phenomena regardless of the investigators' beliefs as to the nature of underlying processes.

#### SUMMARY

An attempt has been made to clarify some issues in current learning theory by giving a statistical interpretation to the concepts of stimulus and response and by deriving quantitative laws that govern simple behavior systems. Dependent variables, in this formulation, are classes of behavior samples with

common quantitative properties; independent variables are statistical distributions of environmental events. Laws of the theory state probability relations between momentary changes in behavioral and environmental variables.

From this point of view it has been possible to derive simple relations between probability of response and several commonly used measures of learning, and to develop mathematical expressions describing learning in both classical conditioning and instrumental learning situations under simplified conditions.

No effort has been made to defend the assumptions underlying this formulation by verbal analyses of what "really" happens inside the organism or similar arguments. It is proposed that the theory be evaluated solely by its fruitfulness in generating quantitative functions relating various phenomena of learning and discrimination.

#### REFERENCES

1. BRUNSWIK, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, 25, 175-197.
2. ESTES, W. K. Effects of competing reactions on the conditioning curve for bar-pressing. *J. exp. Psychol.* (in press).
3. GRAHAM, C. H., & GAGNÉ, R. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *J. exp. Psychol.*, 1940, 26, 251-280.
4. GUTHRIE, E. R. Psychological facts and psychological theory. *Psychol. Bull.*, 1946, 43, 1-20.
5. HOAGLAND, H. The Weber-Fechner law and the all-or-none theory. *J. gen. Psychol.*, 1930, 3, 351-373.
6. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
7. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
8. SPENCE, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, 51, 47-68.
9. WIENER, N. *Cybernetics*. New York: Wiley, 1948.

[MS. received July 15, 1949]