VOL. 58, NO. 5 SEPTEMBER 1951

THE PSYCHOLOGICAL REVIEW

A MATHEMATICAL MODEL FOR SIMPLE LEARNING

BY ROBERT R. BUSH [1] AND FREDERICK MOSTELLER

Harvard University [2]

Introduction

Mathematical models for empirical phenomena aid the development of a science when a sufficient body of quantitative information has been accumulated. This accumulation can be used to point the direction in which models should be constructed and to test the adequacy of such models in their interim states. Models, in turn, frequently are useful in organizing and interpreting experimental data and in suggesting new directions for experimental research. Among the branches of psychology, few are as rich as learning in quantity and variety of available data necessary for model building. Evidence of this fact is provided by the numerous attempts to construct quantitative models for learning phenomena. The most recent contribution is that of Estes (2).

In this paper we shall present the basic structure of a new mathematical model designed to describe some simple learning situations. We shall focus attention on acquisition and extinction in experimental arrangements using straight runways and Skinner boxes, though we believe the model is more general; we plan to extend the model in order to describe multiple-choice problems and experiments in generalization and discrimination in later papers. Wherever possible we shall discuss the correspondence between our model and the one being developed by Estes (2), since striking parallels do exist even though many of the basic premises differ. Our model is discussed and developed primarily in terms of reinforcement concepts while Estes' model stems from an attempt to formalize association theory. Both models, however, may be re-interpreted in terms of other sets of concepts. This state of affairs is a common feature of most mathematical models. An example is the particle and wave interpretations of modern atomic theory.

We are concerned with the type of learning which has been called "instrumental conditioning" (5), "operant behavior" or "type R conditioning" (10), and not with "classical conditioning" (5), "Pavlovian conditioning" or "type S conditioning" (10). We shall follow Sears (9) in dividing up the chain of events as follows: (1) perception of a stimulus, (2) performance of a response or instrumental act, (3) occurrence of an environmental event,

[1] SSRC-NRC Post-doctoral Fellow in the Natural and Social Sciences.

[2] This research was supported by the Laboratory of Social Relations, Harvard University, as part of a program of the Laboratory's Project on Mathematical Models.

The authors are grateful to many persons for helpful advice and constant encouragement, but especially to Drs. W. O. Jenkins, R. R. Sears, and R. L. Solomon.

and (4) execution of a goal response. Examples of instrumental responses are the traversing of a runway, pressing of a lever, etc. By environmental events we mean the presentation of a "reinforcing stimulus" (**10**) such as food or water, but we wish to include in this category electric shocks and other forms of punishment, removal of the animal from the apparatus, the sounding of a buzzer, etc. Hence any change in the stimulus situation which follows an instrumental response is called an environmental event. A goal response, such as eating food or drinking water, is not necessarily involved in the chain. It is implied, however, that the organism has a motivation or drive which corresponds to some goal response. Operationally speaking, we infer a state of motivation from observing a goal response.

Probabilities and How They Change

As a measure of behavior, we have chosen the probability, p, that the instrumental response will occur during a specified time, h. This probability will change during conditioning and extinction and will be related to experimental variables such as latent time, rate, and frequency of choices. The choice of the time interval, h, will be discussed later. We conceive that the probability, p, is increased or decreased a small amount after each occurrence of the response and that the determinants of the amount of change in p are the environmental events and the work or effort expended in making the response. In addition, of course, the magnitude of the change depends upon the properties of the organism and upon the value of the probability before the response occurred. For example, if the probability was already unity, it could not be increased further.

Our task, then, is to describe the change in probability which occurs after each performance of the response being studied. We wish to express this change in terms of the probability immediately prior to the occurrence of the response and so we explicitly assume that the change is independent of the still earlier values of the probability. For convenience in describing the step-wise change in probability, we introduce the concept of a mathematical operator. The notion is elementary and in no way mysterious: an operator Q when applied to an operand p yields a new quantity Qp (read Q operating on p). Ordinary mathematical operations of addition, multiplication, differentiation, etc., may be defined in terms of operators. For the present purpose, we are interested in a class of operators Q which when applied to our probability p will give a new value of probability Qp. As mentioned above, we are assuming that this new probability, Qp, can be expressed in terms of the old value, p. Supposing Qp to be a well-behaved function, we can expand it as a power series in p:

$$Qp = a_0 + a_1 p + a_2 p^2 + \cdots \quad (1)$$

where $a_0, a_1, a_2, \cdots$ are constants independent of p. In order to simplify the mathematical analysis which follows, we shall retain only the first two terms in this expansion. Thus, we are assuming that we can employ operators which represent a linear transformation on p. If the change is small, one would expect that this assumption would provide an adequate first approximation. Our operator Q is then completely defined as soon as we specify the constants a_0 and a_1; this is the major problem at hand. For reasons that will soon be apparent, we choose to let $a_0 = a$ and $a_1 = 1 - a - b$. This choice of parameters permits us

to write our operator in the form

$$Qp = p + a(1 - p) - bp. \quad (2)$$

This is our basic operator and equation (2) will be used as the cornerstone for our theoretical development. To maintain the probability between 0 and 1, the parameters a and b must also lie between 0 and 1. Since a is positive, we see that the term, $a(1 - p)$, of equation (2) corresponds to an increment in p which is proportional to the maximum possible increment, $(1 - p)$. Moreover, since b is positive, the term, $-bp$, corresponds to a decrement in p which is proportional to the maximum possible decrement, $-p$. Therefore, we can associate with the parameter a those factors which always increase the probability and with the parameter b those factors which always decrease the probability. It is for these reasons that we rewrote our operator in the form given in equation (2).

We associate the event of presenting a reward or other reinforcing stimulus with the parameter a, and we assume that $a = 0$ when no reward is given as in experimental extinction. With the parameter b, we associate events such as punishment and the work required in making the response. (See the review by Solomon [11] of the influence of work on behavior.) In many respects, our term, $a(1 - p)$, corresponds to an increment in "excitatory potential" in Hull's theory (6) and our term, $-bp$, corresponds to an increment in Hull's "inhibitory potential."

In this paper, we make no further attempt to relate our parameters, a and b, to experimental variables such as amount of reward, amount of work, strength of motivation, etc. In comparing our theoretical results with experimental data, we will choose values of a and b which give the best fit. In other words, our model at the present time is concerned only with the form of conditioning and extinction curves, not with the precise values of parameters for particular conditions and particular organisms.

Continuous Reinforcement and Extinction

Up to this point, we have discussed only the effect of the occurrence of a response upon the probability of that response. Since probability must be conserved, *i.e.*, since in a time interval h an organism will make some response or no response, we must investigate the effect of the occurrence of one response upon the probability of another response. In a later paper, we shall discuss this problem in detail, but for the present purpose we must include the following assumption. We conceive that there are two general kinds of responses, overt and non-overt. The overt responses are subdivided into classes A, B, C, etc. If an overt response A occurs and is neither rewarded nor punished, then the probability of any mutually exclusive overt response B is not changed. Nevertheless, the probability of that response A is changed after an occurrence on which it is neither rewarded nor punished. Since the total probability of all responses must be unity, it follows that the probability gained or lost by response A must be compensated by a corresponding loss or gain in probability of the non-overt responses. This assumption is important in the analysis of experiments which use a runway or Skinner box, for example. In such experiments a single class of responses is singled out for study, but other overt responses can and do occur. We defer until a later paper the discussion of experiments in which two or more responses are reinforced differentially.

With the aid of our mathematical

operator of equation (2) we may now describe the progressive change in the probability of a response in an experiment such as the Graham-Gagné runway (3) or Skinner box (10) in which the same environmental events follow each occurrence of the response. We need only apply our operator Q repeatedly to some initial value of the probability p. Each application of the operator corresponds to one occurrence of the response and the subsequent environmental events. The algebra involved in these manipulations is straightforward. For example, if we apply Q to p twice, we have

$$Q^2p = Q(Qp) = a + (1 - a - b)Qp$$
$$= a + (1 - a - b) \times [a + (1 - a - b)p]. \quad (3)$$

Moreover, it may be readily shown that if we apply Q to p successively n times, we have

$$Q^np = \frac{a}{a+b} - \left(\frac{a}{a+b} - p\right) \times (1 - a - b)^n. \quad (4)$$

Provided a and b are not both zero or both unity, the quantity $(1 - a - b)^n$ tends to an asymptotic value of zero as n increases. Therefore, Q^np approaches a limiting value of $a/(a + b)$ as n becomes large. Equation (4) then describes a curve of acquisition.

It should be noticed that the asymptotic value of the probability is not necessarily either zero or unity. For example, if $a = b$ (speaking roughly this implies that the measures of reward and work are equal), the ultimate probability of occurrence in time h of the response being studied is 0.5.

Since we have assumed that $a = 0$ when no reward is given after the response occurs, we may describe an extinction trial by a special operator E which is equivalent to our operator Q of equation (2) with a set equal to zero:

$$Ep = p - bp = (1 - b)p. \quad (5)$$

It follows directly that if we apply this operator E to p successively for n times we have

$$E^np = (1 - b)^np. \quad (6)$$

This equation then describes a curve of experimental extinction.

Probability, Latent Time, and Rate

Before the above results on continuous reinforcement and extinction can be compared with empirical results, we must first establish relationships between our probability, p, and experimental measures such as latent time and rate of responding. In order to do this, we must have a model. A simple and useful model is the one described by Estes (2). Let the activity of an organism be described by a sequence of responses which are independent of one another. (For this purpose, we consider doing "nothing" to be a response.) The probability that the response or class of responses being studied will occur first is p. Since we have already assumed that non-reinforced occurrences of other responses do not affect p, one may easily calculate the mean number of responses which will occur before the response being studied takes place. Estes (2) has presented this calculation and shown that the mean number of responses which will occur, including the one being studied, is simply $1/p$. In that derivation it was assumed that the responses were all independent of one another, *i.e.*, that transition probabilities between pairs of responses are the same for all pairs. This assumption is a bold one indeed (it is easy to think of overt responses that *cannot* follow one another), but it appears to us that any other assumption would

require a detailed specification of the many possible responses in each experimental arrangement being considered. (Miller and Frick [8] have attempted such an analysis for a particular experiment.) It is further assumed that every response requires the same amount of time, h, for its performance. The mean latent time, then, is simply h times the mean number of responses which occur on a "trial":

$$L = \frac{h}{p}. \tag{7}$$

The time, h, required for each response will depend, of course, on the organism involved and very likely upon its strength of drive or motivation.

The mean latent time, L, is expressed in terms of the probability, p, by equation (7), while this probability is given in terms of the number of trials, n, by equation (4). Hence we may obtain an expression for the mean latent time as a function of the number of trials. It turns out that this expression is identical to equation (4) of Estes' paper (2) except for differences in notation. (Estes uses T in place of our n; our use of a difference equation rather than of a differential equation gives us the term $(1 - a - b)$ instead of Estes' e^{-a}.) Estes fitted his equation to the data of Graham and Gagné (3). Our results differ from Estes' in one respect, however: the asymptotic mean latent time in Estes' model is simply h, while we obtain

$$L_\infty = h\left(\frac{a+b}{a}\right). \tag{8}$$

This equation suggests that the final mean latent time depends on the amount of reward and on the amount of required work, since we have assumed that a and b depend on those two variables, respectively. This conclusion seems to agree with the data of Grindley (4) on chicks and the data of Crespi (1) on white rats.

Since equation (7) is an expression for the mean time between the end of one response of the type being studied and the end of the next response of the type being studied, we may now calculate the mean rate of responding in a Skinner-box arrangement. If $\bar{t}$ represents the mean time required for the occurrence of n responses, measured from some arbitrary starting point, then each occurrence of the response being studied adds an increment in $\bar{t}$ as follows:

$$\frac{\Delta \bar{t}}{\Delta n} = \frac{h}{p}. \tag{9}$$

If the increments are sufficiently small, we may write them as differentials and obtain for the mean rate of responding

$$\frac{dn}{dt} = \frac{p}{h} = \omega p, \tag{10}$$

where $\omega = 1/h$. We shall call ω the "activity level" and by definition ω is the maximum rate of responding which occurs when $p = 1$ obtains.

The Free-Responding Situation

In free-responding situations, such as that in Skinner box experiments, one usually measures rate of responding or the cumulative number of responses versus time. To obtain theoretical expressions for these relations, we first obtain an expression for the probability p as a function of time. From equation (2), we see that if the response being studied occurs, the change in probability is $\Delta p = a(1 - p) - bp$. We have already assumed that if other responses occur and are not reinforced, no change in the probability of occurrence of the response being studied will ensue. Hence the expected change in probability during a time interval h is merely the change in probability times the probability p that the re-

sponse being studied occurs in that time interval:

$$\text{Expected } (\Delta p) = p\{a(1-p) - bp\}. \quad (11)$$

The expected rate of change of probability with time is then this expression divided by the time h. Writing this rate as a derivative we have

$$\frac{dp}{dt} = \omega p\{a(1-p) - bp\} \quad (12)$$

where, as already defined, $\omega = 1/h$ is the activity level. This equation is easily integrated to give p as an explicit function of time t. Since equation (10) states that the mean rate of responding, dn/dt, is ω times the probability p, we obtain after the integration

$$\frac{dn}{dt} = \frac{\omega p_0}{p_0(1+u) + [1 - p_0(1+u)]e^{-a\omega t}} = V \quad (13)$$

where we have let $u = b/a$. The initial rate of responding at $t = 0$ is $V_0 = \omega p_0$, and the final rate after a very long time t is

$$V_\infty = \left[\frac{dn}{dt}\right]_{t=\infty} = \frac{\omega}{1+u} = \frac{\omega}{1+b/a}. \quad (14)$$

Equation (13) is quite similar to the expression obtained by Estes except for our inclusion of the ratio $u = b/a$. The final rate of responding according to equation (14), increases with a and hence with the amount of reward given per response, and decreases with b and hence with the amount of work per response. These conclusions do not follow from Estes' results (2).

An expression for the cumulative number of responses during continuous reinforcement is obtained by integrating equation (13) with respect to time t. The result is

$$n = \frac{1}{1+u}\left\{\omega t + \frac{1}{a}\log\left[p_0(1+u) \times (1 - e^{-a\omega t}) + e^{-a\omega t}\right]\right\}. \quad (15)$$

As the time t becomes very large, the exponentials in equation (15) approach zero and n becomes a linear function of time. This agrees with equation (14) which says that the asymptotic rate is a constant. Both equations (13) and (15) for rate of responding and cumulative number of responses, respectively, have the same form as the analogous equations derived by Estes (2) which were fitted by him to data on a bar-pressing habit of rats. The essential difference between Estes' results and ours is the dependence, discussed above, of the final rate upon amount of work and amount of reward per trial.

We may extend our analysis to give expressions for rates and cumulative responses during extinction. Since we have assumed that $a = 0$ during extinction, we have in place of equation (12)

$$\frac{dp}{dt} = -\omega b p^2 \quad (16)$$

which when integrated for p and multiplied by ω gives

$$\frac{dm}{dt} = \frac{\omega p_e}{1 + \omega b p_e t} \quad (17)$$

where p_e is the probability at the beginning of extinction. The rate at the beginning of extinction is $V_e = \omega p_e$. Hence we may write equation (17) in the form

$$V = \frac{dm}{dt} = \frac{V_e}{1 + V_e b t}. \quad (18)$$

An integration of this equation gives for the cumulative number of extinc-

tion responses

$$m = \frac{1}{b}\log\,[1 + V_e bt] = \frac{1}{b}\log\left(\frac{V_e}{V}\right). \quad (19)$$

This result is similar to the empirical equation $m = K \log t$, used by Skinner in fitting experimental response curves (10). Our equation has the additional advantage of passing through the origin as it must.

It may be noted that the logarithmic character of equation (19) implies that the total number of extinction responses, m, has no upper limit. Thus, if our result is correct, and indeed if Skinner's empirical equation is correct, then there is no upper limit to the size of the "reserve" of extinction responses. For all practical purposes, however, the logarithmic variation is so slow for large values of the time t, it is justified to use some arbitrary criterion for the "completion" of extinction. We shall consider extinction to be "complete" when the mean rate of responding V has fallen to some specified value, V_f. Thus, the "total" number of extinction responses from this criterion is

$$m_T = \frac{1}{b}\log\frac{V_e}{V_f}. \quad (20)$$

We now wish to express this "total" number of extinction responses, m_T, as an explicit function of the number of preceding reinforcements, n. The only quantity in equation (20) which depends upon n is the rate, V_e, at the beginning of extinction. If we assume that this rate is equal to the rate at the end of acquisition, we have from equations (4) and (10)

$$V_e = \frac{dn}{dt} = \omega p_n = V_{\max} - (V_{\max} - V_0)(1 - a - b)^n \quad (21)$$

where we have let

$$V_{\max} = \omega\frac{a}{a + b}, \quad (22)$$

and where $V_0 = \omega p_0$ is the rate at the beginning of acquisition. If we now substitute equation (21) into equation (20), we obtain

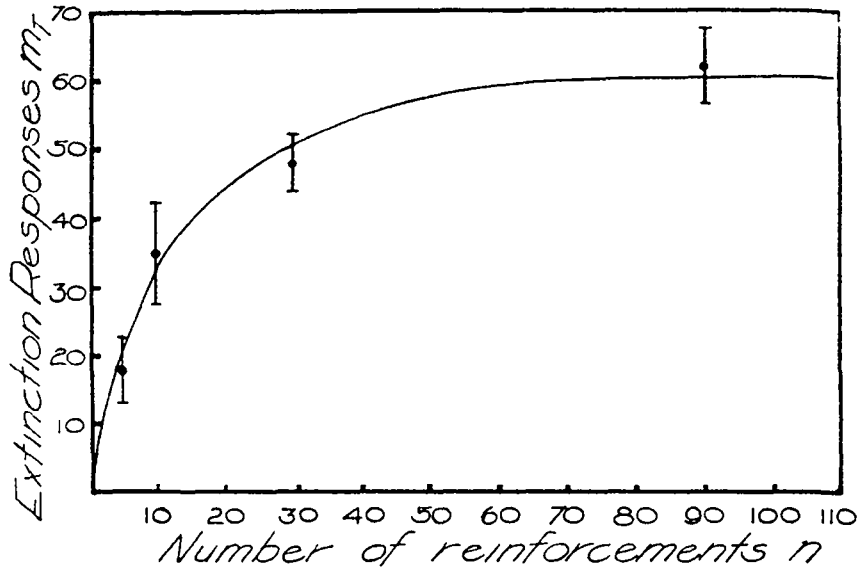
$$m_T = \frac{1}{b}\log\left\{\frac{V_{\max}}{V_f} - \left[\frac{V_{\max}}{V_f} - \frac{V_0}{V_f}\right] \times (1 - a - b)^n\right\}. \quad (23)$$

This result may be compared with the data of Williams (12) obtained by measuring the "total" number of extinction responses after 5, 10, 30 and 90 reinforcements. From the data, the ratio $V_{\max}/V_f$ was estimated to be about 5, and the ratio V_0/V_f was assumed to be about unity. Values of $a = 0.014$ and $b = 0.026$ were chosen in fitting equation (23) to the data. The result is shown in the figure.

Fixed Ratio and Random Ratio Reinforcement

In present day psychological language, the term "fixed ratio" (7) refers to the procedure of rewarding every kth response in a free-responding situation ($k = 2, 3, \cdots$). In a "random ratio" schedule, an animal is rewarded on the average after k responses but the actual number of responses per reward varies over some specified range. We shall now derive expressions for mean rates of responding and cumulative numbers of responses for these two types of reinforcement schedules. If we apply our operator Q, of equation (2), to a probability p, and then apply our operator E, of equation (5), to Qp repeatedly for $(k - 1)$ times, we obtain

$$(E^{k-1}Q)p = (1-b)^{k-1}[p + a(1-p) - bp] = p + a'(1-p) - b'p \quad (24)$$



"Total" number of extinction responses as a function of the number of reinforcements. Curve plotted from equation (23) with $b = 0.026$, $a = 0.014$, $V_{max} = 5V_0$, $V_f = V_0$. Data from Williams (12).

where

$$a' = a(1-b)^{k-1} = a\{1-(k-1)b+\cdots\} \cong a \quad (25)$$

and

$$b' = 1-(1-b)^k = kb\left\{1-\frac{k-1}{2}b+\cdots\right\} \cong kb. \quad (26)$$

The symbol $\cong$ means "approximately equal to." In the present case the exact approach would be to retain the primes on a and b throughout; however the approximations provide a link with the previous discussion. The approximations on the right of these two equations are justified if kb is small compared to unity. Now the mean change in p *per response* will be the second and third terms of equation (24) divided by k:

$$\Delta p = \frac{a'}{k}(1-p) - \frac{b'}{k}p \cong \frac{a}{k}(1-p) - bp. \quad (27)$$

This equation is identical to our result for continuous reinforcement, except that a'/k replaces a and b'/k replaces b

We may obtain a similar result for the "random ratio" schedule as follows: After any response, the probability that Q operates on p is $1/k$ and the probability that E operates on p is $(1 - 1/k)$. Hence the expected change in p per response is

$$\text{Expected } (\Delta p) = \frac{1}{k}Qp + (1-1/k)Ep - p. \quad (28)$$

After equations (2) and (5) are inserted and the result simplified, we obtain from equation (28)

$$\text{Expected } (\Delta p) = \frac{a}{k}(1-p) - bp. \quad (29)$$

This result is identical to the approximate result shown in equation (27) for the fixed ratio case. Since both equations (27) and (29) have the same

form as our result for the continuous reinforcement case, we may at once write for the mean rate of responding an equation identical to equation (13), except that a is replaced by a'/k. Similarly, we obtain an expression for the final rate of responding identical to equation (14) except that a is replaced by a'/k. This result is meant to apply to both fixed ratio and random ratio schedules of reinforcement.

In comparing the above result for the asymptotic rates with equation (14) for continuous reinforcement, we must be careful about equating the activity level, ω, for the three cases (continuous, fixed ratio and random ratio reinforcements). Since $1/\omega$ represents the minimum mean time between successive responses, it includes both the eating time and a "recovery time." By the latter we mean the time necessary for the animal to reorganize itself after eating and get in a position to make another bar press or key peck. In the fixed ratio case, presumably the animal learns to look for food not after each press or peck, as in the continuous case, but ideally only after every k response. Therefore both the mean eating time and the mean recovery time *per response* are less for the fixed ratio case than for the continuous case. In the random ratio case, one would expect a similar but smaller difference to occur. Hence, it seems reasonable to conclude that the activity level, ω, would be smaller for continuous reinforcement than for either fixed ratio or random ratio, and that ω would be lower for random ratio than for fixed ratio when the mean number of responses per reward was the same. Moreover, we should expect that ω would increase with the number of responses per reward, k. Even if eating time were subtracted out in all cases we should expect these arguments to apply. Without a quantitative estimate of the mean recovery time, we see no meaningful way of comparing rates of responding under continuous reinforcement with those under fixed ratio and random ratio, nor of comparing rates under different ratios (unless both ratios are large). The difficulty of comparing rates under various reinforcement schedules does not seem to be a weakness of our model, but rather a natural consequence of the experimental procedure. However, the importance of these considerations hinges upon the orders of magnitude involved, and such questions are empirical ones.

Aperiodic and Periodic Reinforcement

Many experiments of recent years were designed so that an animal was reinforced at a rate aperiodic or periodic in time (7). The usual procedure is to choose a set of time intervals, $T_1, \cdots, T_n$, which have a mean value T. Some arrangement of this set is used as the actual sequence of time intervals between rewards. The first response which occurs after one of these time intervals has elapsed is rewarded.

To analyze this situation we may consider k, the mean number of responses per reward, to be equal to the mean time interval T multiplied by the mean rate of responding:

$$k = T\frac{dn}{dt} = T\omega p. \qquad (30)$$

Equation (29) for the expected change in probability per response is still valid if we now consider k to be a variable as expressed by equation (30). Thus, the time rate of change of p is

$$\frac{dp}{dt} = \frac{a}{T}(1 - p) - \omega b p^2. \qquad (31)$$

With a little effort, this differential equation may be integrated from 0 to

t to give

$$\frac{dn}{dt} = \omega p = \frac{\omega}{z}\frac{(s-1)+(s+1)Ke^{-\omega at/T}}{1-Ke^{-\omega at/T}} \quad (32)$$

where

$$z = 2\omega Tb/a, \quad (33)$$

$$s = \sqrt{1+2z}, \quad (34)$$

$$K = (1+zp_0-s)/(1+zp_0+s). \quad (35)$$

For arbitrarily large times t, the final rate is

$$\left(\frac{dn}{dt}\right)_{t=\infty} = \frac{\omega}{z}(s-1). \quad (36)$$

For sufficiently large values of T, z becomes large compared to unity and we may write approximately

$$\left(\frac{dn}{dt}\right)_{t=\infty} = \omega\sqrt{2/z} = \omega\sqrt{a/b\omega T}. \quad (37)$$

Thus, for large values of T, the final rate varies inversely as the square root of T.

Periodic reinforcement is a special case of aperiodic reinforcement in which the set of time intervals, $T_1, \cdots, T_n$, discussed above, consists of a single time interval, T. Thus, all the above equations apply to both periodic and aperiodic schedules. One essential difference is known, however. In the periodic case the animal can learn a time discrimination, or as is sometimes said, eating becomes a cue for not responding for a while. This seems to be an example of stimulus discrimination which we will discuss in a later paper.

Extinction After Partial Reinforcement Schedules

In the discussion of extinction in earlier sections, it may be noted that the equations for mean rates and cumulative responses depended on the previous reward training only through V_e, the mean rate at the beginning of extinction. Hence, we conclude that equations (18) and (19) apply to extinction after any type of reinforcement schedule. However, the quantities V_e and b in our equations may depend very much on the previous training. Indeed, if our model makes any sense at all, this must be the case, for "resistance" to extinction is known to be much greater after partial reinforcement training than after a continuous reinforcement schedule (7).

Since the rate at the start of extinction, V_e, is nearly equal to the rate at the end of acquisition, it will certainly depend on the type and amount of previous training. However, the logarithmic variation in equations (19) and (20) is so slow, it seems clear that empirical results demand a dependence of b on the type of reinforcement schedule which preceded extinction. We have argued that b increases with the amount of work required per response. We will now try to indicate how the required work might depend upon the type of reinforcement schedule, even though the lever pressure or key tension is the same. For continuous reinforcement, the response pattern which is learned by a pigeon, for example, involves pecking the key once, lowering its head to the food magazine, eating, raising its head, and readjusting its body in preparation for the next peck. This response pattern demands a certain amount of effort. On the other hand, the response pattern which is learned for other types of reinforcement schedules is quite different; the bird makes several key pecks before executing the rest of the pattern just described. Thus we would expect that the average work required per *key peck* is considerably less than for continuous reinforcement. This would imply that b is larger and thus "resistance" to extinction is less

for continuous reinforcement than for all other schedules. This deduction is consistent with experimental results (7). However, this is just part of the story. For one thing, it seems clear that it is easier for the organism to discriminate between continuous reinforcement and extinction; we have not handled this effect here.

Summary

A mathematical model for simple learning is presented. Changes in the probability of occurrence of a response in a small time h are described with the aid of mathematical operators. The parameters which appear in the operator equations are related to experimental variables such as the amount of reward and work. Relations between the probability and empirical measures of rate of responding and latent time are defined. Acquisition and extinction of behavior habits are discussed for the simple runway and for the Skinner box. Equations of mean latent time as a function of trial number are derived for the runway problem; equations for the mean rate of responding and cumulative numbers of responses versus time are derived for the Skinner box experiments. An attempt is made to analyze the learning process with various schedules of partial reinforcement in the Skinner type experiment. Wherever possible, the correspondence between the present model and the work of Estes (2) is pointed out.

REFERENCES

1. CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.
2. ESTES, W K. Toward a statistical theory of learning PSYCHOL. REV., 1950, **57**, 94–107.
3. GRAHAM, C., AND GAGNÉ, R. M. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *J. exp. Psychol.*, 1940, **26**, 251–280.
4. GRINDLEY, C. C. Experiments on the influence of the amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929–30, **20**, 173–180.
5. HILGARD, E. R., AND MARQUIS, D. G. *Conditioning and learning*. New York: D. Appleton-Century Co., 1940.
6. HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
7. JENKINS, W. O., AND STANLEY, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
8. MILLER, G. A., AND FRICK, F. C. Statistical behavioristics and sequences of responses. PSYCHOL. REV., 1949, **56**, 311–324.
9. SEARS, R. R. Lectures at Harvard University, Summer, 1949.
10. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century-Crofts, 1938.
11. SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45**, 1–40.
12. WILLIAMS, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–521.

[MS. Received September 21, 1950]