ALL-OR-NONE PROCESSES IN LEARNING AND RETENTION[1]

W. K. ESTES

Stanford University

IF I should pull a match from my pocket and strike it, no one in the audience would be surprised at the sudden appearance of fire and smoke. The reason is not simply that we are all accustomed to predicting and controlling the behavior of matches, but rather that we have available a satisfying account of what occurs during the short interval when a bit of sulphur disappears and the air around it bursts into flame. To less civilized individuals, unprepared to visualize what happens during the kindling of a fire in terms of the breakdown of molecules into their constituents and recombination into new molecules, the same phenomenon remains a perpetual source of mystery. But despite all our sophistication concerning molecules and atoms, the situation contains fully as much mystery for us. Although we may feel that we adequately understand the lighting of the match, we can offer no comparable account of the events that occur when an individual is learning to strike one. Eliminating this conceptual blind spot is one of the tasks, indeed to my mind the principal task, of learning theory.

How may we best direct our efforts if we hope to achieve an adequate picture of the processes and events involved in any instance of learning? Simply to the industrious accumulation of parametric data? This seems rather dubious when the mazes and memory drums, the cumulative recorders and variance analyzers already are producing data at a rate well exceeding our capacity to absorb them. When our purpose is to understand a complex system, sheer quantity of information may obstruct more than it illuminates. Suppose we were set the task of comprehending the workings of a metropolitan telephone system. We would make a slow job of it if we proceeded by determining the additive and interactive effects of factorially combined independent variables upon dollar volume of telephone bills.

Perhaps we could more strategically concentrate attention on particular types and aspects of data that seem likely to be of special diagnostic value. If, as a start in this direction, we consider analogies between properties of the learning organism and those of other organized systems that we know rather more about, such as communication networks and computing machines, we can scarcely fail to be impressed by the fact that it is of the very nature of organized systems to exhibit discontinuities—that is, sharp departures from proportionality of causes and effects. Thus locating discontinuities, despite the fact that they are normally masked by noise, i.e., experimental error, may well be one of our principal objectives. Even in our earliest efforts toward deciphering the organization of the learning process, one clear bit of evidence for a quantal change of state of the system may be of more diagnostic value than a mountain of data exhibiting gradual changes in output as a function of input.

These introductory remarks are intended, not as a prescription or admonition to anyone else, but as a brief reconstruction of the line of thinking that has led some colleagues and me to find special fascination in the uncovering of new sources of evidence for discontinuous changes in the organism's system of behavioral dispositions during learning.

The particular bit of evidence which seemed most compelling to me when I had occasion to address a division of this Association some 4 years ago (Estes, 1960) resulted from detailed analyses of the changes in response probabilities effected by a single reinforced trial in several standard learning situations. The observed pattern of changes uniformly agreed closely with that expected on the assumption of all-or-none association. The experiments were of the RTT design, that is, a single reinforcement followed by two successive unreinforced test trials. The results indicated that a single reinforcement left an item in one or the other of

[1] Preparation of this paper was supported in part by the Personnel and Training Branch, Office of Naval Research, under Contracts 908(16) with Indiana University and 225 (73) with Stanford University, and by a grant from the Carnegie Corporation of New York to Stanford University.

two quite distinct states: a learned state in which the correct response had a high probability on both tests, or an unlearned state in which the correct response had only chance probability on either test.

This is not to say that the observed pattern of test data could not at the same time agree with any other conception. Alternative interpretations of data can always be produced; consequently, the massive demonstrations (see, e.g., Postman, 1963; Underwood & Keppel, 1962) that such exist in the present instance have impressed me as being useful, but not necessarily of great theoretical import. Certainly we need to find ways of purifying experimental situations: for example, in the case of paired-associate learning, by reducing individual differences in conditioning rates and differences of item difficulty. But no one investigator can do everything, and I have been inclined to leave this purification to individuals expert in the particular areas while I look in other directions that I personally find more congenial.

Sometimes one can make most rapid progress in differentiating among possible theoretical interpretations by looking for many alternative sources of evidence. If the same type of assumption continues to generate simple and satisfying accounts of results within a wide variety of situations, or in the same situation with a wide variety of variations in procedure, we naturally come to believe that the scheme is essentially correct. And long-term scientific experience seems to indicate that this is not a bad strategy even though it cannot be fully justified to the satisfaction of logicians.

When, in particular, we are engaged in analyzing the effect of a single learning trial, we can expect to obtain independent sources of evidence concerning the process involved by looking in different temporal directions from the point of reinforcement. Thus it seemed natural to follow up the studies of response shifts immediately following the reinforcement by examining the effects of events immediately *preceding* the point of reinforcement.

To the extent that the process of learning a response is basically akin to the flipping of a binary switch, we might expect the effect of a single learning trial to be relatively independent of the immediate reinforcement history. This expectation is quite in contrast with that of strong dependence which would, for example, be entailed by an interpretation in terms of thresholds and oscillation distributions of competing responses.

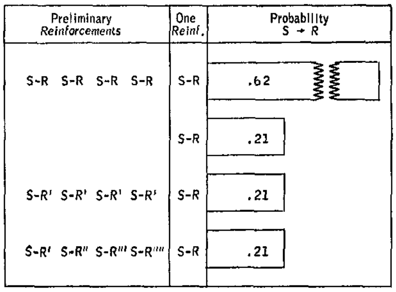

FIG. 1. Effect of a single learning trial in relation to differing reinforcement histories. (Probability of a response R to a stimulus S is assessed following five reinforcements of R, one reinforcement of R, four reinforcements of a single competing response and one of R, or one reinforcement of each of four competing responses and one reinforcement of R.)

One type of experiment designed to yield evidence pertinent to these disparate expectations is schematized in Figure 1. This experiment was conducted at Indiana University with the assistance of Judith Crooks. Forty subjects, undergraduate students, were run on two paired-associate items under each condition (the stimuli being consonant trigrams and the responses English words). The common focus for all conditions was a single reinforcement of a response to a particular stimulus followed by a test to assess the effect of the reinforcement.

For items of the type illustrated in the second row of the figure, neither the stimulus nor the response member of the item had been involved in the experiment in any way prior to this single reinforced trial. In the type shown in the first row, the only difference in procedure was that the given response was reinforced four times in the presence of the same stimulus on preceding trials so that this item had five reinforcements of the correct response prior to the test trial. Inclusion of this condition provides a baseline measure of the effectiveness of additional reinforcements of the same item.

The item type exhibited in the third row of the figure differs from the last described only in that on the four trials preceding the reinforcement of the given response R, some one different response R′ was reinforced. Finally, in the fourth row is exhibited an item type for which a different competing response was reinforced on each of the preceding four trials. From the standpoint of an all-or-none conception of the associative process, the items shown in the second, third, and fourth rows are all alike in having a single opportunity for learning of the correct response immediately prior to the test trial; whereas for the type shown in the top row, the correct response has had a larger number of opportunities to become associated with the stimulus present on the test trial.

The results portrayed in the figure speak quite well for themselves. I might add that, being properly impressed with the dangers of accepting null hypotheses, we replicated the second and fourth conditions as part of a later study with a larger number of observations and obtained quite comparable results (correct response probabilities of .25 and .24 based on *N*s of 160 and 320, respectively, on the test trial). The charm of these data for me is not that they confirm any one previously formulated model, or even that they raise problems for other extant theories, but that they provide a relatively direct experimental demonstration of an independence-of-path property which has hitherto been assumed in stimulus sampling theories (see, for example, Bush & Mosteller, 1955; Estes, 1959) almost solely on grounds of mathematical simplicity.

It is only natural that new information concerning all-or-none processes should often come from especially contrived experiments, such as the one just discussed. A more surprising, and in some ways even more rewarding, development of the last few years has been the uncovering of previously unsuspected, or at least undemonstrated, evidence for the existence of steady states and discontinuities in the data of standard experiments. The methods of quantitative analysis largely responsible for these findings have grown directly out of attempts to formulate mathematical models of the associative process based on assumptions derived, in part, from the "one-trial" experiments.

Individuals who have been imprinted in the classical tradition of schools and systems of psychology seem to have almost insuperable difficulty in appreciating the function of models in this type of research enterprise. According to the conventional view, a model is chosen to represent a theoretical position; then the representatives of different positions are pitted against each other in crucial experiments, the winner being "accepted" and the losers fading gracefully into oblivion, all much in the spirit of a Miss America contest.

But for those who actually use them in scientific research, the function of models is quite different. The set of theoretical ideas suggested by an array of experimental facts can never be adequately represented by any one formal model. On the contrary, to bring originally vague and incompletely defined theoretical notions to fruition, it is necessary to go through a series of successive approximations. At each stage, a specific, usually highly simplified and idealized realization of the general assumptions, that is, a particular mathematical model, is examined for its theoretical implications, its successes or shortcomings in describing data, its suggestiveness in guiding further experimental and quantitative analysis.

Since standard learning situations characteristically (perhaps universally) include important sources of variation in performance other than the effects of the reinforcing operation, raw data can be expected to reveal the outlines of underlying behavioral processes only if examined on the viewing screen, so to speak, of a model which selectively filters signal from noise. Thus, in the case of all-or-none learning theory, a major step conceptually, though a minor one from a mathematical viewpoint, was the formulation of what has been termed the one-element model.

The mathematical formalism of the one-element model evolved independently in the hands of several different investigators: Bower (1961a), Bush and Mosteller (1959); Estes, Hopkins, and Crothers (1960). But it was Bower who first saw the full potentiality of the model for analyzing verbal learning experiments and spelled out its implications in such detail as to make it a workable tool of analysis.

The assumptions of this model are illustrated in Figure 2. The graph represents probability of correct response versus number of training trials for a number of individual subjects. It is assumed for simplicity that all subjects begin at the same initial level of correct response probability (operant level, or guessing rate), and that on each training trial the subject has some fixed probability of having his correct response probability jump from the initial level to unity. Because of the probabilistic nature of the process, different subjects jump to unity on different trials. Thus if data are pooled for a group of subjects, even though all have identical initial levels and identical probabilities of conditioning on each trial, a curve representing proportion of correct responses per trial should average

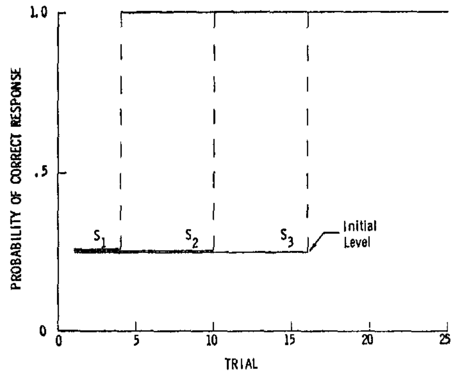

FIG. 2. Illustrative learning functions for individual subjects as conceived in the one-element model.

out to something like the familiar growth curve. In some experiments the initial level of response probability may be expected to be zero; but in many, for example multiple-choice or paired-associate experiments, the initial probability of a correct response by guessing may have some value other than unity, frequently simply the reciprocal of the number of available response alternatives.

Some predictions derivable from this simple model are immediately obvious upon consideration of the assumptions; others are not so obvious until one has done some actual calculations and may even be missed by experts. In Figure 3 we show a sample of predictions derivable from the model. Assuming that all subjects start from an initial response probability of .25 and the probability of forming a correct association is .50 on each trial, we readily derive the mean curve for proportion of correct responses per trial which is shown as the heavy line, and which has the form that we anticipated.

The probability that a correct response will occur on any trial if the given subject made an incorrect response to the same item on the preceding trial is constant, as indicated by the horizontal dashed line toward the bottom of the graph. This prediction follows from the model, of course, only if experimental conditions and previous experience of the subjects are such that there are no learning-to-learn effects which would cause the probability of conditioning to change over trials. In the upper part of the graph is shown the curve representing probability of correct recalls as a function of trials, that is, the probability that a correct response to a given stimulus on one trial will be followed by a correct response to the same stimulus on the next trial. This result is of special interest because many investigators have believed it intuitively evident that this function should be constant over trials; and, in fact, data which follow quite nicely the trend of the dashed line in the figure have been taken to refute the one-element learning model. Such quaint illogicalities point up quite sharply the need for checking our intuitive ideas as to the consequences of theoretical assumptions by embodying the assumptions in specific models and deriving their implications by exact methods.

Finally, at the top of the graph, is shown the function representing probability of correct response on any trial given that it has been correct on the entire sequence of preceding trials; this function again has sometimes been assumed by the unwary to be constant if the one-element model holds; but that is true only in the special case when the initial level of response probability is zero. In all other cases, one has, following the first reinforced trial, a mixture of subjects who are attaining correct responses by guessing and subjects who have learned; as trials proceed, the balance shifts in the direction of the latter, thus giving rise to the rising curve for probability of correct recalls.

Now, having a model clearly formulated, what should we do with it? Proceed to apply it to various experiments with the plan of retaining it as long as it succeeds and rejecting it upon the first failure? This course would entail an absurd waste of time, for we can be certain in advance that no simple model will fit all sets of data even within a relatively limited area.

It is, however, of interest to see whether the model fits any data at all. Even the laws of mo-

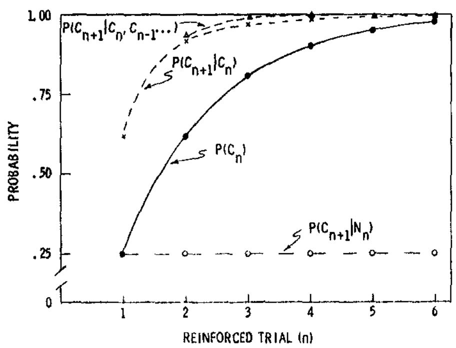

FIG. 3. Theoretical functions derivable from the one-element model, the indicated statistic in each case representing data pooled over subjects on each trial. (The conditioning parameter is taken equal to .50.)

TABLE 1

BOWER VERBAL DISCRIMINATION DATA COMPARED WITH PREDICTIONS FROM ONE-ELEMENT MODEL

Error distribution						
	M			*SD*		
Observed	1.99			1.94		
Predicted	(c = .25)			1.98		
Reset on error[a]						
n	1	2	3	4	5	Aver.
Observed	1.52	1.40	1.56	1.46	1.48	1.48
Predicted	1.49 →		(constant)		→	1.49
Constancy of learning parameter[b]						
k	1	2	3	4–5		
p	.60	.62	.56	.60		
Precriterion stationarity[c]						
	First half			Second half		
Observed	.492			.521		
Predicted	.500			.500		

[a] Mean errors to criterion after error on trial n.
[b] Proportion correct on trial following k successive errors.
[c] Proportion correct over precriterion sequence.

tion rarely provide highly accurate predictions of the behavior of real objects; but in specially contrived situations, where perfect vacuums and frictionless planes are approximately realized, the Newtonian model comes off well enough to convince us that its abstract properties represent more than figments of our imagination.

In the case of the one-element model the first thorough-going applications were made by Gordon Bower (1961a, 1961b, 1962) and with successes far exceeding any expectations that I, at least, would have considered realistic.

A small sample of data from one of Bower's studies (1961b) is summarized in Table 1. The situation giving rise to the results summarized here was a simple verbal discrimination learning task, in which the materials were cards each containing a pair of nonsense syllables, one of which was arbitrarily designated as correct; and the subject's task on each trial was to guess which member of the pair shown on the given trial was correct. With knowledge of results following each correct or incorrect guess, subjects continued to cycle through the cards until they reached a criterion of correct performance. Estimating the value of the conditioning parameter, c, that is the probability that an association between stimulus and correct response would occur on a given trial for a given item, from the observed mean total errors, Bower proceeded to predict quantitative values for numerous statistics of the data. An example, shown on the top line of Table 1, is the prediction of the standard deviation of total errors per item. Investigators who are familiar by much experience with the difficulties of predicting variances as opposed to means may be somewhat impressed by the correspondence between the predicted and observed values. Similar agreement was found by Bower for a large number of statistics, including distribution of trials before the first success, frequencies of error runs of various lengths, and autocorrelations of errors with various lags.

Some additional results which point up particular aspects of the model are shown in the remainder of this table. The heading "reset on error" refers to the important property of the one-element model that the entire process starts over after each error. That is, the occurrence of an error on any trial signifies that up to that point no learning has occurred for that subject on that item. Consequently it is predicted that such statistics as mean errors to criterion after an error on trial n should be constant over trials. For the present data we can even see readily what the constant value should be. Since mean observed total errors, shown at the top, were 1.99, and the probability of an error on the first trial before the subject had received any reinforcements must have been .50 on the average, we can subtract the .50 from the total of 1.99 and arrive at a prediction of 1.49 for the constant mean errors to criterion following an error on any trial. The observed values seem to vary around this prediction with little indication of any systematic trend. If it is the case not only that learning is occurring on an all-or-none basis but that probability of learning is constant over trials, then the proportion of correct responses on any trial following a sequence of preceding errors should be independent of the number of preceding errors. The next to last row of the table shows the value of this proportion following one, two, three, etc., successive preceding errors (the data for 4 and 5 preceding errors being pooled since the number of cases was falling off).

The remaining type of analysis represented in the table arises from a valuable insight growing out of the reset-on-error property. The gist of the idea, developed fully by Suppes and Ginsberg (1963), is that despite the fact that one cannot observe directly the point at which learning occurs, one can define unequivocally a last point before which learning certainly has not occurred if the one-element model holds. Referring to the illustrative protocols in Figure 4, one can see that this critical point in each protocol is the trial of the last error. If the learning process is correctly described by the model, then on all trials prior to the last error, the subject's responses are in effect being generated by a Bernoulli process with all the properties of coin tossing. Correct responses and errors during this precriterion sequence should occur at random with constant probability (not necessarily one-half except in certain two-choice situations) and independence from trial to trial. Thus cogent evidence concerning the existence of an initial steady state of random responding in any situation may be obtained by analyzing individual protocols, locating the last error in each, and performing various analyses on the data of the precriterion sequence.

One of the coarsest, though at the same time most intuitively appealing, of these analyses is a test for precriterion stationarity, that is, constancy of the probability of correct responding over the precriterion sequence, illustrated in the remaining section of the table for Bower's data. Because learning was rather rapid, the data have been pooled into observed and predicted proportions of correct responses in the first and second halves of

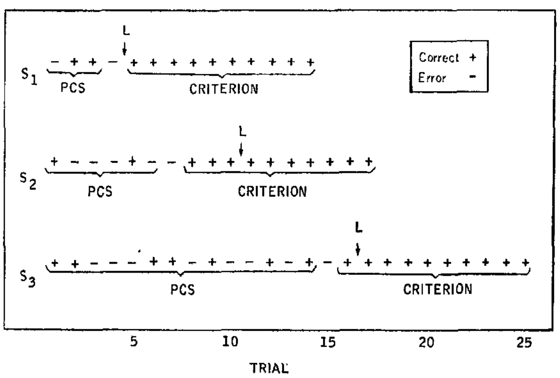

FIG. 4. Illustrative protocols for individual subjects, with the precriterion sequence (PCS), the point of learning (L), and the criterion sequence indicated for each.

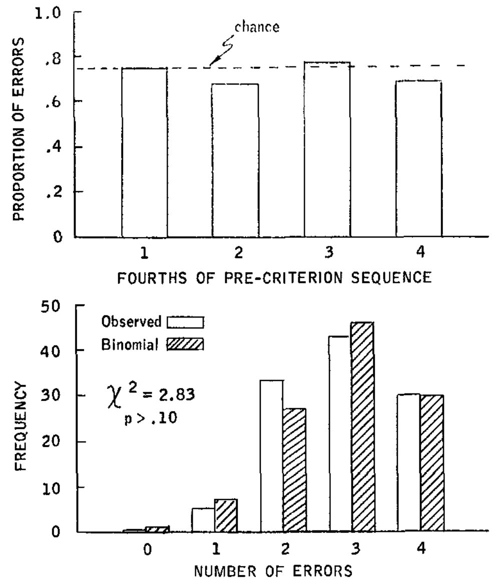

FIG. 5. Precriterion analysis of data from Suppes' study of the learning of mathematical proofs by a group of 50 young children. (The upper diagram exhibits relative constancy of error probability; the lower diagram compares the distribution of error frequencies per four-trial block with the corresponding binomial distribution.)

the precriterion sequence. Departures from the predicted constancy at .5 do not seem great.

Many other results comparable to these have been reported during the last 2 or 3 years. Following are a few illustrations showing some of the situations in which these results arise. Figure 5 shows an analysis of precriterion data from an experiment by Suppes on the learning of mathematical proofs by young school children.[2] The upper portion of the figure shows the constancy of proportion of errors over successive fourths of the precriterion sequence predicted by the one-element model; the lower panel shows close agreement between the data and the binomial distribution of different numbers of errors predicted when data are pooled by four-trial blocks over the entire precriterion sequence. Up to the point when they are able to give several correct responses in a row, the behavior of these children in trying to provide successive steps of proofs in a miniature logical system are much like the data that would be obtained by tossing so many biased pennies.

In Figure 6 are shown data from a study of concept identification (Trabasso, 1963). This figure

[2] I am indebted to Patrick Suppes for making available these unpublished data.

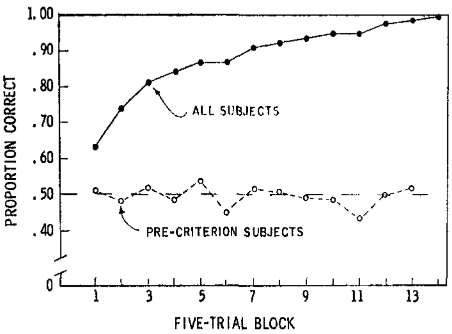

FIG. 6. Learning curve averaged over all subjects and items and curve representing precriterion data only, from Trabasso's (1963) study of concept identification.

shows the dramatic contrast between the curve representing proportion of correct responses per trial block for all subjects over all trials and the plot for the same statistic limited to data from precriterion sequences. Whereas the overall mean curve rises steadily and smoothly, the proportion of correct responses per trial block is about as constant as one could ask for over the precriterion sequences. These data are of especial interest since it has sometimes been suggested that data conforming to predictions from all-or-none models arise only in situations where learning is very rapid; it can be seen that in this situation learning was rather slow, yet precriterion stationarity was nicely realized.

I shall not go on giving examples of all but perfect fits of the one-element model to data because these, although perhaps impressive, cease being particularly instructive.

For an example of a case in which a rather interesting deviation occurs, we may consider Figure 7,

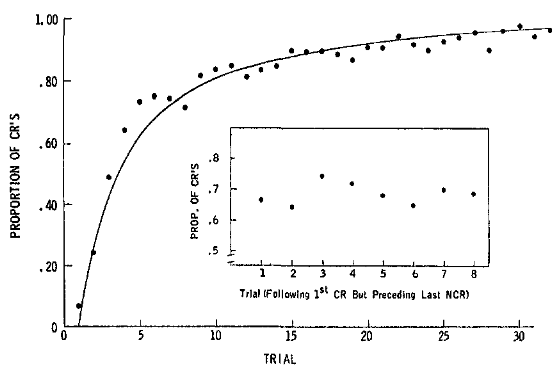

FIG. 7. Acquisition curve for eyelid conditioning data. (The inset shows stationarity of conditioned response proportion on trials following the first CR, but preceding the last non-CR.)

which shows pooled data from a series of eyelid conditioning experiments with human subjects.[3] The curve for proportion of conditioned responses per trial is again nicely negatively accelerated, but this time, when a stationarity analysis is performed, it is found immediately that proportion of correct responses does *not* remain constant over the precriterion sequence; rather the curve exhibits a steady rise, thus showing at once that the one-element model cannot give an adequate representation of the data. However, upon further analysis, stationarity reappears in a rather interesting way. If we truncate the precriterion sequence by considering only trials following the first conditioned response for each subject, then, as shown in the inset, the proportion of correct responses per trial

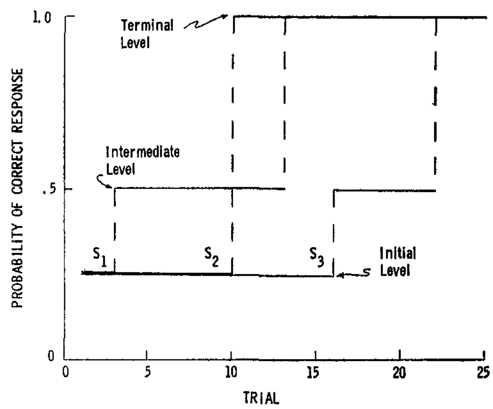

FIG. 8. Illustrative learning functions for individual subjects as conceived in three-state model.

is found to be relatively stationary between the first conditioned response and the last precriterion trial.

A similar observation had been made previously by Theios (1963), working with data from rats in an avoidance situation, and led him to consider the extension of the one-element model illustrated in Figure 8. Here each subject begins at some initial level of response probability and learning occurs by discrete jumps, but subjects may go to an intermediate level before going to the final level of perfect performance. Bower and Theios (1963) investigated the properties of this three-state model and found that predictions from it provided excel-

[3] This analysis was made possible by the cooperation of I. Gormezano, who provided protocols from a series of studies conducted in his laboratory. For the details of apparatus and procedure, see Moore and Gormezano (1961).

lent accounts of the data from the learning of rats in a shuttle box and also of the data on eyelid conditioning from which the preceding illustration was drawn.[4]

Now what is one to conclude from results such as these concerning "one-trial versus incremental learning"? The learning can scarcely be characterized as "one trial," and thus I suppose some will be happy simply to label it as "incremental" and let the matter rest. But if our interest is in understanding the behavior rather than in defending a position, we may wish to pursue the fact that this evidence points to the existence of all-or-none processes even in cases where the structure of the learning process is evidently more complex than that of simple paired-associate or verbal discrimination learning.

From even the brief sketch I have been able to give here, I think it must be evident that the investigators currently working constructively with notions of all-or-none learning have not formed a committee to set down a new theory. Rather they have been largely preoccupied with the always rewarding business of following up some unexpectedly fruitful research leads. At the same time a body of new theory has begun to take form, although only irregularly, and jagged in outline, much like the crystallization of rock candy in a saturated sugar solution. At some stage in this process one begins to raise questions of a broader character than those having to do with the ability of specific models to fit specific data. In particular, one may begin to wonder whether a body of theory based on all-or-none conceptions could ever be extended and elaborated enough to guide research on forms of learning more complex than conditioning and verbal association without becoming unwieldly and unmanageable.

Some recent researches suggest that, to the contrary, theory based on stimulus sampling and all-or-none association may offer some special virtues

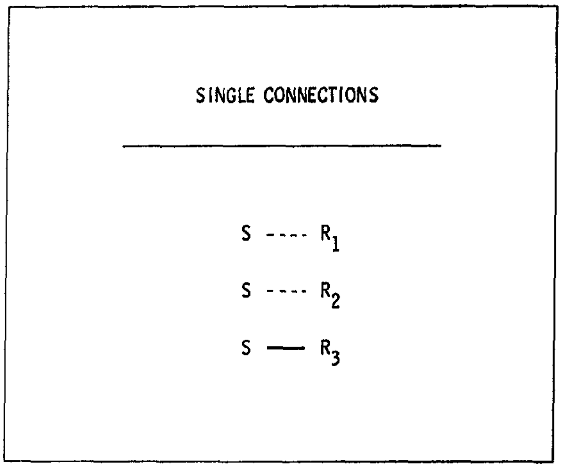

FIG. 9. Schema for stimulus-response associations according to the one-element pattern model. (Effective reinforcement of response R_3 to pattern S establishes an association—heavy line—supplanting the associations previously existing between S and R_1 or R_2.)

in this respect. The one-element process, abstracted from the original context, may constitute a conceptual pattern which can be identified by similar methods at many different levels of behavioral organization.

The first few published studies involving new techniques for revealing all-or-none acquisition have, understandably, dealt only with simple, unitary responses (usually spoken letters or digits) to discrete stimulus patterns. In Figure 9 is shown the paradigm for individual learned associations assumed in applications of the one-element pattern model at this elementary level of behavioral organization. A stimulus pattern S can be conditioned to one response at a time only. Thus, if first R_1, then R_2, then R_3 were successively reinforced in the presence of S, only the last would remain conditioned. On a test trial at the end of this training sequence, R_3 would be evoked, and no trace of the earlier learned experiences could be detected.

Considered outside the limited context of paired-associate experiments, this last property of the model is not very palatable. We know, for example, that an individual often can recognize a previously learned, correct response to a stimulus even when he can no longer supply the response when presented with the stimulus alone. An earlier suggestion on my part (Estes, 1960) that association theory should allow for the learning and unlearning of recognition and recall responses as

[4] The intermediate state of conditioned response probability has a natural interpretation in stimulus sampling theory (Atkinson & Estes, 1963) on the assumption that the stimulus situation associated with onset of the CS overlaps only in part with the stimulus situation existing at the point of reinforcement (onset of the US). Theios and Bower's analyses bear out the prediction that, although the organism has constant probability of leaving the initial, "unlearned," state on any reinforced trial, it can go from the intermediate to the terminal state only on a trial when the conditioned response occurs at CS onset.

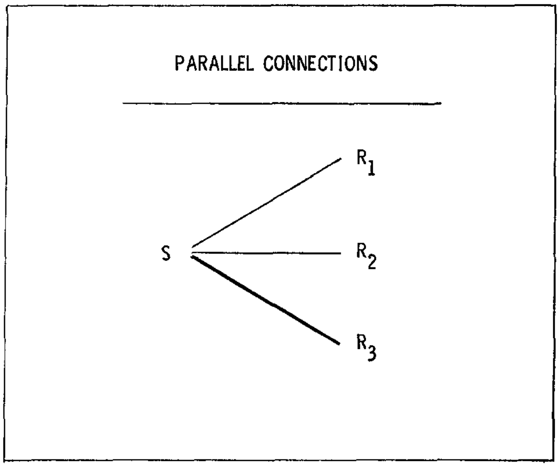

FIG. 10. Schema permitting coexisting associations (light lines) between a stimulus pattern, S, and successively reinforced responses, R_1, R_2, and R_3. (The tendency to recall, or retrieve, a previously learned association—represented by the heavy line—is assumed to be governed by an independent, one-element learning process.)

independent, even though often correlated, processes has met a chorus of complaints about lack of parsimony (see, e.g., Postman, 1963), parsimony being identified with the conception of a unitary underlying construct of associative strength. However, I believe that in this instance the lack of parsimony is in the organism, rather than in the theory.

Numerous considerations[5] suggest modification of the unitary-connection schema to the one illustrated in Figure 10 for the same sequence of events as that of Figure 9. Here the lighter lines represent associations established by earlier reinforcements of R_1 and R_2 (the one to the last reinforced response, R_3, being masked). These associations may be demonstrable by suitable procedures, e.g., recognition tests, even after R_3 has later been learned to the same stimulus. The heavy line represents an association between the stimulus and the response of recalling (selecting, emitting, retrieving) a particular one of the previously reinforced responses upon presentation of the stimulus. In this conception, the formation of each of the associations and the "switching" of the recall, or retrieval tendency, from one response to another are independent one-element processes. A currently active line of research in my laboratory is concerned with the experimental separation of these processes (DaPolito, Casseday, Kegel, McCollum, & Estes, 1961; Estes, 1964).

[5] Including recent work on information storage and retrieval in relation to short-term memory (Broadbent, 1958; Sternberg, 1963) and on nonassociative factors in recall (Asch & Ebenholtz, 1962).

Departing in another direction from the elementary schema of Figure 9, we encounter numerous possible levels of response complexity: The behavior segment which is counted as a single "response" in a learning experiment may be a single spoken digit, a multisyllable word in a new or artificial language, a combination of words, a rule or strategy. When a learning situation involves both the development of a relatively complex response unit and the association of this unit with particular cues or stimulus properties, one must expect that, regardless of the nature of the learning process, the probability of "correct responses" will change gradually over learning trials. However, some current researches are yielding encouraging signs of progress toward analyzing such complex forms of learning into simpler constituent processes. For example, the development of response compounds via the chaining of initially distinct component responses has been investigated by Crothers (1962, 1963) and the formation of chains of observing and instrumental responses by Atkinson (1961). In each of these instances, evidence has been forthcoming in support of the notion that the constituent associative processes conform to the one-element, all-or-none model. Once established, even complex response chains, rules, or strategies (Restle, 1962, 1963) may become associated with, or dissociated from, stimulus patterns on an all-or-none basis.

From the general tenor of this report, you might suspect that I think of *all* learning in terms of the pyramiding and branching of simple component all-or-none processes. This would not be far from the truth. But at the same time, I hasten to add, I have no intention of giving up such time tested theoretical devices as linear operator models (Bush & Mosteller, 1955; Estes & Suppes, 1959) even though these are based on different underlying conceptions. However much we might prefer things to be otherwise, learning, as measured by available techniques, appears sometimes to be an essentially continuous, sometimes a sharply discontinuous process. Thus in current research we need models suitable to represent both kinds of data, while we await further evidence to determine which aspect is the more fundamental and which the derivative.

There may turn out to be no basic inconsistencies between what at present appear to be extremely different mathematical models. This will, for example, be the case if we find that in situations involving numerous concurrent one-element processes, the overall functioning of the system should be expected to conform approximately to a continuous model.

While awaiting more definitive clues concerning underlying properties, I myself have come to operate on the working assumption that all instances of apparently incremental changes in behavioral dispositions during learning are simply cases of incomplete analysis. I am well aware that this assumption is untestable in a formal sense. Nonetheless, by following through its implications for a variety of learning situations, we are beginning to evolve quantitative techniques which may even come to rival the scalpel and the electrode as tools for exploring the organization of behavior.

REFERENCES

ASCH, S. E., & EBENHOLTZ, S. M. The process of free recall: Evidence for non-associative factors in acquisition and retention. *J. Psychol.*, 1962, **54**, 3–31.

ATKINSON, R. C. The observing response in discrimination learning. *J. exp. Psychol.*, 1961, **62**, 253–262.

ATKINSON, R. C., & ESTES, W. K. Stimulus sampling theory. In R. D. Luce, E. Galanter, & R. R. Bush (Eds.), *Handbook of mathematical psychology*. Vol. 2. New York: Wiley, 1963. Pp. 121–268.

BOWER, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, **26**, 255–280. (a)

BOWER, G. H. All-or-none theory applied to verbal discrimination learning. *Amer. Psychologist*, 1961, **6**, 466. (b) (Abstract)

BOWER, G. H. A model for response and training variables in paired-associate learning. *Psychol. Rev.*, 1962, **69**, 34–53.

BOWER, G. H., & THEIOS, J. A learning model for discrete performance levels. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1963. Pp. 1–31.

BROADBENT, D. E. *Perception and communication*. London: Pergamon Press, 1958.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.

BUSH, R. R., & MOSTELLER, F. A comparison of eight models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959. Pp. 293–307.

CROTHERS, E. J. Paired-associate learning with compound response. *J. verb. Learn. verb. Behav.*, 1962, **1**, 66–70.

CROTHERS, E. J. All-or-none paired-associate learning with compound responses. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1963. Pp. 95–115.

DAPOLITO, F., CASSEDAY, J., KEGEL, P., MCCOLLUM, K., & ESTES, W. K. Acquisition of paired-associates under differential instructions with recall and recognition tests. Paper read at Midwestern Psychological Association, Chicago, May 1961.

ESTES, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959. Pp. 9–52.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

ESTES, W. K. Information storage in behavior. In R. W. Gerard (Ed.), *Proceedings of symposium on information processing in the nervous system*. Amsterdam: Excerpta Medica Foundation, 1964, in press.

ESTES, W. K., HOPKINS, B. L., & CROTHERS, E. J. All-or-none and conservation effects in the learning and retention of paired-associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

ESTES, W. K., & SUPPES, P. Foundations of linear models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959. Pp. 137–179.

MOORE, J. W., & GORMEZANO, I. Yoked comparisons of instrumental and classical eyelid conditioning. *J. exp. Psychol.*, 1961, **62**, 552–559.

POSTMAN, L. One trial learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning*. New York: McGraw-Hill, 1963. Pp. 295–320.

RESTLE, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, **69**, 329–343.

RESTLE, F. Sources of difficulty in learning paired-associates. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1963. Pp. 116–172.

THEIOS, J. Simple conditioning as two-stage all-or-none learning. *Psychol. Rev.*, 1963, **70**, 403–417.

TRABASSO, T. R. Stimulus emphasis and all-or-none learning in concept identification. *J. exp. Psychol.*, 1963, **65**, 398–406.

STERNBERG, S. Retrieval from recent memory: Some reaction-time experiments and a search theory. Paper read at Psychonomic Society, Bryn Mawr, Pennsylvania, August 1963.

SUPPES, P., & GINSBERG, ROSE. A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning, with application to concept formation in children. *Psychol. Rev.*, 1963, **70**, 139–161.

UNDERWOOD, B. J., & KEPPEL, G. One-trial learning? *J. verb. Learn. verb. Behav.*, 1962, **1**, 1–13.