A THEORY OF THE SERIAL POSITION EFFECT

Edward A. Feigenbaum*
University of California, Berkeley

and

Herbert A. Simon*
Carnegie Institute of Technology

P-2375

July 18, 1961

*Consultant, The RAND Corporation

SUMMARY

Of major interest in the study of rote serial learning is the serial position effect. If one plots errors made at a syllable position vs syllable position, the resulting serial error curve, as it is called, will be seen to be bowed, with least errors occurring on syllables at the ends of the list, and most errors occurring on syllables near the middle of the list. In 1953 McCrary and Hunter observed that if percentage of total errors is taken as the unit of measurement, then all the empirical serial error curves for lists of a given length are substantially identical.

The paper presents an information processing theory of rote serial learning sufficient to predict (qualitatively and quantitatively) the shape of the serial error curve. The theory also explains other rote learning phenomena.

The theory postulates a serial information processing mechanism which learns (on the average) one item from a serial list every k seconds; has a very small immediate memory span; and utilizes an anchor-point processing strategy for organizing its learning effort over time.

Predictions from the postulates have been made in two ways. First, a computer has been programmed to process information in the manner prescribed by the theory, and this computer program has been used to simulate rote serial learning experiments. Second, the theory has been described by a simple mathematical model, from which the predictions were also generated.

A THEORY OF THE SERIAL POSITION EFFECT *

BY

Edward A. Feigenbaum
University of California, Berkeley

and

Herbert A. Simon
Carnegie Institute of Technology

Intraserial phenomena have been a major focus of interest in the study of serial learning. McGeoch (14) for example, devoted 50 pages of his Psychology of Human Learning to such phenomena. And among intraserial phenomena, one of the most prominent is the serial position curve, depicting the relative number of errors made with the various syllables in a list while learning the list to some criterion.

In 1953 McCrary and Hunter (13) observed that if percentage of total errors is taken as the unit of measurement, then all the empirical serial position curves for lists of a given length are substantially identical. Earlier investigators, measuring numbers of errors, had concluded that relatively more errors occurred for the middle syllables when the lists were hard than when they were easy, more with slow learners than with fast learners, more with rapid presentation of syllables than with slow presentation, and so on. One can find in the literature numerous theoretical explanations of these differences.

*We are indebted to our colleague Allen Newell for numerous helpful discussions about this project, and to the Ford Foundation for financial assistance that made it possible.

The findings of McCrary and Hunter leave us in the embarrassing position of having explained phenomena that don't exist, i.e., the supposed differences in amount of the position effect, and having failed to explain a striking uniformity that does exist--the substantial identity of curves derived under a variety of experimental conditions. McCrary and Hunter themselves reach the peculiar conclusion that a single principle can hardly be expected to account for uniformity of effect under diversity of conditions, hence that some multiple-factor is needed to explain the outcome.

The thesis of this paper is the opposite one--that if a uniformity underlies experiments performed under a wide variety of conditions, this uniformity should be traceable to a single simple mechanism that is invariant under change of conditions. We shall propose such a model of the information processing activity of a subject as he organizes his learning effort in a serial learning task. The serial position effect will be shown to be a consequence of the information processing strategy postulated in the model; the model predicts both qualitatively and quantitatively the shape of the curve and the percentages reported by McCrary and Hunter and by others.

Some Relevant Data

Before we state the theory of the serial position effect we shall review some important empirical findings on serial learning of nonsense syllables:

1. Under the usual experimental conditions and with experienced subjects there is generally a characteristic curvilinear relation between number of errors to criterion for a given syllable and the serial position of that syllable in the list. The syllable with the largest number of errors is generally beyond the middle of the list, though this effect becomes less noticeable as the length of list increases (20); the first syllable almost always exhibits the fewest errors (8).

2. McCrary and Hunter (13) have shown that, for lists of a given number of syllables, all serial position curves obtained with the usual experimental procedures are virtually identical when errors are plotted on a percentage, rather than an absolute, basis. About the same degree of bowing is exhibited with nonsense syllables as with names, with massed as with distributed practice, with slow as with fast learners, with rapid as with slow presentation. Typical data for lists of 12 and 14 syllables are given in Tables 1 and 2.

3. In spite of this uniformity under normal conditions, it is easy to produce large deviations from the characteristic curve. Such deviations can be produced in at least the following four ways: (a) by varying the difficulty of particular items in the list; (b) by introducing an item sharply distinguishable from those that precede it or follow it (14, p. 107); (c) by introducing distinguishable sublists within the main list (24); (d) by explicit instructions

to the subjects (12, 22). Not surprisingly, difficult items are learned with more errors than easy items in the same serial position; distinguishable items are learned with fewer errors; items that the subject is instructed to learn first are learned with fewer errors.

4. For lists of a given length, average learning time per syllable is almost independent of the rate of presentation*, and the order in which subjects are instructed to learn the items.** Hence, number of trials to criterion is inversely proportional to seconds per syllable.

5. Distribution of practice reduces the number of trials to criterion, but not sufficiently to compensate for the additional total time. The advantage of distribution, measured in trials to criterion, almost disappears when the presentation rate is as slow as four seconds per syllable (8).

In the next section we shall propose a theory of the serial learning process that accounts quantitatively for the data mentioned in items 1, 2, and 4, above, and qualitatively for the observations of item 3. In the present paper we shall

---

*Through this result is not given explicitly by Hovland (8), we have used his reported data to compute the average learning time per syllable. (This constancy has been [redacted] independently reported by Wilcoxon, Wilson and Wise. (23))

**Similarly, though this result is not given explicitly by Krueger (12), we have computed it from his data.

not discuss the effects of distribution of practice, since these effects almost certainly derive from mechanisms that go beyond the simple theory proposed here. We wish merely to observe that when time to criterion is taken as the measure of learning rate these effects are of rather small magnitude compared with those we shall consider.

An Information-Processing Theory of Serial Learning

We hypothesize that serial learning is an active, complex process involving the manipulation and storage of symbols by means of an interacting set of elementary information processes; and that these processes are qualitatively similar to those used in problem solving, concept formation, and other higher mental processes (17). Thus, we shall argue that the stimulus-response sequences postulated by S-R theory are simple only in surface appearance--that beneath them lies an iceberg of complex information processing activity.*

---

*We shall not defend this viewpoint in detail here although it has proved exceedingly fruitful in research in which we and our associates have been engaged. See, for example, (4, 16, and 17). We should like to offer three brief observations to persuade the reader that our conjecture does not entirely fly in the face of common sense or previous psychological observation. First, expectancy and mediation theories, like those of Tolman (7, p. 185-221) or of Osgood (7, p. 464-465), attribute as much complexity to the stimulus-response connection as does our conjecture; what they fail to indicate is the nature of the mechanisms that might provide the complexity. Second, equally elaborate and more explicit mechanisms are postulated in concept-formation theories like those of E. J. Gibson (6) and the recent one of Bruner, Goodnow, and Austin (2). Indeed, we shall see that one of our postulates involves a conception closely related to Bruner's notion of "cognitive

Underlying Assumptions of the Model

For the purposes of this paper, we shall not need to examine the elementary information processes in detail, for the shape of the serial position curve will prove to be independent of their micro-structure.* We require, instead, the assumption that in order for a connection between a stimulus and a response to be formed, a certain (unspecified) sequence of elementary processes needs to be carried out, and that the execution of this sequence requires a definite interval of time, the length of the interval depending on the "difficulty" of the task and other parameters of the experimental situation.

We suppose the information processing mechanism to be operating predominantly in a serial rather than a parallel manner--it is capable of doing only one, or a few things at a time. The narrowness of the span of attention is a familiar aspect of conscious activity; we assume that it is also an attribute of the subconscious.

---

strain." Third, the time an experienced subject needs, per syllable, to memorize a list of a dozen nonsense syllables is of the order of thirty seconds. In comparison with the times required by familiar electronic systems for simple processes, this is an enormous time interval. It is large--by a factor of 500 or more--even in comparison with the 50 milliseconds or thereabouts required for the central processes in the simplest responses to stimuli. If a theory is to fill up this thirty-second time interval in at all a plausible manner, it will have to attribute considerable complexity to the processes that take place. In (4) we report on such a theory of verbal learning, dealing in a complete manner with discrimination learning, association learning, responding, etc. The theory predicts a variety of the phenomena of rote learning of nonsense syllables in serial and paired-associate learning tasks.

*This point is examined in detail in (4), where a distinction is drawn between macroprocesses of verbal learning and micro-processes.

Information Processing Postulates

The structure of the theory is embodied in four postulates about the processing mechanism.

Postulate 1. Serial Mechanism. The central processing mechanism operates serially and is capable of doing only one thing at a time. Thus, if many things demand processing activity from the central processing mechanism, they must share the total processing time available. This means that the total time required to memorize a collection of items, when there is no interaction among them, will be the sum of the times of the individual items.*

Postulate 2. Average Unit Processing Time per Syllable. The fixation of an item on a serial list requires the execution of a sequence of information processes that, for a given set of experimental conditions, requires a definite amount of processing time per syllable. The time per syllable varies with the difficulty of the syllables, the length of list, the ability of the subject and other factors. In a well-known series of experiments by Hovland (8), for example, it averaged approximately 30 seconds.

---

*In serial learning of syllables there is, in fact, interaction among individual items; and total learning time increases more than proportionately with number of items. We will not be concerned with this point in the present paper because we are not dealing with total learning time or total errors, but with the relative number of errors made on different syllables in a list.

<u>Postulate 3</u>. <u>Immediate Memory</u>. There exists in the central processing mechanism an immediate memory of limited size capable of storing information temporarily; and all access to an item by the learning processes must be through the immediate memory. There is a great deal of experimental evidence to support the concept of an immediate memory. The evidence points to a span of immediate memory of about five or six symbols (15). We postulate that each symbol stored separately in the immediate memory must be a familiar, well-learned symbol. For unfamiliar nonsense syllable materials, the familiar symbols are the letters. Thus, for the three letter nonsense syllables ordinarily used, we postulate that the immediate memory has the capacity to hold two syllables (six letters). This means that it will ordinarily hold at any moment one S-R pair being learned.

<u>Postulate 4</u>. <u>Anchor Points</u>. In the absence of countervailing conditions--the nature of which will be specified presently--the information processing will be carried out in a relatively systematic and orderly way which will limit the demands that are placed on the small immediate memory. This postulate is related to the generalization, which Bruner and his associates (2) have tested in certain concept-forming experiments, that subjects develop strategies for limiting the "cognitive strain" involved in concept formation, and that these strategies involve handling newly acquired information in a systematic and orderly way.

We assume that subjects learning the syllables of a serial list will reduce the demands on memory by treating the ends of the list as "anchor points," and by learning the syllables in an orderly sequence, starting from these anchor points and working toward the middle. This procedure reduces demands on memory because, at each stage of the learning task, the next syllable to be learned is readily identified as being adjacent to (just before or after) a syllable that has already been learned. Thus, no special information about position in list needs to be remembered.*

The first three postulates differ from the fourth in that the former describe built-in characteristics of the processing mechanism that are probably not learned or readily modified; while the latter describes a method of proceeding that is apparently habitual with most subjects, at least in our culture, but which is modifiable by experimental instructions, and by certain attention-directing stimuli.

It has been observed frequently that in serial memorization subjects not only develop associations between syllables, but also use various position cues and other cues. They learn, for example, that a particular syllable occurs in the early or

---

*The idea of learning from anchor points is not new, though it does not seem to have been previously formalized. Woodworth (25), for example, makes use of it in describing the process by which a list of digits is learned. Wishner, et al., (24) mention it in their discussion of the serial position curves obtained in their list-sublist experiment.

in the late part of the list.* The use of position cues gives a unique status to the beginning and end of a serial list, for these items have the special property that they have no neighbors "before" and "after" respectively, i.e., the first item is always preceded, and the last item succeeded, by the intertrial inactivity. Once the items at the anchor points are memorized, the items contiguous to them become the first unlearned items "after" or "before" syllables already memorized; and so on, as the learning proceeds. More than this, the first two items are unique in that they represent the first S-R pair presented to the subject in the experiment. Thus, we can make out at least a plausible case that a learner can reduce the demands on immediate memory by memorizing in a more or less systematic fashion from the ends of the list toward the middle.

This postulate is sufficient to explain the bowed form of the serial position curve--although it says only a little more than the observed fact of the bowing. Its advantage over explanations like the Lepley-Hull hypothesis (which will be discussed later) is that it is not inconsistent with the ease (item 3, above) with which changes can be induced by the experimenter in the serial position curve.

In order to make quantitative predictions as to the amount of bowing that will be observed, we use the notion of anchor

---

*From this reliance on "irrelevant" cues, one can develop an explanation for such phenomena as anticipatory errors and "remote associations," that is much simpler than the usual one; but these topics would take us beyond the scope of our present task.

points to strengthen postulate 4, as follows:

Postulate 4a. Processing Sequence. We postulate the following information processing strategy for organizing the serial learning task using anchor points: (a) the first two items presented in the experiment are learned first; (b) attention is next focused for learning on an item immediately adjacent to an anchor point. (In the ordinary serial list, this will be the third item or the last.) The probability that any specific item adjacent to an anchor point will be selected for learning next is $1/p$, where p is the number of anchor points. (in the ordinary serial list, $p=2$. Thus, for example, the probability that the last item will be learned after the second item is .5); (c) attention is focused, and learning proceeds, item by item in this orderly fashion until the criterion trial is completed.

One can picture the subject building up over time an internal representation of the serial list he is learning. It will be seen, then, that the postulate specifies only a minimal amount of organizational activity: namely, the ability to add an item immediately after or immediately before an item already learned (or a "special" stimulus like the intertrial interval). Our explorations with other processing strategies have shown that this strategy reduces greatly the information processing demands on the learner.

Predictions from the Postulates

The postulates describe a learning mechanism that memorizes serial lists in a prescribed way. This mechanism generates a

serial error curve as it learns (i.e., some serial error curve is a consequence of the postulates). We wish to compute this serial error curve and compare it with the McCrary and Hunter curve.

Computer simulation is the most general and powerful method for doing this. We and others have used this method extensively in building theories of problem solving (18), binary choice behavior (5), concept attainment (10), and other cognitive phenomena. It is described in detail elsewhere (16). Briefly the idea is this. The digital computer is a universal information processing device, capable of carrying out any precisely specified information process. Thus a computer can carry out exactly the information processing required by the postulates of the model. We program the model on a computer, use it *qua* subject in verbal learning experiments (simulated inside the computer), observe the learning behavior of the model, and thereby generate the consequences of the postulates in particular verbal learning situations. We have used this method in constructing and exploring an information processing theory of verbal learning.* In particular, for the purposes of this paper, we have generated the serial error curve for a few simple serial learning experiments. We have done this in two different ways: first, following postulate 2, we introduced a unit processing time

---

*A discussion of the complete theory is not appropriate here. It is given in (4). The computer program which simulates verbal learning processes is called EPAM (Elementary Perceiver and Memorizer), and it has been tested in a variety of verbal learning experiments.

per syllable without specifying the microprocesses of the learning that takes place during this time interval; second, we removed the latter artificiality and substituted instead the full complement of microprocesses postulated by the more complete theory.

For the particular case of the serial position curve, the postulates are simple enough so that there is no real need to employ computer simulation to generate the predictions. The postulates can be formalized in a simple mathematical model, from which the quantitative predictions can be generated. As this method is likely to be more familiar to the reader, we give the mathematical model in the Appendix, and present the serial error curves which it predicts (for lists of twelve and fourteen syllables) in Tables 1 and 2. The results obtained by the computer simulation technique are substantially identical (though slightly more discontinuous).

What can we say specifically about the fit? First, the ordinates of the first and last syllable of the predicted curves are in almost exact agreement with those of the empirical curves. Second, the syllable position of the peak of the predicted curves is substantially the same as that of the observed curves. Third, the ordinates of the predicted curves and the empirical curves at each syllable position are

Table 1

Table showing percentage of total errors made during acquisition at each syllable position of a 14 item serial list of nonsense syllables predicted and observed

Syllable Position

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Predicted	.95	1.9	4.7	5.6	8.1	8.9	10.5	10.8	10.8	10.5	8.9	8.1	5.6	4.7
Observed [a]	1.0	3.5	4.3	6.0	8.0	8.9	9.2	10.0	9.5	10.6	8.8	8.9	7.2	4.0

[a] These values are approximate

The data were taken from Figure 4 of McCrary and Hunter (13, p. 133)

Table 2

Table showing percentage of total errors made during acquisition at each syllable position of a 12 item serial list of nonsense syllables predicted and observed

Syllable Position

	1	2	3	4	5	6	7	8	9	10	11	12
Predicted	1.3	2.6	6.3	7.5	10.5	11.3	12.5	12.5	11.3	10.5	7.5	6.3
Observed [b]	1.5	2.3	4.5	7.0	10.3	11.6	14.0	12.4	11.0	10.0	8.5	7.0

[b] These values are approximate values for the median percentages at each position for the four curves presented in Figure 2 of McCrary and Hunter (13, p. 132)

very close, especially in the critical first and last third of each list, where very good agreement is important to any claims about goodness-of-fit. Furthermore, this fit was obtained without any arbitrary parameters, other than the specification of the sequence in which incoming syllables are processed.*

Elaboration and Discussion

In this section, we wish to compare the predictions of our information processing theory with those derived from the Lepley-Hull Hypothesis, and extend our predictions to two important experiments, one of which was published after we had specified our model.

1. The Lepley-Hull Hypothesis

There have been few attempts to account for the serial position effect in quantitative terms. Hull attempted to do so (11), on the assumption of some inhibitory processes, or intra-list "interference." Atkinson (1), drawing on statistical learning theory, has recently exhibited a stochastic process which generates a curve of the general shape of the serial position curve. We shall discuss Hull's results in some detail, and comment briefly on Atkinson's.

Although Hull's equations provide a good fit to the empirical data, this fit is not a convincing test of the

---

*The goodness-of-fit of the observed frequency distribution to the predicted distribution was tested by the Kolmogorov-Smirnov test of association, a non-parametric test. The test accepted the null hypothesis at the 99 per cent level of significance.

Lepley-Hull theory for the following reasons:

Hull's theory leads to a set of equations having three free parameters; the reaction threshold, the ratio of inhibitory potential to excitatory potential per trial, and the remoteness reduction factor. These are used to fit the serial error curve, or, more precisely, the curve of number of repetitions to reaction threshold. See (11), pp. 103-107. Hull fits the theoretical curve by passing it through three points of the empirical curve. Since the empirical data form a relatively smooth, bow-shaped curve, it is not surprising that a three-parameter curve can be made to fit them closely; an equally good approximation can be obtained by fitting a parabola empirically to the data.

This means that Hull's hypothesis will fit almost any data (provided the serial position curve has the characteristic bowed pattern), and hence is almost impossible to disprove from the data. It is therefore an exceedingly weak hypothesis. By the same token--because of the three free parameters--Hull's theory does not predict the constancy on a percentage scale observed by McCrary and Hunter.

Conversely, given the constancy observed by McCrary and Hunter, we can draw certain conclusions from Hull's theory regarding the growth of excitatory and inhibitory potential and the reaction threshold. For example, it can be shown by an examination of Hull's equations that the ratio of the increment of inhibitory potential per trial to the increment of

excitatory potential must be a constant (independent, for example, of intra-list similarity).* This is surprising and contrary to the whole spirit of the Lepley-Hull hypothesis. For we would expect that with high intra-list similarity the inhibitory potential would rise more rapidly than with low intra-list similarity. On the contrary, if Hull's model is correct the only parameter that changes as lists become more difficult to learn is the ratio of the threshold to the increment of excitatory potential per trial.** Finally, the Lepley-Hull hypothesis does not explain how a subject can voluntarily or through a shift in his attention greatly alter the shape of the curve.

There are four reasons, therefore, why Hull's mathematical model for the serial position effect is unsatisfactory; since it contains three adjustable parameters its predictions are very weak; it does not predict the constancy in the percentage-error curve; this constancy hardly seems compatible with the mechanism assumed as a basis for the model; and finally, the model is difficult to reconcile with well-known attention-shift and set-change phenomena.

---

*See (11), pp. 104-105. The McCrary and Hunter result implies that the Rs are related: they are proportional to each other. The variables q_i are homogeneous of degree two in R, and the D_i are homogeneous of degree one in R. The Js are homogenous of degree zero in R. By equation (3) of page 104, $\Delta e/\Delta k$ is homogenous of degree zero in R, hence a constant.

**By equation 2, p. 104, of (11).

The preceding discussion of the curve-fitting aspects of Hull's equations applies also to Atkinson's equations. Atkinson has available four free parameters. He estimates these parameters from data for an 18 syllable list, and uses these estimated values to make predictions for lists of 8 and 13 syllables. However, a careful examination of Atkinson's equations shows that even after the parameters have been estimated, there are enough degrees of freedom left in the system almost to insure a reasonably good fit to the other curves. Thus, his theory suffers the same infirmity that we have pointed out in Hull's.

Furthermore, to make workable the difficult mathematics of the stochastic process, Atkinson has had to introduce a number of very constraining assumptions. His equations hold only for serial lists of highly dissimilar words which are familiar and easily pronounced; the presentation must be at a moderate rate, with a long interval provided at the conclusion of each trial. Yet the same bowed curve is obtained empirically when these conditions are not met as when they are. Finally, as in Hull's theory, there is difficulty in predicting the constancy observed by McCrary and Hunter. Given a set of parameter values, Atkinson's theory does not predict the constancy over the experimental conditions reported by McCrary and Hunter. On the other hand, if one admits that these values may change from situation to situation, then he must reestimate them for each experimental condition, and therein the theory loses much of its power.

2. <u>The Two-part List</u>

A simple extension of ordinary nonsense syllable experiments is to differentiate the first half of the list from the second half by printing the former in one color (say black) and the latter in another color (say red), dividing the total list perceptually into two smaller sub-lists.

What will be the shape of the serial position curve? One can predict this from Hull's theory by the unsatisfying procedure of assuming that the total list is learned as two separate sub-lists, and by fitting each sublist with a three-parameter curve such as we have discussed previously. The total fit will have six free parameters.

We should like to be able to predict the shape of the curve from the theory we have already presented. There are two important issues involved in making such a prediction:

1) Consider those subjects who perceive the total list as being constructed of two sublists. One possible reasonable strategy for dealing with the learning task is to use the end points of each sublist as anchor points, in the type of learning process described earlier. Another plausible strategy is to use as anchor points the ends of the total list and one point in the center to identify the point of bifurcation, say the first red syllable.

In making a prediction of the serial position curve for the two-part list, we have assumed simply that of those subjects who perceived the list as being two sublists, one half used the first strategy and the other half used the second strategy.

The assumption is, of course, a relatively crude one, but the prediction is not very sensitive to the actual percentages assumed. Alternatively, we could have estimated the percentage from the data.

2) In the experiment which we shall discuss shortly, we have no way of knowing precisely how many subjects perceived the task as one of memorizing two sublists and how many perceived it as learning one long serial list. In the absence of this knowledge, we can estimate these percentages from the observed ordinate of the first red syllable and weight our prediction at each syllable position appropriately using this estimate. This procedure essentially "fits" our predicted curve to the empirical curve at one point, the "break" between black and red syllables. But it guarantees nothing about the quality of the prediction at the other points.

Recently, as part of a larger experiment, Wishner, Shipley, and Hurvich (24) performed an experiment with a two-part list. They had an experimental group memorize lists of fourteen syllables, half of which were printed in black capitals, the other half in red lower case letters. The experimental group was told that the object of the experiment was to discover how people learned two lists simultaneously.

In Table 3, the predicted values for the percentage of errors at each syllable position are compared with the

observed values.* The agreement at all syllable positions is very close.** As contrasted with our other predictions, in this one we had available the one free parameter already mentioned.

3. The Experiments of Krueger

We turn now to some important experiments, the results of which were alluded to previously, and which have important implications for the information-processing theory.

In a well-known series of experiments, Krueger (12) presented various kinds of lists of "easy" and "hard" paired nouns to subjects who either received instructions to learn

---

*The predictions were generated by the methods previously described. Three anchor points were used. In the mathematical treatment the three-anchor-point predictions were corrected by a factor which insured an exact fit of the ordinate of syllable 9 (the middle anchor point). In the computer simulation method, whole-list and sub-list strategies were both run and the predicted ordinates averaged in a weighted fashion such that the ordinate of syllable 9 fit exactly. What this procedure comes down to is the assumption that approximately two thirds of the subjects learned the list as two sub-lists and approximately one third learned it as a single long list (note that these fractions were not assumed *a priori* but were obtained by working backwards from the observed ordinate of the ninth syllable).

**The goodness-of-fit of the observed frequency distribution to the predicted distribution was tested by the Kolmogorov-Smirnov test of association, a non-parametric test. The test accepted the null hypothesis at the 99 per cent level of significance.

Table 3

Table showing percentage of total errors, predicted and observed, made at each syllable position during acquisition, for a two-color, 14-item serial list of nonsense syllables

	Syllable Position													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Predicted	1.9	3.1	4.7	6.7	9.1	10.1	8.9	5.9	6.0	8.3	8.7	9.8	9.9	5.5
Observed [c]	2.2	4.3	6.5	7.3	8.2	8.5	8.1	6.0	6.0	9.0	9.5	9.9	8.5	6.0

[c] These values are approximate, and were derived from data taken from Wishner, et al., (24, p. 260)

the list in some specified order or no instructions at all. These studies demonstrated that the order in which the various items were learned was influenced markedly by the instructions given to the subjects. As McGeoch puts it (14, p. 102), "The relation between rate of learning and position in the series is, then, a function of the direction of the subject's effort or attention." As we have indicated, this is entirely consistent with the information processing theory, which regards particular learning sequences as "strategies" for dealing with the learning problem--as adaptive response to task.

What about our more specific hypothesis that in the usual serial learning experiment the "end points" of the list will be taken as anchor points in the learning process? Krueger's experiments showed that subjects given <u>no</u> instructions produced essentially the same serial position curves as those subjects who were instructed to learn the <u>ends</u> of the list first.

Because the fixation of an item requires a fixed amount of processing time, and because the sequence of learning is considered a "strategy" and not a built-in characteristic of the learning process, our theory predicts that the total number of syllables learned will be proportional to the total learning time and independent of the order of learning. On this point, Krueger reports, "When the attention given is constant, the total amounts learned are the same, irrespective

of whether this effort is directed to the beginning, center, or the final sections of the unit which is to be memorized." (12, p. 527)

Although we assume that there is a constant fixation time associated with each particular item on a list, items of different kinds (e.g., "easy" items vs "difficult" items) will have different processing times. The theory we have proposed predicts that the total learning time will be the sum of these processing times per syllable, and as such will be independent of the order in which the syllables are learned. Confirming this prediction, Krueger reports, "When materials of unequal difficulty appear within the same unit to be mastered, the total number of trials required to memorize the unit is approximately the same whether attention is given at first to the more difficult or the easier sections of the unit." (12, p. 527)*

Thus, Krueger's experiments, though they were performed with serial lists of paired-nouns rather than nonsense syllables, demonstrate a) that the serial position curve can be "shaped" by the experimenter with suitable instructions to subjects, so that the order of learning syllables is itself a learned

---

*If one plots the McCrary and Hunter curve by ordering the abscissa values not by serial position in the list but by the apparent order in which the syllables were learned, the ordinates lie on a straight line. This, of course, is in exact agreement with our model. Recent additional information of this phenomenon was obtained by Jensen at the University of California, Berkeley, for the learning of nonsense figures (personal communication).

response; b) that the total amount of material learned within a given time is independent of the order in which various items, sometimes heterogeneous with respect to difficulty, are learned.

Some Recent Results

Subsequent to the specification of the model proposed in this paper, some important new experiments have been published on the effect of replacing syllables on a list during learning. Rock (21) used the following procedure on his experimental groups: on each trial, those syllables incorrectly responded to by the subject were removed and new syllables were substituted in their place for the next trial. Rock found no impairment of the rate of learning for the experimental groups (as compared with the control groups). This important result casts further doubt on Hull's incremental buildup hypothesis. Criticism of Rock's technique led Estes, Hopkins, and Crothers (3) to replicate and extend Rock's experiment, but their results substantiate Rock's.

Estes, Hopkins, and Crothers say of these experiments, "No hitherto published theory with which we are familiar gives a reasonable account of our principal findings." (p. 338) The information processing theory we have proposed here predicts the Rock result. A computer simulation of the Rock Experiment using the information processing model generated behavior substantially identical with that reported by Rock. In terms of our theory (postulate 4a) the explanation, of course, is

that items on a list are learned one at a time in the processing sequence. Items presented when attention is focused on some other particular item are simply ignored by the learner, and are picked up on a later trial, as determined by the processing sequence. Hence, no time is lost by the learner if the experimenter replaces an item that has not yet been processed.

Concluding Remarks

In this paper we have surveyed the principal known facts about the shape of the serial position curve in serial learning of nonsense syllables by the anticipation method. We have examined the Lepley-Hull hypothesis as an explanation for the shape of the curve, and have concluded that the hypothesis is unsatisfactory. We have proposed an alternative hypothesis formulated in terms of information processes. We have shown that the hypothesis not only predicts the constancy of the serial position curve when the ordinates are plotted in percentage terms, but also predicts the quantitative values of the ordinates. Since the hypothesis allows no free parameters, its success in fitting the observed data provides rather persuasive evidence for its validity.

The information processing hypothesis is built on the following assumptions:

1) that the brain is a serial processing mechanism with a limited span of processing attention;

2) that the fixation of an item uses up a definite amount of processing time;

3) that there is a small immediate memory which holds information to be processed;

4) that the subject employs a relatively orderly and systematic method for organizing the learning task, using items with features of uniqueness as anchor points.

In this paper, we have offered no explanation of the fixation process itself, i.e., we have talked not at all about what occurs during the processing time assumed in item (2) above.

APPENDIX

Given the unit processing time per syllable, one can, from the postulates, compute the average time after the beginning of a learning experiment that will pass before any specified syllable is learned. This time, in turn, determines uniquely the number of errors that will be made with that syllable. While the actual number of errors will be a function of the unit processing time, the percentage that this number represents of the expected total errors, is independent of the unit processing time.

The numerical estimates given in the text tables were obtained as follows: By the postulates, the syllables will be learned in an orderly sequence, each syllable requiring a certain processing time, say $\underline{k}$. Each syllable can be identified by its serial order, $\underline{i}$, in the list as presented by the experimenter, and also by the order, $\underline{r}$, in which it is learned by the subject. Since learning takes place from both ends of the list, these two orders will not, in general, be identical. Thus, s_i, the i^{th} syllable in order of presentation, may be the same syllable as s'_r, the r^{th} syllable in order of learning. (Technically, the list of syllables in order of learning is a permutation of the original list.) Let T'_r be the time that elapses before the first successful response to syllable s'_r--that is, until the r^{th} syllable is learned. Then $T'_r=kr$. The number of errors, W'_r, the subject will make on the r^{th} syllable is equal to the number of learning trials prior to the trial on which

that syllable is learned and will be proportional to r:

(1) $W'_r = mr$, where m is a proportionality constant, equal to k divided by the time per trial.

The numerical value of m, is a function, of course, of the difficulty of the items, and the rate of presentation. However, we are only concerned with the fraction of total errors made on a given syllable, and this fraction is clearly independent of m. For let W be the total number of errors, summed over all syllables, and $w'_r = W'_r/W$, the fraction of total errors made on the r^{th} syllable.

(2) $$w'_r = mr/\sum_{r=1}^{n} mr = r/\sum_{r=1}^{n} r = \frac{2r}{n(n+1)}$$, where n is the number of syllables in the list.

Suppose, for example, that we are dealing with a list of twelve syllables, then $\sum_{r=1}^{12} r = 78$. Hence, the fraction of total errors that will be made on the 4th syllable learned will be $4/78 = .051$.

Now, to obtain the serial position curve, we need merely to relabel the syllables from the order in which they are learned to their order of presentation. That is, if r_i is the rank, in order of learning, of the i^{th} syllable in order of presentation, then the fraction of total errors for the i^{th} syllable will be simply:

(3) $$w_i = w'_{r_i} = r_i/\sum_{r} r_i$$

To apply this result, we must calculate the rank, r_i, of the i^{th} syllable in order of presentation, as determined by

Postulate 4a. We assume (Postulate 3) that the immediate memory capacity is two syllables and for simplicity in the calculation shall assume that the items are picked up pairwise in the processing sequence, and stored in the immediate memory for learning. The first two syllables in the list will be learned first (will have rank, r=1 and r=2, respectively), followed either by the last two syllables on the list, or by syllables three and four, each with probability one-half. Then, subsequent pairs will be chosen from the beginning and end of the list of those not yet learned, always with probability one-half. The result is that in a list of twelve syllables the third syllable, for example, will have a probability of one-half of being the 3rd syllable in order of learning, a probability of one-quarter of being the 5th syllable, a probability of one-eighth of being 7th, and of one-sixteenth of being 9th or 11th.* Averaging these ranks, weighted by their respective probabilities, we find that the average rank of the 3rd is $r_3 = 4.875$. The fraction of total errors on the 3rd syllable will then be $w_3 = 4.875/78 = .063$. (See Table 2). All the other predicted values in the tables were computed in the same way.

---

*The fact that the 3rd syllable has zero probability of being learned fourth, sixth, eighth, etc., is artificially introduced by the calculation simplication introduced above of handling the syllables in pairs. It does not materially affect the serial position curve prediction, as is shown by the fact that the computer simulation (which does not use the pairwise learning simplification) generated the same serial position curve prediction.

REFERENCES

1. Atkinson, [redacted] A Stochastic Model for Rote Serial Learning," Psychometrika, 22, 1957, 87-95.

2. Bruner, J. S., Goodnow, J. J., and Austin, G. A., A Study of Thinking, New York; Wiley, 1956.

3. Estes, W. K., Hopkins, B. L., and E. J. Crothers,"All-Or None and Conservation Effects in the Learning and Retention of Paired Associates,"Jour. Exp. Psychol., 60, 6, 329-39.

4. Feigenbaum, E., An Information Processing Theory of Verbal Learning, The RAND Corporation Paper, P-1817, Oct., 1959.

5. Feldman, J., Analysis of Behavior in Two Choice Situations, unpublished doctoral dissertation, Carnegie Institute of Technology, 1959.

6. Gibson, E. J., "A Systematic Application of the Concepts of Generalization and Differentiation to Verbal Learning," Psychol. Rev., 47, 1940, 196-229.

7. Hilgard, E. R., Theories of Learning (2nd ed.). New York: Appelton-Century-Crofts, 1956.

8. Hovland, C. I., "Experimental Studies in Rote Learning Theory, III, Distribution of Practice with Varying Speeds of Syllable Presentation," J. Exp. Psychol., 23, 1938, 172-190.

9. Hovland, C. I., "Human Learning and Retention," ed. S. S. Stevens, Handbook of Experimental Psychology, New York: Wiley, 1951.

10. Hovland, C. I. and E. B. Hunt, "The Computer Simulation of Concept Attainment," Behavioral Science, 5, 1960, 265-267.

11. Hull, C. L., Hovland, C. I., et. al., Mathematico-deductive Theory of Rote Learning, New Haven: Yale Univ. Press, 1940.

12. Krueger, W. C. F., "Learning During Directed Attention," J. Exp. Psychol., 15, 1932, 517-527.

13. McCrary, J. W. and Hunter, W. S., "Serial Position Curves in Verbal Learning," Science, 117, 1953, 131-134.

14. McGeoch, J. A., The Psychology of Human Learning, New York: Longmans, Green, 1942.

15. Miller, G. A., "The Magical Number Seven, Plus or Minus Two; Some Limits on Our Capacity for Processing Information," Psychol. Rev., 63, 1956, 81-97.

16. Newell, A., and H. A. Simon, The Simulation of Human Thought, Current Trends in Psychological Theory, Pittsburgh: Univ. of Pittsburgh Press, 1961, 152-179.

17. Newell, A., Shaw, J. C., and Simon, H. A., The Elements of a Theory of Human Problem Solving, Psych. Rev., 65, May, 1958, 151-156.

18. Newell, A., Shaw, J. C., and H. A. Simon, The Processes of Creative Thinking, The RAND Corporation Paper, P-1320, September 1958.

19. Osgood, C. E., Method and Theory in Experimental Psychology, New York: Oxford Univ. Press, 1953.

20. Ribback, A., and B. V. Underwood, "An Empirical Explanation of the Skewness of the Bowed Serial Position Curve," Jour. Exp. Psych., 40, 1950, 329-335.

21. Rock, I., "The Role of Repetition in Associative Learning," Amer. J. Psychol., 70, 1957, 186-93.

22. Welch, G. B., and Burnett, "Is Primacy a Factor in Association-Formation?," Amer. Jour. Psych., 35, 1924, 396-401.

23. Wilcoxon, H. C., W. R. Wilson, and D. A. Wise, "Paired-Associate Learning as a Function of Percentage of Occurrence of Response Numbers and Other Factors," J. Exp. Psychol., 61, April, 1961, 284-289.

24. Wishner, J., Shipley, T. E., Jr., and Jurvich, M. S., "The Serial Position Curve as a Function of Organization," Amer. J. Psychol., 70, 1957, 258-262.

25. Woodworth, R. S., Experimental Psychology, New York: Holt, 1938.