PRESENTATION ORDERS FOR ITEMS

FROM DIFFERENT CATEGORIES

by

Edward J. Crothers

TECHNICAL REPORT NO. 71

March 10, 1965

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

PRESENTATION ORDERS FOR ITEMS
FROM DIFFERENT CATEGORIES

Edward J. Crothers
Stanford University

Abstract

We conclude that under quite general learning assumptions certain statements can be made regarding the theoretical efficiency of two radically different orders of presenting dependent items. In one schedule (massed), all items from the same category are presented successively, followed by all members of the next category. By contrast, the other schedule (distributed) takes one exemplar of the first category, then one of the second, etc., recycling through the categories in the same order after one representative of each has been presented. The most interesting finding is for the case where both schedules entail one reinforcement per item, followed by one reinforcement of each other item in the list, followed by the test on the original item. Given this routine of reinforced test trials, we prove the parameter-free result that the predicted mean test proportion correct is higher if the members of a category are massed rather than systematically distributed among exemplars of other categories. The

result holds regardless of whether proactive facilitation is admitted. Contrariwise, for the case of nonreinforced test trials, we have been unable to discover general conditions regarding the relative efficiency of the two schedules. Numerical computations were offered to supplement the algebraic results.

I. Introduction

This paper is the third in a series of theoretical comparisons of test performance after different presentation orders during training. The general approach has been to construct systematic training sequences, each followed by the identical test. We apply a learning model to predict which sequence will yield the maximum proportion of correct responses on the test. The motivation for examining this problem is twofold. First, many interesting questions can be formulated that generate strong tests of the learning models. In terms of predicted test proportions correct, ordinal relationships among the sequences can often be established independently of the values of the model's parameters. Secondly, the proposed presentation orders are orderly, plausible training sequences, not arbitrarily contrived permutations. Hence the model's predictions have potential applicability. Later this practical relevance will be clarified when the models are illustrated.

One investigation (Crothers, 1965) involved two types of trials - rule presentations and examples illustrating the rules. Assuming a fixed number of trials of each type, several learning models were employed to rank-order the sequences on the basis of predicted proportion correct on the terminal test. According to a variety of models, the theoretically optimal sequence consists of all trials of one type, and then all those of the other type. That is, if a function of the parameters exceeded zero the optimal sequence was all presentations of the rule, then all examples. Otherwise the reverse order was most efficient, but under no circumstances could interspersing rules with examples be optimal.

The second investigation (Suppes, 1964) provided the impetus for the present paper. He examined the case of m independent items, each presented n times. An example would be paired-associates, selected so as to be learned independently of one another. The independent variable of interest was <u>block-size</u>, defined as one plus the number of other items intervening between successive presentations of the same item. For example, take a list where m = 4 items, and for specificity suppose that each gets two reinforcements. Then the schedule $i_1\ i_1\ i_2\ i_2\ i_3\ i_3\ i_4\ i_4$ has block-size one. Similarly, schedules $i_1\ i_2\ i_1\ i_2\ i_3\ i_4\ i_3\ i_4$ and $i_1\ i_2\ i_3\ i_4\ i_1\ i_2\ i_3\ i_4$ exhibit sizes two and four, respectively. In general, each divisor of m is an admissible block-size. Following the mn reinforced trials, each item received one unreinforced test trial. Assuming a mean learning curve characteristic of a one-element or single-operator linear model, Suppes proved that predicted mean test performance is a monotonic function of block-size.

A key assumption in his development is that the items are learned independently of one another. Of course, the tenabality of this assumption depends heavily on the particular set of items being presented. But intuition suggests that instructional material usually consists of items that are designed to teach the same skill, and hence are highly interdependent. For example, a child acquires facility in multiplication by exercises with different numerals. Again, one may learn to hear the difference between a Russian voiced and a voiceless consonant by a procedure wherein the particular consonants and adjoining vowels differ from one item to the next. Since the distinctive feature of voicing characterizes all contrasts, the items would not be learned independently.

In cases such as these where dependent items comprise the list, the prediction of the block-size model does not appear wholly realistic. To accent the issue more sharply, consider a paired-associates task in which the subject learns the Russian equivalents of the numbers from 1 to 5, and also the translations of five animal names. Intuitively, it might well be more efficient to present the numbers in succession, then the animal names, instead of interspersing numbers with names. More generally, the topical ordering of instructional material reflects the belief that one task should be mastered before another is undertaken. On the contrary, topical ordering is not recommended by the block-size model. Since the model does not include dependent items, it deals only with <u>how many</u> other items should intervene between successive reinforcements of the same item - nothing is said about <u>what</u> they should be. The present issue is: should an item be surrounded by other members of the same category or by items from other categories?

How can a model be formulated to capture the notion of item dependency? One treatment would be to postulate a concept-learning process. (Bower and Trabasso, 1964; Suppes and Ginsberg, 1963). Each exemplar would be viewed as affording an opportunity for the concept to be mastered; after one concept was mastered, all its exemplars would be responded to correctly. But for phoneme discrimination, acquisition of Russian vocabulary, etc. this model cannot be empirically adequate. A less extreme hypothesis is that by virtue of their unique properties the members of a category cannot be learned simply by mastering a single concept. Rather, we maintain that when an item is reinforced, there is positive transfer to other instances of the same category. If the transfer were 100%, we would have the concept-learning version. We also adopt a forgetting axiom analogous to that incorporated by Suppes. The s-r association for a particular exemplar may be forgotten when exemplars of other categories are presented.

In the interest of mathematical simplicity, most of the findings to be reported involve the restriction that an item be followed by at most one reinforcement of every other item in the list. If parameter-free predictions cannot be made here, then _a fortiori_ they cannot be made when the list is repeated a number of times. At one pole we have a schedule to be called _massed_ exemplars, and at the opposite pole we have _distributed_ exemplars. A massed sequence has all presentations of one category, then all of another category, etc. A distributed order begins with a subsequence containing one member of each category in succession, then a second member of each category, with the same category order as before. These two schedules are extremes in the sense that the number of other categories intervening between consecutive reinforcements of the same category is zero in the massed schedule and a maximum in the distributed schedule.

Parallel derivations for the two schedules appear in the next section. Two sets of derivations are carried out, depending on whether or not responses on the criterion test are reinforced. The salient finding is that when the second (test) cycle through the list is reinforced and replicates the first cycle presentation order, then massing of exemplars is more efficient than distributing them. This prediction is independent of the model's parameters, and the numbers of categories and exemplars. No general result has been found for the method of nonreinforced test trials, although some insight is furnished by numerical computations.

II. Derivations

1. Notation

A particular item is denoted as a_{ij}, where i and j index the category and exemplar, respectively ($1 \leqq i \leqq C$, $1 \leqq j \leqq E$). For each item a_{ij} the probability of a correct response can be transformed by three path-independent operators: the acquisition operator A applies when a_{ij} is presented, the (positive) transfer operator T applies when a_{ik} ($k \neq j$) is presented, and the forgetting operator applies when an item not belonging to category i is presented.

Now we define the schedules in terms of sequences of the a_{ij}'s. In all subsequent work the sequences are described by strings of operators. Written in matrix form, schedules M (massed) and D (distributed) given by the matrix

$$\begin{bmatrix} a_{11} & a_{12} & \text{-----} & a_{1E} \\ a_{21} & a_{22} & \text{-----} & a_{2E} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \text{----} & a_{iE} \\ a_{C1} & a_{C2} & \text{----} & a_{CE} \end{bmatrix}.$$

Reading from left to right within a row, taking first the top row, then the second row, etc. gives schedule M. To obtain schedule D, we transpose, reading down the columns beginning with the leftmost column.

2. Operator Sequences

The first step in the derivations is to write the operator sequence for item a_{ij} under each schedule. At this more abstract level we generalize to N cycles through the list, but when we restrict the operators we also find it convenient to restrict N. It is easier to divide the sequence into two subsequences. The first begins with an item's initial presentation and ends on the trial before its final presentation. This constitutes the N-1 complete cycles for each item, where N is the number of reinforcements per item. The second sequence is a partial cycle beginning with the item's last reinforcement and ending with the last reinforcement of the terminal item in the list.

Calculation of the operator sequences is straightforward for the N-1 complete cycles. Schedule M yields

$$M: \quad [AT^{E-j}\ F^{(C-1)E}\ T^{j-1}]^{N-1}. \qquad (1)$$

The product in brackets represents the sequence for one cycle. Independent of i, item a_{ij} gets one presentation, then E-j exposures to other members of the same concept, then (C-1)E trials on other concepts, and then exemplars a_{i1} through $a_{i\ j-1}$.

The expression for schedule D can be written two ways; sometimes one computational form is more convenient and sometimes the other, so we give both. The first is

$$D: \quad [AF^{C-1}\ (TF^{C-1})^{E-1}]^{N-1}. \qquad (2a)$$

That is, the cycle for a_{ij} is initiated with a presentation of that item. Next there are C-1 forgetting trials, consisting of a presentation of one exemplar from each of the remaining C-1 concepts. The next block of C trials is the same as the first, except that a_{ij} gets replaced by a_{ik} $(k \neq j)$, yielding TF^{C-1}. And if there are E exemplars of each concept, the TF^{C-1} sequence must intervene E-1 times between successive appearances of a_{ij}. Hence the product in brackets denotes the sequences for a single cycle. In the alternative form for the N-1 complete cycles, i appears explicitly:

$$D: \quad [AF^{C-i}\ (F^{i-1}\ TF^{C-i})^{E-1}\ F^{i-1}]^{N-1}. \qquad (2b)$$

Here the AF^{C-i} factor signifies the effect of the sequence from the a_{ij} to a_{Cj}, inclusive. The factor in parenthesis means that a member of concept i is preceded by i-1 exemplars of other concepts and followed by one exemplar of each of the remaining C-i concepts.

The equivalence of Eqs. 2a and 2b is obvious. Note that both are independent of j. If in Eq. 2b we move the first F^{i-1} term to the left of the parenthesis, the expression becomes

$$D: [AF^{C-i} F^{i-1} (TF^{C-1})^{E-2} TF^{C-i} F^{i-1}]^{N-1}$$

which is the right side of Eq. 2a.

Now we derive the operator product for the partial cycle that begins with an item's last presentation and ends with the final reinforcement of the experiment. Here the expressions will depend on both i and j, since an item's position in the sequence determines the number of subsequent presentations. For massed exemplars it is clear by comparison with Eq. 1 that the operator product is

$$M: AT^{E-j} F^{(C-i)E}. \qquad \text{(partial cycle) (3)}$$

For distributed presentations the useful expression derives from Eq. 2b and is

$$D: AF^{C-i} (F^{i-1} TF^{C-i})^{E-j}. \quad \text{(partial cycle) (4)}$$

This reflects the fact that after exemplar j has been reinforced the final E-j exemplars of each concept remain to be presented.

3. Linear Operators.

We proceed by assuming that all operators are linear. As in the Suppes' model for the effect of vocabulary block-size, our sole concern is with the mean learning curve. Hence we can assume either a one-element or a single-operator linear process. Letting q_{ij} equal the probability of an error upon presentation of item a_{ij}, the (path-independent) linear operators for all i and j are for $0 \leq a, b \leq 1$.

$$\begin{aligned} A(q_{ij}) &= aq_{ij} \\ T(q_{ij}) &= bq_{ij} \\ F(q_{ij}) &= (1 - f)\, q_{ij} + f(1 - g), \end{aligned} \tag{5}$$

where 1-g has the customary one-element or linear interpretations. Therefore $0 \leq q_{ij} \leq 1 - g$, and we anticipate that $a \leq b$. Routine computations of the operator products in Sec. 2 are facilitated if we record a well-known result.

If φ is any operator such that for response probability q,

$$\varphi(q) = Rq + S,$$

where R and S are suitably constrained constants, then

$$\begin{aligned} \varphi^2(q) &= R(Rq + S) + S = R^2 q + S(1 + R) \\ &\vdots \\ \varphi^k(q) &= R^k q + S\left(\frac{1 - R^k}{1 - R}\right). \end{aligned} \tag{6}$$

Now we derive the complete-and partial cycle expressions for schedule M, followed by those for schedule D. Major interest in all four derivations focuses on the case of only one cycle (complete or partial) through the list. This restriction materially simplifies the comparison of the two schedules. That is, the comparison must be based on mean performance over all items, and the averaging over i and j becomes complex if we permit more than one cycle.

Schedule M. Equations 1 and 6 are relevant for the complete cycle. Let $\psi_M(q)$ be the outcome of a complete cycle under schedule M, starting at an initial level q. Abbreviate $(1 - f)^{C-1}$ as x. By Eq. 1, q gets transformed into $ab^{E-j}q$, then into $x^E ab^{E-j} q + (1 - x^E)(1 - g)$. The final T^{j-1} in Eq. 1 carries this latter quantity into

$$\psi_M(q) = x^E ab^{E-1}q + b^{j-1}(1 - x^E)(1 - g). \qquad (7)$$

The outcome of N - 1 complete cycles can be written at once via Eq. 6. However, we omit this, since it does not seem to offer any new insight into the analysis. Instead, we find the mean over items of $\psi_M(q)$, denoted $\bar{q}_M$. Averaging over j (and i) gives

$$\bar{q}_M = x^E ab^{E-1} q + (1 - x^E)(1 - g)\left[\frac{1 - b^E}{E(1 - b)}\right]. \qquad (8)$$

Using primes to denote partial cycles, Eq. 3 implies that

$$\psi'_M(q) = (1 - f)^{(C-i)E} ab^{E-j} q + [1-(1 - f)^{(C-i)E}](1 - g).$$

Letting q = 1 - g, the mean over i and j is found thus:

$$\psi'_M(q) = (1 - f)^{(C-i)E} (1 - g) [ab^{E-j} - 1] + 1 - g,$$

$$\bar{q}'_M = \left[\frac{1-(1 - f)^{EC}}{1-(1 - f)^E}\right] \frac{(1 - g)}{C} \left[\frac{a}{E}\left(\frac{1 - b^E}{1 - b}\right) - 1\right] + 1 - g. \quad (9)$$

Schedule D. As before, we derive the mean q-value for a complete cycle and then that for a partial cycle. These will be the analogs of Eqs. 8 and 9 respectively. From Eq. 2a we have that the initial AF^{C-1} transforms a q-value into xaq + (1 - x)(1 - g). Then Eq. 6 tells us that E - 1 applications of TF^{C-1} transform our new q-value into

$$(xb)^{E-1}[xaq + (1 - x)(1 - g)] + (1 - x)(1 - g) \left[\frac{1 - (xb)^{E-1}}{1 - xb}\right].$$

Again letting the initial q-value be 1 - g, and collecting the coefficients of (1 - x)(1 - g) gives

$$\bar{q}_D = (xb)^{E-1} xa(1 - g) + (1 - x)(1 - g) \left[\frac{1 - (xb)^E}{1 - xb}\right]. \quad (10)$$

We are justified in replacing $\psi_D(1 - g)$ by $\bar{q}_D$ on the left side of the above equation, since the former is found to be independent of i and j.

Turning now to the partial cycle, Eq. 4 tells us that as a preliminary step we must compute the effect of $(F^{i-1}\ TF^{C-i})^{E-j}$ operating on some q-value. Straightforward calculation gives for the effect of $F^{i-1}\ T$:

$$q \rightarrow (1 - f)^{i-1}\ bq + [1-(1 - f)^{i-1}]\ b\ (1 - g).$$

This is transformed by the F^{C-i} operator, so the outcome of one of the E − j replications is

$$q \rightarrow (1 - f)^{C-1}\ bq + (1 - f)^{C-i}\ b(1 - g) - (1 - f)^{C-1}\ b(1 - g) + [1-(1 - f)^{C-i}](1 - g).$$

$$= bxq + (1 - f)^{C-i}\ (1 - g)(b - 1) + (1 - g)(1 - bx). \quad (11)$$

So by the general Eq. 6, E − j replications yield

$$q \rightarrow (bx)^{E-j}\ q + h_i \left[\frac{1 - (bx)^{E-j}}{1 - bx}\right],$$

where h_i represents all terms to the right of bxq in Eq. 11. The average over j is

$$\left[\frac{1 - (bx)^E}{E(1 - bx)}\right]\left[\frac{q(1 - bx) - h_i}{1 - bx}\right] + \frac{h_i}{1 - bx}\ .$$

Averaging over i, we need to determine

$$\left[\frac{1 - (bx)^E}{E(1 - bx)^2}\right]\frac{1}{C} \sum_{i=1}^{C} q(1 - bx) - h_i + \frac{1}{C} \sum_{i=1}^{C} \frac{h_i}{1 - bx}. \quad (11a)$$

From Eq. 4, q is the outcome of AF^{C-i} operating on 1 − g, so $q = (1 - f)^{C-i}\ a(1 - g) + [1-(1 - f)^{C-i}](1 - g)$. To forestall a possible confusion, we note that when q represented the error probability at the start of the experiment (as in deriving Eqs. 9 and 10), we had $q = 1 - g$. Algebraic manipulation of the terms inside the first summation above yields $q(1 - bx) - h_i = (1 - f)^{C-i} (1 - g)[a(1 - bx) - b(1 - x)]$. Hence the means over i in Eq. 11a are

$$\frac{1}{C}\left[1 - \frac{(1 - f)^C}{f}\right](1 - g)\,[a(1 - bx) - b(1 - x)] + \frac{(1 - g)}{C(1 - bx)}\left\{(b - 1)\left[\frac{1-(1 - f)^C}{f}\right] + C(1 - bx)\right\}.$$

After simplification, the entire average over i is

$$\bar{q}_D' = \left[\frac{1-(1 - f)^C}{1-(1 - f)}\right]\left(\frac{1 - g}{C}\right)\left\{\left[\frac{1-(bx)^E}{E(1 - bx)}\right]\left[a - \frac{b(1 - x)}{1 - bx}\right] - \frac{1 - b}{1 - bx}\right\} + 1 - g. \qquad (12)$$

4. <u>Comparisons between massed and distributed schedules</u>.

The complete cycle comparison is between Eq. 8 for schedule M and Eq. 10 for schedule D. The corresponding results for a partial cycle are Eqs. 9 and 12. With either type of item cycling, our main question is: which schedule produces a higher predicted proportion correct on the test? As we have already remarked, the answer depends heavily on how we conceive of the test. If we view the test as simply a second run through the list, with reinforcement continuing to follow each response, then we have the complete cycle procedure. In our vocabulary experiments reinforcements indeed occurred on the test but there the items were not selected from common categories. On the other hand, if we conform to the more customary method, that of withholding reinforcement on the test, then the learning phase is a partial cycle. As we shall now demonstrate, under a complete cycle schedule M is more efficient than schedule D for all values of a, b, and f, and all number of concepts and exemplars. Unfortunately, with a partial cycle no analytic condition has been discovered that would guarantee the superiority of one schedule to the other.

To verify these two conclusions we cite the derivations in the preceding section. For a partial cycle the comparison of Eqs. 9 and 12 is not very enlightening. Of course, the trivial case E = 1 yields $\bar{q}'_M = \bar{q}'_D$. Also, it is hardly surprising that when b = 0 (perfect learning on a transfer trial) massed presentations are less effective than spaced, as the reader can quickly show. Another clue concerns the upper bound on $\bar{q}'_D - \bar{q}'_M$:

$$\bar{q}'_D - \bar{q}'_M \leq u \qquad (13)$$

where we abbreviate from Eq. 12 thus:

$$y = \frac{1-(1-f)^C}{f}, \qquad z = \frac{1-(bx)^E}{E(1-bx)}, \qquad u = \frac{y(1-g)}{C}\,\frac{b(1-x)(1-z)}{(1-bx)}.$$

The proof of the above inequality proceeds by subtracting u from $\bar{q}'_D$ and showing that the difference, although positive, is less than $\bar{q}'_M$.

$$\bar{q}'_D - u = y\left(\frac{1 - g}{C}\right)\left\{ z\left[a - \frac{b(1 - x)}{1 - bx}\right] - \frac{(1 - b)}{1 - bx}\right\} + 1 - g - y\frac{(1 - g)}{C}\frac{b(1 - x)(1 - z)}{1 - bx}$$

$$= y\frac{(1 - g)}{C}\left[za - 1\right] + 1 - g.$$

Here and in Eq. 9, the product before the final $1 - g$ is negative. But the absolute value of the product in the above line exceeds that in Eq. 9, so $\bar{q}'_D - u \leq \bar{q}'_M$, which verifies Eq. 13. One would like to proceed by comparing $\bar{q}'_M - (\bar{q}'_D - u)$ with u, but the author has not found any interesting results, even when $b = 1 - f$. Recall that in Suppes' development for independent items the largest block-size is optimal when the learning parameter exceeds the forgetting parameter, whereas the smallest size is optimal when forgetting is faster than learning. But in the present development the relative effects of E and C introduce complications that were absent from the block-size model. Note that $u \to 0$ as $C \to \infty$, although this limit is not of practical importance.

To explore further the partial cycle comparison between schedules, numerical computations of $\bar{q}'_M$ and $\bar{q}'_D$

were performed. Arbitrarily fixing a at .3 and 1 - g at .67, we took all combinations of b = .3, .6, .9, with f = .3, .6, .9 and C = 3, 6 and E = 3, 6. Inserting these values in Eqs. 9 and 12, we made 36 computations of $\bar{q}_M'$ and 36 of $\bar{q}_D'$. Figures 1A - 1C plot the

---

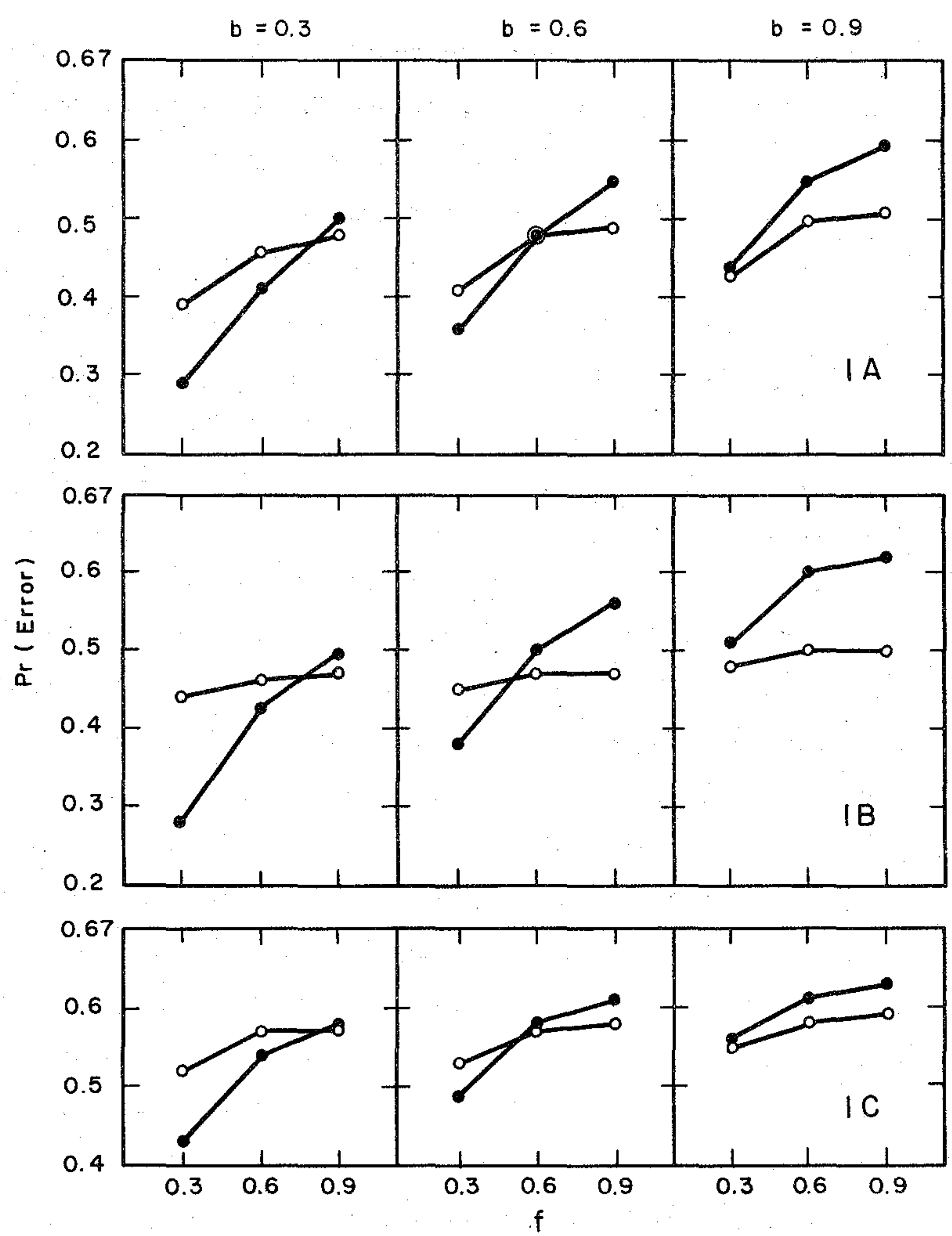
Insert Figs. 1A - 1C here

---

probability of an error against f, each panel corresponding to a particular value of b, C, and E. [2]

One striking generalization is that low values of b and f produce $\bar{q}_D' < \bar{q}_M'$, whereas high values of these two parameters reverse the inequality. When there is little positive transfer (b = .9), then massed exemplars are more efficient than distributed for most explored values of the other parameters. But when positive transfer is high (b = .3, moreover, b = a) then in general distributed exemplars are more efficient than massed. By and large, Figures 1B and 1C also suggest that the difference between $\bar{q}_M'$ and $\bar{q}_D'$ is more pronounced when the number of exemplars exceeds the number of concepts than when the converse holds.

Another instructive point is that within the range



Figs. 1A - 1C. Theoretical probability of an error plotted against f. Panels in a figure represent different values of b, while E and C vary from one figure to another. All computations were made with a = .3 and 1 - g = .67.

of sampled values, $\bar{q}_M'$ depends mostly on C and is relatively insensitive to changes in f, b, and E. Figures 1A and 1B reveal that when C = 3, $\bar{q}_M'$ ranges only from .41 to .50, except for the first point in Figure 1A. Figure 1C and the results for C = 6, E = 6 indicates that M varies only from .52 to .59. Within the set of values studied, then, performance under the massed schedule changes little as we vary b, f, and E while holding C constant.

By way of summarizing the numerical analysis, we note that with a high degree of learning schedule D is superior to schedule M, and conversely when a single cycle produces little learning. Thus from a practical angle schedule D is superior to schedule M. Our rationale for this evaluation is that the comparison between schedules possesses more than academic interest only if at least one schedule is effective.

For partial cycles, how can we reconcile this evaluation with intuition and with any data that might favor schedule M? One tentative answer is that for a difficult task schedule M might be preferable, but more than one cycle might be required to reach a low probability of error. A second possibility is that the underlying learning model may be empirically inadequate. For example, the transfer operator may not be path-independent. Future theoretical efforts might include an exploration of these two suggestions.

A much more satisfying conclusion emerges when we examine the situation with reinforced test trials. The relevant equation for schedule M comes from setting $q = 1 - g$ in Eq. 8, while that for schedule D appears in Eq. 10. Hence

$$\bar{q}'_M \leqq \bar{q}_D \iff \frac{(1 - x^E)(1 - b^E)}{E(1 - b)} \; \frac{1 - xb}{(1 - x)[1-(xb)^E]} \leqq 1$$

$$\iff \sum_{j=1}^{E} x^{j-1} \sum_{j=1}^{E} b^{j-1} \leqq E \sum_{j=1}^{E} (xb)^{j-1} \quad .$$

But now we can show that the last inequality holds for all values of E, C, and the three learning parameters. A simple proof by induction on E is given in the appendix. Therefore if a complete cycle is given and the presentation order is repeated on the test, schedule M is more efficient than schedule D.

Let us mention one objection that can perhaps be raised to the learning assumptions underlying our comparison of the two schedules. If two members of a common category do indeed share structural properties then reinforcement of the one exemplar might well facilitate performance on the second <u>prior to</u> the first reinforcement of the latter. To minimize formal complexity we have ignored such "proactive" facilitation. Upon casual reflection, it might appear that "proactive" interference could also result, due to other concepts preceding the first presentation of a particular item. However, the argument for interference does not seem so cogent as that for facilitation. The item affected by negative transfer would have its initial probability of a correct response lowered below the chance level. More important, our interpretation that intervening concepts produce forgetting rather than negative transfer rules out "proactive" interference.

Therefore if an item is affected at all prior to its first reinforcement, the influence would seem to be facilitory. The initial probability of an error on the jth exemplar would be reduced from $1 - g$ to $(1 - g)b^{j-1}$, independently of the schedule. A little thought reveals that this modification does not affect our conclusion for a complete cycle. Referring to Eqs. 7 and 10, we merely replace $1 - g$ by $b^{j-1}(1 - g)$. Then the average over j is $\frac{(1 - b^E)}{E(1 - b)}(1 - g)$ for both schedules. With q originally set at $1 - g$, the term containing q was identical in the two schedules. All we have done is to multiply each of two identical expressions by b^{j-1}, so they remain identical. The import of this derivation is that we have generalized the complete cycle result to include proactive facilitation, as the term is used here.

References

Bower, G. H., & Trabasso, T. Concept identification. In R. C. Atkinson (Ed.) Studies in mathematical psychology. Stanford University, 1964.

Crothers, E. J. Learning model solution to a problem in constrained optimization. J. math. Psychol., 1965, 2, in press.

Suppes, P.C. Problems of optimization in the learning of a simple list of items. In M. Shelly and Bryan (Eds.) Human judgments and optimality. New York: Wiley, 1964.

Suppes, P.C., & Ginsberg, R. A fundamental property of all-or-none models. Psychol. Rev., 1963, 70, 139-161.

Apostol, T. Mathematical analysis. Addison-Wesley, 1957.

Footnotes

[1] This investigation was supported by Contract 5-14-013 between Stanford University and the U. S. Office of Education.

[2] The figure for C = E = 6 is what we would expect from the others and is omitted for brevity.

[3] This is exercise 1.16 on p.20 of Mathematical Analysis, Apostol (1957).

Appendix

<u>Theorem</u>.[3] For all $x,b < 1$,

$$\sum_{j=1}^{E} x^{j-1} \sum_{j=1}^{E} b^{j-1} \leqq E \sum_{j=1}^{E} (xb)^{j-1} . \quad (1A)$$

The proof is straightforward and proceeds by induction on E. When E = 2, the above line is simply

$$(1 + x)(1 + b) \leqq 2(1 + xb), \text{ or}$$
$$x + b \leqq 1 + xb,$$

which is obviously true. Next, we separate out the final term of each sum in Eq. 1A.

$$\left(\sum_{j=1}^{E-1} x^{j-1} + x^{E-1}\right)\left(\sum_{j=1}^{E-1} b^{j-1} + b^{E-1}\right) \leqq E \sum_{j=1}^{E-1} (xb)^{j-1} + E(xb)^{E-1} .$$

By the induction hypothesis it suffices to prove that

$$x^{E-1} \sum_{j=1}^{E-1} b^{j-1} + b^{E-1} \sum_{j=1}^{E-1} x^{j-1} + x^{E-1} b^{E-1} \leqq \sum_{j=1}^{E-1} (xb)^{j-1} + E(xb)^{E-1} .$$

Subtracting $(xb)^{E-1}$ from both sides, the above line reduces to

$$\sum_{j=1}^{E-1} (x^{E-1} b^{j-1} + b^{E-1} x^{j-1}) \leqq \sum_{j=1}^{E-1} [(xb)^{j-1} + (xb)^{E-1}].$$

A sufficient condition is that for each value of j the term on the left be less than or equal to the corresponding term on the right. That is, for all j such that $1 \leq j \leq E$,

$$x^{E-1} b^{j-1} + b^{E-1} x^{j-1} \leq (xb)^{j-1} + (xb)^{E-1}.$$

Dividing through by $(xb)^{j-1}$:

$$x^{E-j} + b^{E-j} \leq 1 + (xb)^{E-j}, \text{ or}$$
$$x^{E-j} (1 - b^{E-j}) \leq 1 - b^{E-j}$$

The last statement is indeed valid, and the proof is complete. Equality arises only when $E = 1$.