Bell Laboratories Technical Memorandum
TM-73-1221-8 (12/5/73)

EVIDENCE AGAINST SELF-TERMINATING MEMORY SEARCH FROM PROPERTIES OF RT DISTRIBUTIONS[1]

Saul Sternberg
Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

Models that incorporate a self-terminating series of component operations are among the most popular explanatory devices in contemporary cognitive psychology. In instances where the order of the components may vary from trial to trial and are not under experimental control, it has proved somewhat difficult to discriminate such models from others.

Examples of such models are the self-terminating memory-search processes recently claimed to underlie character-classification and 1-1 choice (e.g., Theios, 1973) and context-recall (Sternberg, 1967a).

The present paper introduces two new properties of the reaction-time (RT) distributions produced by a broad class of self-terminating processes. (A third property is introduced in an appendix.) These properties, which are not captured by the usual mean and variance statistics, are represented by dominance relations among cumulative distribution functions (cdf's) of RTs, or among cdf's that have been transformed by rescaling the probability axis. Tests of the properties applied to data from several experimental paradigms yielded negative results. Models embodying self-terminating memory-search processes that have been proposed for classification and choice behavior must therefore be questioned.

The paper is organized in two main sections: an informal text that presents main ideas and results, and an appendix containing derivations, elaborations of the ideas, and supplementary analyses of data.

---

[1] Paper presented at the annual meeting of the Psychonomic Society, St. Louis, November 1973.

CONTENTS

EVIDENCE AGAINST SELF-TERMINATING MEMORY SEARCH FROM PROPERTIES OF RT DISTRIBUTIONS[1]

Saul Sternberg[2]
Bell Laboratories
Murray Hill, New Jersey 07974

1. INTRODUCTION

This paper is concerned with certain models of reaction time (RT) that incorporate a process of self-terminating memory search. By self-terminating I mean that the search stops as soon as the stimulus is found, so that not all of the stored items need be searched on every trial. Such models are very attractive, and have been receiving a good deal of attention lately.[3] They are

---

[1]Paper presented at the annual meeting of the Psychonomic Society, St. Louis, November 1973.

[2]I thank the John Simon Guggenheim Memorial Foundation for its fellowship support during the period in which the reported work was begun, and the Department of Psychology at University College London for its hospitality.

[3]Models that incorporate a self-terminating series of component operations are among the most popular explanatory devices in contemporary cognitive psychology. They have been applied, for example, to visual search in large arrays (e.g., Neisser, 1963), to scanning of post-exposure images of small arrays (e.g., Sternberg, 1967b), to same-different judgments of simple multidimensional stimuli (e.g., Egeth, 1966; Nickerson, 1972, Sec. 4), of letter strings (e.g., Bamber, 1969), of faces (Bradshaw & Wallace, 1971), and of letters (e.g., Posner & Mitchell, 1967), to 1-1 choice (e.g., Falmagne & Theios, 1969; Falmagne et al., 1973), to binary classification (e.g., Theios et al., 1973), to retrieval from categorized word lists (e.g., Naus, Glucksberg, & Ornstein, 1972), to context recall and recognition (e.g., Sternberg, 1969a), to lexical decisions about multiple letter strings (e.g., Meyer & Schvaneveldt, 1971), and to categorizing words in multiple semantic categories (Meyer, 1973). In most of these instances, the effects of important experimental factors are attributed, not to differences induced in the component operations themselves, but to

attractive because the self-terminating property seems plausible, and because the same search mechanism that explains effects of the number of items stored might also, in a natural way, explain effects of stimulus probability and stimulus sequence.

One reason for the attention that self-terminating search models have been receiving most recently is the set of ingenious arguments that John Theios has presented in several papers (e.g., Theios et al., 1973; Theios, 1973). Theios and his colleagues have proposed such models to account for results from a variety of experimental tasks, including the traditional choice-reaction paradigm, where each stimulus is assigned to a different response, and also the two-choice character-classification paradigm, such as I have studied (Sternberg, 1963; 1966; 1969a).

In this paper I report some new methods of testing models of self-terminating search. Their most unusual feature is that they permit testing the basic idea of self termination, with virtually no additional assumptions.[4] That is, the tests are relevant to very broad classes of models.

---

[3](cont'd) changes in the distribution of the number of components required. In those rare cases when it is believed that the components can be specified with certainty on each trial, this assumption, and therefore the general model, can be tested relatively directly. In the more usual case, the RT for even a specified trial must be regarded as having been drawn from a probability mixture of distributions; this is the case for which the tests presented in the present paper were designed.

[4]The additional assumptions that are often made include, e.g., (1) equality of mean durations of successive operations, (2) rules about the order of search and how it changes in response to stimulus events, and sometimes (3) the stochastic independence of the durations of component operations.

2. A SEARCH MODEL FOR 1-1 CHOICE REACTIONS

Let us start by considering search models for the choice-reaction paradigm, as shown in Fig. 1. This is a generalization of the kind of model advanced by Theios (1973) for this paradigm. Suppose that the stimuli in any

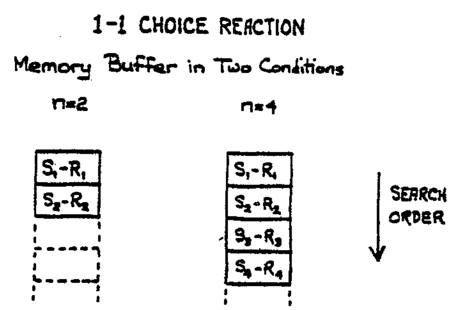

Figure 1

condition are presented from trial to trial with equal probability. The figure shows the assumed contents of a set of positions in a memory buffer in two conditions, one involving two different S-R pairs, and the other four. Only the first two positions in the buffer are used in the two-alternative condition. These same two positions, along with two others, are used in the four-alternative condition. On half the trials in the two-pair condition, the stimulus is found in the first position, along with the correct response, and the second position is not searched; on the other half of the trials, both positions are searched. On a quarter of the trials in the four-pair condition the stimulus is found in the first position, and the others are not searched, and so forth.

Three assumptions are needed for our purposes: First, the presence of information in more remote positions in the buffer has no effect on the time to find a stimulus that is at any particular higher position. For example, on trials where the test stimulus is in the first buffer position, the RT-distribution is the same, regardless of the number of alternatives.[5] A second critical assumption for the properties that I shall present is that the high positions in the buffer be tightly packed. For example, in the two-alternative condition, although the pairs are free to change places in the buffer according to any rule, neither of them can move to the third buffer position.

This description covers a very broad class of models. Note that I have said nothing about search time as a function of number of locations searched. The strongest assumption we need here is that moving a stimulus further down in the buffer cannot shorten the time to find it.[6]

3. TWO IMPLICATIONS OF THE MODEL FOR RT DISTRIBUTIONS

It is easy to see the intuitive basis of the two properties of this kind of model that I shall discuss. (As far as I know, these properties are necessary, but not sufficient, for the process to be self-terminating. That is, if a set of data exhibited these properties, this would not require the process to be self-terminating, although it would increase the credibility of such a process.)

The first property has to do with short RTs. Whatever the number of alternatives, the stimulus is found in the first buffer position on some trials. Therefore,

---

[5] See Section A3 for a method of estimating how large the required RT changes would be if this assumption were relaxed.

[6] See Section A.1 for a more precise statement of the model.

both conditions should generate some RTs that are equally short.

This idea suggests examining the minimum RT in a series of trials in each condition. But the sample minimum is a biased estimate of the population minimum, and I do not know of any way to evaluate its bias without making unacceptably strong distributional assumptions. The minimum has occasionally been used in RT studies--as long ago as Donders (1868), and most recently by Barry Lively (Lively & Sanford, 1972; Lively, 1972) and by D. H. Taylor (1965). Donders was aware even then of the estimation problem. As far as I know, the statistic I shall use to capture the intuition about short RTs avoids this difficulty.

The second property has to do with long RTs. The lowest buffer position in the four-alternative case is no more accessible than either position in the two-alternative case. Therefore the four-alternative condition should generate at least as many long RTs as the two-alternative condition.

Model: Mixtures of Nested Equiprobable Components

n = number of objects to search

$n=2$: $G_2(t) = \frac{1}{2}F_1(t) + \frac{1}{2}F_2(t)$

$n=4$: $G_4(t) = \frac{1}{4}F_1(t) + \frac{1}{4}F_2(t) + \frac{1}{4}F_3(t) + \frac{1}{4}F_4(t)$

Properties: ① $2G_4(t) \geq G_2(t)$

② $G_2(t) \geq G_4(t)$

[Requires assuming $F_1(t) \geq F_2(t)$, etc.]

Property ① in terms of quantiles:

$$G_4^{-1}(\tfrac{p}{2}) \leq G_2^{-1}(p)$$

Figure 2

These ideas are made precise in Fig. 2. The F's and G's are cumulative probability distribution functions (cdf's). F_1 is the RT distribution function on trials when the stimulus is found in the first buffer position. The distribution F_2 is generated on trials when both first and second positions are searched, and so on. The G's correspond to observed RT-distributions. Each G is an equal-probability mixture of F's. The word "nested" in the description of the model means that all the F's that contribute to G_2 also contribute to G_4.

The first inequality is mainly of interest for short RTs. Think of the G's as giving the percentage of RTs less than the value t. Suppose 40% of the RTs in the two-alternative condition are less than 300 msec. Then the first property says that at least half of this proportion, or 20%, of the RTs in the four-alternative condition must be that short.[7] For different numbers of alternatives the same kind of inequality holds, but with different multipliers.

The second inequality is mainly of interest for long RTs. It says, for example, that if 20% of the RTs in G_2 are longer than 400 msec, then at least 20% of the RTs in G_4 must also be longer than 400 msec.[8]

Taken together, the two properties mean that at all values of RT, G_2 is not only bounded above by twice G_4, but is bounded below by G_4 itself.

---

[7] In other words, there is a limit to how much smaller G_4 can be than G_2; if it was too small, then doubling G_4 could not reverse the inequality.

[8] It is in order to infer the second inequality that we must add the weak assumption mentioned above about how RT depends on the number of positions searched. The inequality on successive F's, $F_j(t) \geq F_{j+1}(t)$, (all t), says that by any specified time an item in a higher position is at least as likely to have been found than an item in any lower position.

As stated, these bounds apply to the (cumulative) probabilities associated with specified RTs. However, it is sometimes convenient to start instead with specified probabilities, and use the equivalent bounds between the RTs (quantiles) associated with them. The first property is expressed in terms of quantiles at the bottom of the figure. The (p/2)-percent point of the G_4-distribution can be no greater than the p-percent point of the G_2-distribution.[9] Because the quantiles to be compared involve different probability values, such as p versus p/2, I shall call them adjusted quantiles.

The two properties are illustrated in Fig. 3 with

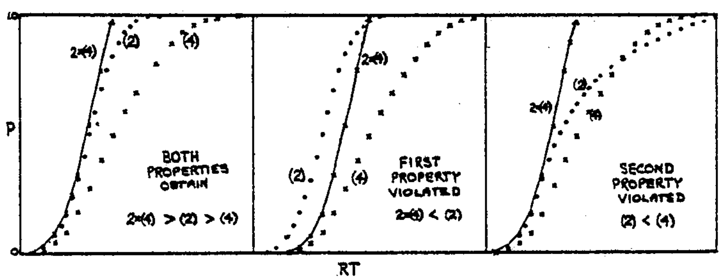

Figure 3

fictitious data. Each panel shows distribution functions for two and four S-R pairs, with dots and crosses. The curve represents $2G_4$. In the leftmost panel both properties obtain: G_2 never goes outside the region bounded above by $2G_4$ and bounded below by G_4 itself. In the middle panel G_2 and G_4 "march along" instead of "fanning out" from the same

[9]For small values of p, this assertion looks similar to the expected invariance of the minimum that I mentioned earlier, although tests are not subject to the same bias problems. The inequality on quantiles must hold for all values of p, $0 \leq p \leq 1$, but when testing a set of data for the property it is sometimes convenient to choose a single low value of p.

point. The result is that the first or short-RT property is seriously violated: for a large range of values G_2 is above $2G_4$. In the right-hand panel this does not occur, but the second or long-RT property is violated, since there are some time values for which G_4 is above G_2.

4. APPLICATION TO A 1-1 NUMERAL-KEYPRESS EXPERIMENT

Now I would like to consider several sets of data in light of these properties. Only the first experiment was designed with the present analysis of empirical distribution functions in mind, so it is the one that should probably be taken most seriously.[10] It was a choice-reaction experiment, with stimuli drawn from among the numerals 1 through 8. Each numeral was assigned to a different finger, in a natural way, and the responses were key presses. The three conditions involved two, four, and eight S-R pairs. Because each condition was run several times per session, each S-R pair could be studied in each condition. Four subjects were run for two sessions each of about 1000 trials; the data that I shall present are from the second session.

Figure 4 contains just the mean RTs of correct responses. It shows that the experiment reproduced the classical findings in this situation (Bricker, 1955; Brainard et al., 1962): the mean increases linearly with

---

[10]This experiment was performed in collaboration with Ronald L. Knoll.

In order that tests of the properties be valid regardless of the rules that govern movement of items within the buffer, the sequence of tested items in an experiment must be an independent-trials sequence. (See Section A1.2 for an explanation.) For tests to be completely valid, therefore, there can be no constraint in the randomizations that determine stimulus sequences. In all the experiments whose data are considered in this paper except the first, randomizations were constrained, so the conclusions must be regarded as tentative.

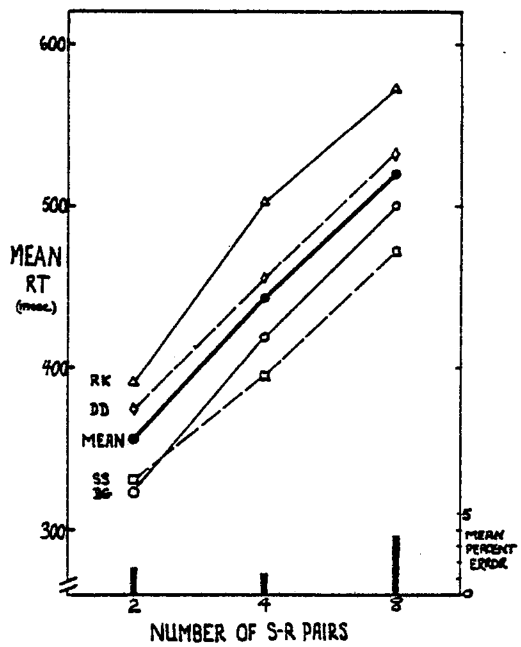

Figure 4

the logarithm of the number of S-R pairs, and the effect is nicely invariant across subjects.[11]

Figure 5 shows a RT distribution from each of the three conditions for one subject. Each distribution includes pooled data from all eight S-R pairs. Consider first the relations among the distributions themselves, before they are multiplied; these are shown by the

---

[11] A simple self-terminating search process produces a linear function, of course (e.g., Sternberg, 1966). However, a nonlinear function that relates mean RT to number of S-R pairs could be produced by a self-terminating search model if the mean times for successive search components (or degrees of accessibility) were permitted to be unequal, e.g., or if it was assumed that one or more registers in a memory buffer could contain more than one S-R pair (as in Theios, 1973).

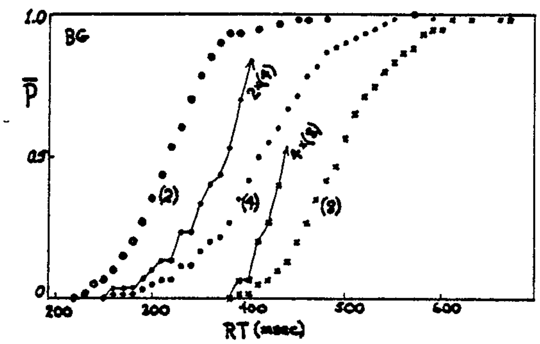

Figure 5

unconnected symbols, and permit a test of the long-RT property. Since the (2)-function is always above the (4)-function, which is always above the (8)-function, this property is satisfied. But the short-RT property is seriously violated. In order for this property to hold, the multiplied (4)-function should be at least as high as the (2)-function everywhere, and the multiplied (8)-function should be at least as high as the multiplied (4)-function everywhere. But neither of these conditions is met. Put roughly, there are not as many short RTs in the more difficult conditions as a self-terminating search-process requires. Data from the other three subjects reveal exactly the same pattern of results: conformity with the long-RT property, but violation of the short-RT property. Figures 6, 7, and 8 display the distributions for these subjects.

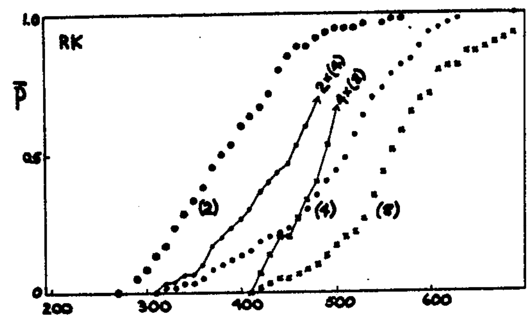

Figure 6

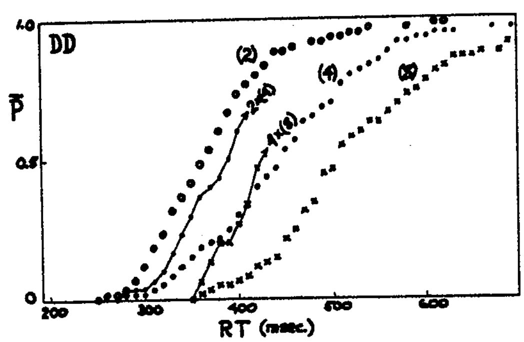

Figure 7

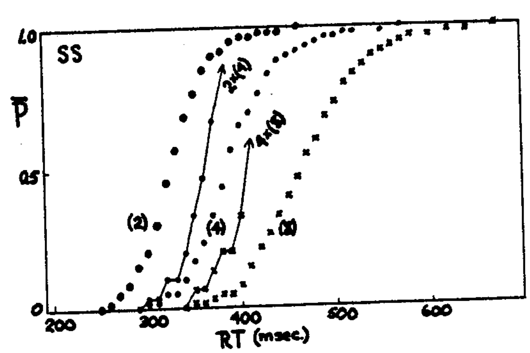

Figure 8

5. APPLICATION TO A 1-1 NUMERAL-NAME EXPERIMENT

The second experiment I would like to consider also used a choice-reaction paradigm with a 1-1 mapping of stimuli onto responses, and again the stimuli were numerals. But here the response to a numeral was the vocalization of its name. Five highly trained subjects each worked in two conditions, one with two S-R pairs, and one with eight.[12]

Again, the pairs of functions from each of the five subjects violated the short-RT property. But in addition, four of the five subjects violated the long-RT property as well. In Fig. 9 data from the two extreme subjects

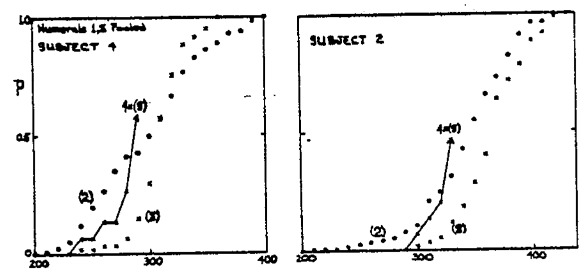

Figure 9

are displayed. The pair of functions on the left are the ones that most strongly violate the long-RT property: not only does the (8)-function contain too few short RTs; it also contains too few long ones. Subject 2, on the right of the figure, is the one subject whose distribution functions violated only one of the two properties.[13]

---

[12]This experiment was reported as Exp. V in Sternberg, 1969b.

[13]The most noteworthy constraint in the randomization that was used to produce the trial sequences in this experiment was that no immediate repeats were permitted in the eight-alternative condition. Although this might contribute to a violation of the short-RT property, however, it is hard to see how it could induce a violation of the long-RT property.

It appears that without elaboration, a self-terminating search model is inappropriate for the 1-1 choice-reaction paradigm.[14] Successive-dichotomization processes, such as those discussed by Hick (1952) and by Welford (1971), would not be embarrassed by violation of the first property, but could not easily be reconciled with violations of the second.

6. APPLICATION TO A CHARACTER-CLASSIFICATION EXPERIMENT: TEST OF A MODEL OF SELF-TERMINATING SEARCH THROUGH THE POSITIVE SET

Next I would like to turn to the character-classification paradigm, and two varieties of self-terminating search that have been considered as possible models. In the first model the items searched are the stimuli in the positive set. In the second model the items searched are stimulus-response pairs for all the stimuli in the experiment. The first model is shown in Fig. 10. The

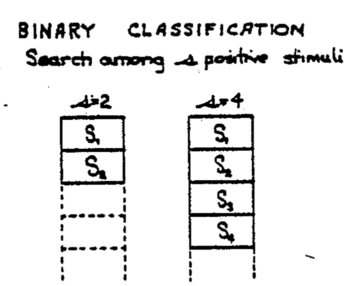

Figure 10

response rule here is to initiate a positive response if and when the test stimulus is found in the buffer, and to initiate a negative response otherwise.

---

[14] Taylor's (1965) report of a minimum that increased markedly with number of S-R pairs suggested this eight years ago, and the present analysis refines the argument.

I have previously argued (Sternberg, 1963; 1966) that search in this paradigm is exhaustive rather than self-terminating. This inference was based on the fact that as positive-set size is increased, mean RTs for positive and negative responses increase at about the same rate. But alternative explanations are available for the relation between RTs for negative and positive responses.[15] Furthermore, if the search process were assumed to be self-terminating, then sequential and stimulus probability effects might be easier to explain without having to invoke properties of other stages of processing. This makes it interesting and useful to test for self-terminating search through the positive set by using data from positive responses only.

The data that I used for the test were collected from 12 subjects, each serving for one hour. The stimuli were the ten numerals, and each subject was tested with

---

[15] An exhaustive search process alone is insufficient to account for some features that have occasionally been observed in RT data from the character-classification paradigm. Without elaboration, such a process cannot explain serial-position effects, sequential effects, or probability effects. Since it has been found necessary to postulate additional processing stages in addition to the search stage (Sternberg, 1969b), it is tempting to explain these effects in terms of properties of the additional stages, and retain the exhaustiveness of the search stage.

But an alternative explanatory line to pursue that might appear more parsimonious is to postulate a self-terminating search stage, and explain the observed relation between times for negative and positive responses in some other way. One possibility is that (self-terminating) search that produces a match with the test stimulus is repeated before a positive response is initiated, whereas a negative response is initiated after only one search. A second possibility is that altogether different processes, possibly operating in parallel, mediate positive and negative responses, just as has been proposed for "same" and "different" responses in letter-string matching (Bamber, 1969).

fixed positive sets of size 1, 2, and 4.[16] For each subject, three empirical distribution functions were determined. The test was limited to the short-RT property only. To simplify the analysis I used a set of adjusted quantiles, rather than the cdf's themselves. The left-hand side of Fig. 11 may help make this clear. The unconnected

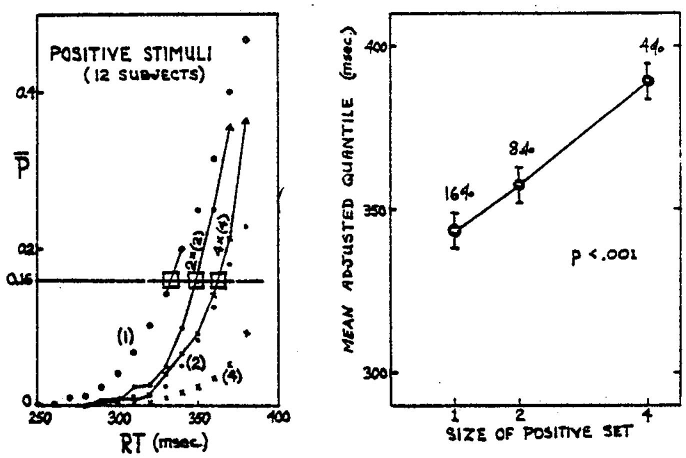

Figure 11

points represent the short-RT portions of the empirical cdf's for the three conditions, averaged over subjects. If each subject's cdf's satisfied the first property, then these average cdf's would satisfy it also. From the behavior of the multiplied cdf's you can see that they do not: the original ordering is maintained, even after multiplication.

Also shown with the cdf's is a 16% line. Squares mark its intersection with G_1, with $2G_2$, and with $4G_4$.

---

[16] This experiment has been reported briefly as Exp. IV in Sternberg, 1969b. In the conditions whose data are considered here, positive and negative responses were equally probable, and within positive and negative sets stimuli were equally probable.

These intersection points define a set of three RTs or adjusted quantiles: in particular, the 16% point of G_1, the 8% point of G_2, and the 4% point of G_4. These adjusted quantiles were determined for each of the 12 subjects, and submitted to analysis of variance.[17] Results (in terms of means and their SE's) are shown on the right-hand side of the figure. If the short-RT property holds, the adjusted quantiles cannot increase with set size. But their observed increase is highly significant.[18] It follows that if it is the positive set that is searched by subjects in the character-classification paradigm, then the search is not self-terminating.

---

[17] It should be noted that sample quantiles are not, in general, unbiased estimators of the corresponding population quantiles. The amount of bias depends on the local curvature of the cumulative distribution function (or the slope of the density function). If one assumes the shape of the tail of the distributions in question be to approximately normal, then an approximate correction for the bias is available in Harter (1961). If one assumes only that for small values of probability, α, the curvature increases with α (as, e.g., in the normal distribution), then for quantiles associated with small probabilities the bias works to favor the short-RT property, so the bias cannot be responsible for violations of the property.

Alternative methods still to be explored for testing, e.g., $2G_4(t) \geq G_2(t)$ in the data from a group of subjects, include (1) Choose a particular low value of α. For each subject find the sample quantile $t_0 = G_2^{-1}(\alpha)$ from the empirical cdf. Now compare the proportions $2G_4(t_0)$ and $G_2(t_0)$ defined by the empirical cdf. The proportions in this comparison are unbiased estimates of the corresponding population probabilities. (2) Comparison of entire functions. (Note that because one of the functions is a transformed cdf, the Kolmogorov-Smirnov test is not directly applicable.) These alternative methods may be preferable to the comparison of adjusted quantiles.

[18] This result refines the finding (Lively & Sanford, 1972; Lively, 1972) that RT sample minima increase with size of the positive set.

7. APPLICATION TO A CHARACTER-CLASSIFICATION EXPERIMENT: TEST OF A MODEL OF SELF-TERMINATING SEARCH THROUGH STIMULUS-RESPONSE PAIRS

Figure 12 shows the second model I would like to

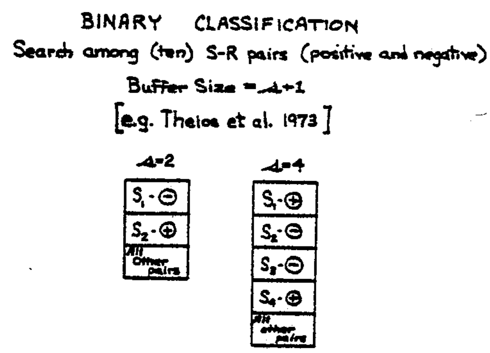

Figure 12

consider for character classification. This is a generalization of the model that has been advanced by John Theios and his colleagues (Theios et al., 1973). They suggest that what is searched in the classification paradigm is a set of S-R pairs involving all stimuli in the experiment, both negative and positive, and furthermore, that this search is self-terminating. (When the stimulus is found, the response associated with it is initiated.) Three other important features of their model are shown in the figure. First, the size of the buffer varies with the size of the positive set, even though the total number of S-R pairs that have to be accessed remains fixed. Thus, as the positive set is increased in size, more distant buffer positions are used. Second, this can be accomplished because the lowest of the occupied buffer positions can contain several S-R pairs, which are accessed in parallel. Finally, positive and negative S-R pairs can be intermingled in the buffer.

Because of the variable buffer size in this model, the cdf-multipliers used to generate the short-RT bounds are somewhat different from those already discussed. The result is that the bounds are more generous. Data from the same 12 subjects were used for this test as for the last one, but this time I pooled data from positive and negative responses.[19] For each subject I then determined adjusted low-order quantiles that would permit making the three pairwise comparisons among functions. Results (in terms of means and their SE's) are shown in Fig. 13. As

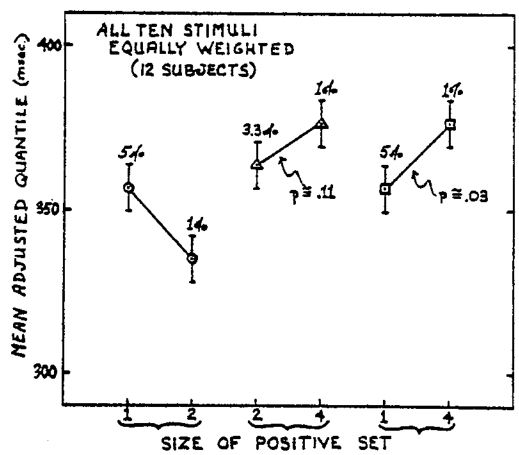

Figure 13

before, if the first property holds, the adjusted quantiles cannot increase with set size. The relation between G_1 and

[19] See Section A2.2 for a statement of the bounds in this case and for a sample derivation. Note the importance, for the validity of this test, of pooling in such a way that all of the ten stimuli have equal weight in each of the three cdf's.

G_2 satisfy the model, but in the other two comparisons the model fails, at moderate levels of significance. With a positive set of size four, too few short RTs are generated.

These conclusions must be somewhat tentative, since, again, the experiment was not designed with this kind of analysis in mind. Within this limitation, however, the relations among the RT distributions are inconsistent with a self-terminating search model for character-classification in which the items are S-R pairs, even when a memory buffer of adjustable size is postulated. I should mention, incidentally, that self-terminating models of this type in which the buffer size is fixed fare even worse in relation to the short-RT property.[20] It seems, then, that to explain performance in character-classification in terms of a search process, we may be forced to stick with an exhaustive search through the positive set. In that case, effects of trial sequence, stimulus probability, and serial position, where they are observed, may have to be explained in terms of their influence on stages of processing other than the search stage.

8. SUMMARY

In summary, I have described two properties that hold for RT distributions from a very broad class of serial self-terminating search models--broad enough to include all the existing models I know of as special cases. Once they are written down, the properties seem simple, obvious, and weak. Yet they are powerful enough to permit us to reject this class of models for both 1-1 choice-reaction tasks and character-classification experiments.

---

[20] See Section A2.1 for a test of models with fixed-size buffers.

APPENDIX

A1. DERIVATION OF THREE PROPERTIES OF THE CDF'S OF RTS FROM SELF-TERMINATING SEARCH AND RELATED PROCESSES

A1.1 The Model and Alternative Interpretations

We suppose that a set of items, indexed by i, occupies a set of states of accessibility, indexed by j. When an item is in state j, the RT cdf for that item is $F_j(t)$. The idea that state k is less accessible than state j is represented by the dominance (or stochastic order) relation:

$$F_j(t) \geq F_k(t), \text{ (all } t\text{)}.$$

This model has a variety of interpretations. The states of an item might correspond, e.g., to its location in different registers of a buffer memory (Theios, 1973), or in different positions in a display (Sternberg, 1967b). Alternatively, the items might correspond to different dimensions in a multidimensional stimulus (Egeth, 1966), whose states differ in accessibility because they are searched, tested, or compared in a particular order. The states of an item might also correspond, e.g., to different degrees of "trace strength" depending on time since the item was last rehearsed in a recycling rehearsal process (Nickerson, 1972, Sec. 2.2.6; Baddeley & Ecob, 1970, 1973). (Note that in the trace-strength interpretation, if the RT depends on the difference between sampled and criterion trace strengths, and the criterion changes as number of items increase, states could not correspond to trace strengths themselves, and the available states would not be invariant across conditions. But it will be seen that if the direction of criterion change is toward lower strength with larger numbers of items, then the criterion change cannot produce violations of the short-RT property.)

The properties to be considered do not depend on the states being distinct. Hence, although we shall require that each state is occupied by at most one item, this requirement does not mean that two or more items cannot be equally accessible [$F_j(t) = F_k(t)$]. (Thus, under the buffer-search interpretation of the model, more than one item can occupy the same register in the buffer, and the associated subset of states would then not be distinct.) If two items can occupy a register, then two different states with the same distribution functions will be associated with that register. Let the states be ordered from high to low, such that

$$F_j(t) \geq F_k(t) \text{ iff } j < k. \tag{1}$$

For most applications we shall also require that the items are "packed" into the higher accessibility states, such that in order for a state, k, to be occupied by an item, all states, j, of higher accessibility ($j<k$) must also be occupied. It follows that if there are m different items, the occupied states are indexed $j = 1,2,\ldots,m$. Subject to this restriction, the items can move freely among the states according to any set of rules, in response to trial events. [For example, if two items occupy the same buffer register in one condition, so that two states are associated with that register, then there can be no condition in which fewer than two items occupy that register while registers of lower accessibility are occupied. Where

this packing property is not assumed, as in the Theios (1973) variable-size buffer model, special arguments can nonetheless sometimes be used, as in Sec. A2.2, to derive the distributional bounds that are the focus of this paper.]

A1.2 Equal Weighting of State Distributions in the Averaged Cumulative RT Distribution, and Its Invariance Relative to State-occupation Probabilities

Movement of items among states will generate for each item, i, and state, j, an occupation probability, α_{ij}. Since, on any trial, each of the items must be in some state,

$$\sum_{j=1}^{m} \alpha_{ij} = 1, \text{ (all i)}, \tag{2}$$

and since each of the states, $j = 1,\ldots,m$, must be occupied by some item,

$$\sum_{i=1}^{m} \alpha_{ij} = 1, \text{ (all j)}. \tag{3}$$

Equation 3 means that the average state-occupation probabilities, averaged over items, are all equal.

Now let us assume that the choice by the experimenter of which item must be retrieved on any trial is independent of which items are occupying which states on that trial. If it is thought that the occupation pattern might depend on the previous sequence of trials then to satisfy this independence requirement the sequence of items tested must be an independent-trials sequence. (Note that this does not mean that items must be tested with equal probabilities.)

If the independence requirement is satisfied it is easy to form the predicted average (cumulative) distribution function over items, where each item is weighted equally in the average. For a particular item, i,

$$G_i(t) = \sum_j \alpha_{ij} F_j(t). \tag{4}$$

Hence, if G(t) is the average distribution,

$$G(t) \equiv \frac{1}{m} \sum_i G_i(t) = \frac{1}{m} \sum_i \sum_j \alpha_{ij} F_j(t). \tag{5}$$

Reversing the order of summation and using Eq. 3, we get

$$G(t) = \frac{1}{m} \sum_j F_j(t). \tag{6}$$

What Eq. 6 shows is that G(t) is given by an equally weighted average of the distributions $\{F_j(t)\}$, $j = 1,2,\ldots,m$, for the occupied states, regardless of the probabilities $\{\alpha_{ij}\}$. This leads to the

> Invariance Property. G(t) depends only on the set of occupied states and not on the occupation probabilities.

(This property was not discussed in the text, but was used in the analysis reported in Sec. 7.) Thus, if the number of items is held constant, but operations are performed that alter the distributions $\{\alpha_{ij}\}$ of state-occupation probabilities, G(t) must be invariant. Properties 1 and 2, derived below, are therefore independent of the $\{\alpha_{ij}\}$. Operations that have been proposed to exert their influence by acting on the distribution of occupation probabilities include variations in the relative frequency with which items are tested, and variations in the sequential properties of trials.

A1.3 Derivation of the "Short-RT" Property (Property 1)

This derivation makes use of the representation provided by Eq. 6. Suppose that two experimental conditions differ in the number of alternative items, and therefore in the number of occupied states. Let the two numbers be m and n, with $m < n$. From the Invariance Property we know that regardless of the $\{\alpha_{ij}\}$ in the two situations,

$$mG_m(t) = \sum_{j=1}^{m} F_j(t) \le \sum_{j=1}^{n} F_j(t) = nG_n(t), \text{ (all } t), \tag{7}$$

so that mG_m is bounded above by nG_n, for all t. (Note that subscripts on G refer here to number of items, and that G has been defined (Eq. 5) as an equally-weighted average over items. Note also that this property does not depend on the dominance relation of Eq. 1.) If $RT < t_0$ defines a short RT, and the proportion of short RTs in condition m is α, this bound implies that the proportion of short RTs in condition n is at least $m\alpha/n$. Inequality (7) on the distribution functions, $\{G\}$, is equivalent to an inequality on the quantile functions, $\{G^{-1}\}$, which give RT as a function of cumulative probability:

$$G_m^{-1}\left(\frac{\alpha}{m}\right) \ge G_n^{-1}\left(\frac{\alpha}{n}\right), \ (0 \le \alpha \le m). \tag{8}$$

A1.4 Derivation of the "Long-RT" Property (Property 2).

Let $0 < m < n$. Then from the "packing" assumption we have

$$G_n = \frac{1}{n}\sum_{1}^{n} F_j = \frac{1}{n}\left(\sum_{1}^{m} F_j + \sum_{m+1}^{n} F_j\right) = \frac{m}{n} G_m + \frac{1}{n}\sum_{m+1}^{n} F_j, \qquad (9)$$

(where the argument, t, of the distribution functions, F and G has been suppressed). Now from the dominance relation of Eq. 1, for $j \leq m$, $F_j \geq F_{m+1}$. Hence,

$$G_m = \frac{1}{m}\sum_{j=1}^{m} F_j \geq \frac{1}{m}\sum_{j=1}^{m} F_{m+1} = F_{m+1}.$$

Furthermore, for $j \geq m + 1$, $F_{m+1} \geq F_j$. Combining this with the inequality above, we have

$$G_m \geq F_j, \; (j \geq m+1). \qquad (10)$$

Replacing F_j by G_m in Eq. 9, and using Eq. 10, we get

$$G_n \leq \frac{m}{n} G_m + \frac{n-m}{n} G_m,$$

or

$$G_n(t) \leq G_m(t), \; (\text{all } t, \; 0<m<n). \qquad (11)$$

In short, the dominance relation among the F's induces a corresponding dominance relation among the G's. Combining inequalities 7 and 11 we have

$$\frac{n}{m} G_n(t) \geq G_m(t) \geq G_n(t), \text{ (all t; } 0<m<n);$$

for all t, $G_m(t)$ must lie between $G_n(t)$ (below) and $\frac{n}{m} G_n(t)$ (above).

The proportion of RTs longer than t is given by $G^*_n(t) \equiv 1 - G_n(t)$; from Eq. 11 it follows that $G^*_n(t) \geq G^*_m(t)$. If RT > t_0 defines a long RT and the proportion of long RTs in condition m is β, this bound implies that the proportion of long RTs in condition n is at least β.

A2. DISTRIBUTION BOUNDS FOR ASSOCIATION-SEARCH MODELS OF CHARACTER CLASSIFICATION

A2.1 Models with Fixed-size Buffers

In one class of models considered, e.g., by Theios et al. (1973) for the character-classification paradigm, the size of the memory buffer is fixed regardless of the size of the positive set.

If the ensemble of test stimuli (union of positive and negative sets) is also fixed, this class of models generates remarkably strong predictions for the average distribution $G_s(t)$ as a function of positive set size, s. Since the only influence s can have is on the distribution of positions occupied by each stimulus within the buffer, the invariance property derived in Section A1.2 applies. As positive set size is varied, the average distribution must remain unchanged. A corollary is that the expectation of $G_s(t)$, or the mean RT averaged over stimuli, should be invariant over changes in s. I shall denote this special mean by $\mu^*(s)$.

For some experiments the mean $\mu^*(s)$ with equal stimulus weighting can be obtained as a weighted combination of the more usual statistics, $\mu_p(s)$ and $\mu_n(s)$, the means for positive and negative responses. Suppose an experiment in which positive and negative responses are each required on half of the trials, and the ensemble of test stimuli is of size m. Suppose that within positive and negative sets, stimuli are presented with equal probability, and that the m-s stimuli not assigned to the positive set in a condition are assigned to the negative set. Then

$$\mu^*(s) = \frac{s}{m}\,\mu_p(s) + \frac{m-s}{m}\,\mu_n(s).$$

For the character-classification experiment whose data were used in the analyses presented in the text, the mean, $\mu^*(s)$, is clearly not invariant, as shown in Table 1,

Table 1

s	1	2	4
$\mu_n(s)$	442.9	480.1	550.5
$\mu_p(s)$	388.5	437.0	498.9
$\mu^*(s)$	437.5±13.3	471.5±13.3	529.9±13.3

which displays means and 95% intervals in msec based on the data from 12 subjects.

If, as in the experiment reported in Theios et al. (1973), the ensemble of test stimuli is not fixed in the character-classification paradigm, but the buffer size is fixed, the properties derived in Sec. A1 apply, as in Eqs. 7 and 11, with m and n denoting number of items in the ensemble.

A2.2 Models with Variable-size Buffers

The model (Fig. 12) most favored by Theios et al. (1973) and presented in Theios (1973) as an account of performance in the character-classification experiment, postulates a buffer of size s+1, where s denotes the number of items in the positive set. At any time, each of the first s registers is assumed to be occupied by one item, and the remaining items are assumed all to occupy the last register, which is identified as a long-term memory state. (Thus, the accessibility of items in long-term memory is assumed to be higher if s is smaller.) The average distribution, G_s, for this model, with an ensemble of size ten, is given by

$$G_s = \frac{1}{10} \sum_{j=1}^{s} F_j + \frac{10-s}{10} F_{s+1}. \qquad (12)$$

Using Eq. 12 together with the dominance relation of Eq. 1, we obtain the following short-RT bounds for s = 1,2, and 4:

$$\begin{aligned} G_1(t) &\le 5G_2(t) \\ G_1(t) &\le 5G_4(t) \\ 3G_2(t) &\le 10G_4(t). \end{aligned} \qquad (13)$$

These bounds were used to determine the probabilities that defined the adjusted quantiles of Fig. 13.

As an example, we derive the last of inequalities (13). We have

$$10G_2 = F_1 + F_2 + 8F_3$$

$$10G_4 = F_1 + F_2 + F_3 + F_4 + 6F_6$$

Hence $$10G_4 \geq F_1 + F_2 + F_3.$$

Now we wish to find a multiplier, k, such that $kG_4 \geq G_2$. This requires $kF_1 + kF_2 + kF_3 \geq F_1 + F_2 + 8F_3$, or $(k-1)F_1 + (k-1)F_2 + kF_3 \geq 8F_3$. From Eq. 1 this is satisfied if $(k-1) + (k-1) + k \geq 8$, or $k = 10/3$.

The long-RT property follows easily from Eq. 12. [Note that if the retrieval-time distribution, F_{s+1}, associated with items in the long-term memory state is assumed to be <u>independent</u> of s in Theios' variable-size buffer model, which is perhaps more plausible, and the ensemble size, m, is fixed, then $G_{s+1}(t) \geq G_s(t)$, implying $\mu^*(s+1) \leq \mu^*(s)$, which is clearly false.]

A3. ESTIMATION OF DIFFERENCES BETWEEN CONDITIONS IN COMPONENT DISTRIBUTIONS

All the self-terminating search models I know of that have been advanced in published articles include something close to our assumption that the component F_j-distributions that are mixed to produce the G-distributions do not change as the number of components change. (The presence of information in more remote positions in a buffer, e.g., is assumed to have no effect on the time to find a stimulus that is at any particular higher position.)

Indeed, the explanatory force of such models lies in their ability to account for effects in terms of changes in the distribution of the number of component operations, rather than changes in the durations of the components themselves.

Given failure of the model, however, one direction to pursue while still retaining the self-terminating feature is to consider relaxing the assumption of invariant F_j's. How much change in the F_j's would be needed?

The size of the violation of the short-RT property can provide a lower bound on changes required in the F_j's. An an example, let us consider the model for the character-classification experiment in which the items searched are the members of the positive set. How much change in F_1 as a function of s is needed to account for the violations of the short-RT property shown in Fig. 11? Let $\{F_j(s)\}$ be the changing component distributions. For $s = 1,2$, and 4 we have

$$\begin{aligned} G_1 &= F_1(1) \\ 2G_2 &= F_1(2) + F_2(2) \geq F_1(2) \\ 4G_4 &= F_1(4) + F_2(4) + F_3(4) + F_4(4) \geq F_1(4). \end{aligned} \tag{14}$$

The multiplied G-distributions thus provide upper bounds on the altered F_1-distribution, which in turn provide lower bounds on the amount of change in F_1:

$$\begin{aligned} G_1 - 2G_2 &\leq F_1(1) - F_1(2) \\ G_1 - 4G_4 &\leq F_1(1) - F_1(4). \end{aligned} \tag{15}$$

It can be seen from Fig. 11 that the change required in F_1 is large enough to rob the search model of much of its explanatory power. In evaluating the size of the change required it is helpful to express it in terms of the means, $\mu_1(1)$, $\mu_1(2)$, and $\mu_1(4)$ of $F_1(1)$, $F_1(2)$, and $F_1(4)$, respectively. To do this we shall assume that the increase in the low-order quantiles of $F_1(s)$, associated ~~with an increase in s, is not accompanied by a decrease in~~ the dispersion of $F_1(s)$. Dispersion is here indexed by $\mu_1(s) - t_{1,\alpha}(s)$, or the difference between the mean, $\mu_1(s)$, of $F_1(s)$ and its α-quantile, $t_{1,\alpha}(s)$, for any low-order quantile. This assumption implies that for small α,

$$\mu_1(s) - \mu_1(1) \geq t_{1,\alpha}(s) - t_{1,\alpha}(1), \quad (s=2,3,\ldots). \tag{16}$$

Now, from Eq. 14,

$$\begin{aligned} t_{1,\alpha}(1) &= G_1^{-1}(\alpha) \\ t_{1,\alpha}(2) &\geq G_2^{-1}\left(\tfrac{\alpha}{2}\right) \\ t_{1,\alpha}(4) &\geq G_4^{-1}\left(\tfrac{\alpha}{4}\right). \end{aligned} \tag{17}$$

Combining Eqs. 16 and 17 we have

$$\begin{aligned} \mu_1(2) - \mu_1(1) &\geq G_2^{-1}\left(\tfrac{\alpha}{2}\right) - G_1^{-1}(\alpha) \\ \mu_1(4) - \mu_1(1) &\geq G_4^{-1}\left(\tfrac{\alpha}{4}\right) - G_1^{-1}(\alpha). \end{aligned} \tag{18}$$

Thus the required effects of s on $\mu_1(s)$ are bounded below by differences among the adjusted quantiles. The relevant adjusted quantiles, for $\alpha = .16$, are those shown in the right-hand panel of Fig. 11. They imply the values in msec shown in Table 2, which shows that even a lower bound on the required increase in $\mu_1(s)$ represents a large percentage of the increase in mean RT that is to be explained.

Table 2

	s=1 to s=2	s=1 to s=4
Lower bound on increase in $\mu_1(s)$	15.0	46.7
Observed increase in mean of G_s	48.5	110.4
Percentage	30.3%	42.3%

For the numeral-key choice reaction experiment discussed in the text, a similar argument can be made, but in this case we must consider required changes produced by increasing the number of alternative S-R pairs in the equal-probability mixture of F_1 and F_2, rather than in F_1 alone. Corresponding to Eq. 14 we have

$$\begin{aligned} G_2 &= [F_1(2)+F_2(2)]/2 \\ 2G_4 &\geq [F_1(4)+F_2(4)]/2 \\ 4G_8 &\geq [F_1(8)+F_2(8)]/2. \end{aligned} \qquad (19)$$

Again, Figs. 5-8 show the change required in the component distributions to be large.

If we let $\mu\dagger(n) = [\mu_1(n)+\mu_2(n)]/2$, we have, corresponding to Eq. 18,

$$\mu\dagger(4) - \mu\dagger(2) \geq G_4^{-1}(\tfrac{\alpha}{2}) - G_2^{-1}(\alpha)$$
$$\mu\dagger(8) - \mu\dagger(2) \geq G_8^{-1}(\tfrac{\alpha}{4}) - G_2^{-1}(\alpha). \quad (20)$$

With α = .20 the adjusted quantiles were determined for each of the four subjects, and averaged, with the results shown in msec in Tables 3 and 4.

Table 3

	n=2	n=4	n=8
Mean Adjusted Quantile	310.5	347.8	407.8
Mean Reaction Time	354.9	443.3	519.7

Table 4

	n=2 to n=4	n=2 to n=8
Lower bound on increase in $\mu\dagger(n)$	37.3	97.3
Observed increase in mean G_n	88.4	164.8
Percentage	42.2%	59.0%

Again, these lower bounds on the required increases in component means indicate much more than a second-order adjustment.

A4. TWO SELF-TERMINATING SEARCHES IN SERIES

One way to explain the failure of the self-terminating search model to account for the effects of number (n) of alternative S-R pairs in the 1-1 choice-reaction experiment, without giving up the idea of self-terminating search altogether, is to argue that n influences the duration of at least one stage of processing in addition to a self-terminating search stage. The notion that n influences more than one processing stage is an old one, supported by evidence from the subtraction method (see, e.g., Jastrow, 1890, p. 35). (Such influences would produce changes in the component distributions, such as F_1, from one n-value to another, as discussed in Sec. A3.) More recently (Sternberg, 1969, Sec. 5) an application of the additive-factor method to a choice-reaction experiment in which the stimuli were numerals led to the same conclusion.

Theios (1973) specifically excludes a two-stage explanation for the effects of n when the stimuli are numbers or letters. But if we nonetheless consider this possibility it would be in the spirit of his approach to experiments with nonalphanumeric stimuli to develop a model in which the two processes each involved self-terminating search, the first to identify the stimulus among the n possible stimuli, and the second to specify the response among the n possible responses. Exactly what such a model predicts for the relations among the G_n-distributions depends on the correlation assumed between the ordering of stimulus identities in the stimulus buffer, and the ordering of S-R pairs in the S-R pair buffer. Since buffer position in both cases is assumed to depend in a similar way on the same aspects of the trial sequence (recency and frequency) the correlation will be positive for any specific realization of the model. If the correlation is perfect, then the expected bounds are the same as for a one-stage self-terminating search. If the two orderings are independent (zero correlation) and the search rates in the two buffers are equal (worst case) then the multiplying constants in Eq. 7 must be squared:

$$\left(\frac{n}{m}\right)^2 G_n(t) \geq G_m(t), \text{ (all t).} \tag{21}$$

Inspection of Figs. 5-8 shows that this bound also fails. (Only if the correlation of orderings is negative will the minimum population RT increase with n.)

A5. INFLUENCE OF A SPEED-ACCURACY TRADE-OFF ON RT COMPONENTS

As conditions (such as the number, n, of S-R alternatives) are changed, both RT distributions and error rates are affected, in general. If we assume, as in the class of models under consideration, that the only processing stage influenced by n is the search process, then both effects are associated with this process. Note, however, that just as other stages make contributions to RT that are independent of n, so they might contribute to the error rate.

Suppose that for each of the component tests that make up the self-terminating search there is a speed-accuracy trade-off such that if β = Pr{correct test outcome} then $F_j(t) = F_j(t;\beta)$. Tests of the model like those we have outlined depend on β being constant from condition to condition. Without an accepted model of error production, of course, we cannot decide from the error data whether β is constant; we have been assuming, implicitly, that any effects of n on β that do occur have relatively small effects on F_j. On the other hand, if a constant value of β led to an unacceptably high error rate at large values of n, and β could be adjusted by the subject, larger n might be associated with higher β, and the corresponding changes in F_j could produce violations of the short-RT property (but not the long-RT property). It seems unlikely to me that the large changes required in the F_j could be produced by any plausible error model for the experiments we have been considering (see Section A3), but it is nonetheless instructive to consider the consequences of one such model (proposed by J.-C. Falmagne, personal communication).

Suppose that all errors during the search process arise from "false alarms"--i.e., matches with the contents of registers that are tested earlier than the register containing the relevant stimulus representation. Then β can be defined as the probability of a correct nonmatch. Suppose further that all errors that occur arise in the search process and that none arise in other stages. Then for the probability $P_n(c)$ of a correct response in the equal-probability n-alternative condition we have

$$P_n(c) = \frac{1}{n}\,\frac{1-\beta^n}{1-\beta}. \tag{22}$$

For the choice-reaction data shown in Figs. 4-8, the mean values of $P_n(c)$ were .984, .987, and .964 for n = 2, 4, and 8, respectively. Solving Eq. 22 for each of these n-values we obtain, for the required values of β_n, $\hat{\beta}_2 = .970$,

$\hat{\beta}_4 = .990$, and $\hat{\beta}_8 = .990$. Hence, although the estimated β increases from $n = 2$ to $n = 4$, there is no increase in β required from $n = 4$ to $n = 8$. Yet the RT data violate the model in both of these comparisons. If we relaxed the assumption that no errors arise from other stages of processing, then we would find that $\hat{\beta}_4 - \hat{\beta}_2$ is smaller than above, and that $\hat{\beta}_8 < \hat{\beta}_4$, an inequality in the wrong direction to account for the short-RT violations.

A6. TESTS OF THE SELF-TERMINATING SEARCH MODEL APPLIED TO CONTEXT-RECALL DATA FROM THREE EXPERIMENTS

Another memory-retrieval paradigm for which a self-terminating search model has been advanced is that of the "context-recall", or "successor-naming" experiment (Sternberg, 1967a; 1969a, Sec 12-14). Here the subject first memorizes an ordered list. In the most typical procedure any item in the list except the last can then be presented as a test item, and the correct response is to vocalize the name of the item that follows the test item in the list. On each trial, a new list is presented.

The model that has been proposed for performance in this task incorporates a self-terminating ordered search for the test item, through the items in the list. But here, the ordered list does not naturally correspond to the buffer with ordered registers that has been considered elsewhere in this paper. Rather than consistently starting with the first item in the list, the search may start at a random position, and then, if necessary, recycle to the first item again. Because of the possibility that search does not consistently start at the first item, two features that have been considered as possibilities for the process in the context-recall paradigm would, in effect, violate the "packing" assumption discussed in Secs. 2 and A1.1.

Both are relevant on those trials on which the search recycles. First, the test item may be compared to the last item in the list, just as it is compared to the other items, even though the last item is never tested. And second, the time between comparisons of the test item to the last and first items may be greater than the time between other pairs of successive comparisons, because of a time increment contributed by the recycling operation. Translated into terms of the model described in the text, these features would mean that, on any trial on which search did not begin with the first item, one (or more) register in the buffer might be effectively empty while less accessible registers contained relevant items.

Whether a self-terminating search with these features could produce data that violated the distributional inequalities would depend on the distribution of points in the list at which the search began, and on how this starting-point distribution changed between conditions. (That such violations might occur underlines the fact that not all possible self-terminating serial processes are covered by the tests discussed in the present paper.)

If an experiment were performed in which the postulated search could be guaranteed to start uniformly with the first item, the two features should not, of course, prevent the distributional inequalities from being satisfied. But such an experiment has not yet been done.

Nonetheless, data from three experiments on context recall were examined for the short-RT property. In Exp. A (Sternberg, 1967a; 1969a, Exp. 6) the list was presented once before test-item presentation. In Exp. B (one of the three conditions in Sternberg, 1969a, Exp. 7; serial-position functions provided in Sternberg, 1969c) the list was presented three times before test-item presentation. Increasing the number of list presentations has been shown to increase the slopes of the serial-position functions, indicating a higher probability of starting the postulated search at the beginning of the list. This, in turn, should reduce the importance of both features mentioned above. Experiments A and B used lists ranging in length from three to seven items, with the last item never tested, so that the number of tested list items ranged from two to six.

In Exp. C (an unpublished pilot study involving only two subjects) the paradigm was modified so that the last item could be tested, with the correct response defined to be the name of the first item. In this "circular list" design, list length, now equal to the number of tested list items, ranged from two to six items. Since all items could be tested, the first of the two features of the more usual paradigm that might produce violations of the short-RT property was absent. Mean adjusted quantiles from the three experiments, $G_n^{-1}(\frac{.6}{n})$, where n = number of tested items, are shown in msec in Table 5.

Table 5

Experiment	Number of Tested Items 2	3	4	5	6	95% Interval	Slope (3≤n≤6)
A (6 Ss)	583	632	660	700	687	±41	+20.5
B (6 Ss)	542	595	633	612	633	±43	+ 9.3
C (2 Ss)	500	615	601	595	643	±41	+ 7.8

The short-RT property is violated in all three experiments, even when the worst offender (data for n=2) is omitted. But considering the large rate of increase of mean RT with n in these experiments (about 125 msec/item) and the large standard errors, and assuming that comparisons across procedures and quantiles are meaningful, the model does not appear to fail as badly here as in some of the other cases we have considered. Furthermore, if the extent of the failure for $n \geq 3$ is compared for the three experiments (which can be done by comparing the tabulated slopes of least-squares lines fitted to mean adjusted quantiles versus n, for $3 \leq n \leq 6$) it can be seen that in a condition where the starting point of the postulated process is more likely to be at the beginning of the list, and in a condition where all items might be tested, the violation is less severe. This finding tends to confirm the ideas that in the usual paradigm the last item may be searched even though it is not tested, and that extra time is taken to recycle from the end to the beginning of the list during the search.

As for many of the questions considered in this paper, more suitable experiments are needed before we can properly interpret these violations of the short-RT property. They may indicate a self-terminating search with the added features mentioned above, or may indicate that an altogether different model is needed for context-recall performance.

A7. HOW WELL SUPPORTED BY HIS OWN DATA IS THEIOS' MODEL OF CHARACTER CLASSIFICATION?

The tests of a model of self-terminating search through a series of S-R pairs that have been applied to character-classification data in Sections 7 and A2 were applied to data not collected under exactly the same conditions as those of Theios and his collaborators (Theios et al. 1973, called "T1" below; Theios, 1973, Sec. F, called "T2" below). This may be particularly important for T1 where the paradigm differed, in a way that is perhaps fundamental, from earlier "item-recognition" or "character-classification" experiments (e.g., Sternberg, 1966) using a fixed-set procedure: in the earlier experiments and in the classification experiment considered in Sections 6, 7, A2.1 and A3, as well as in T2, the positive set is smaller than the negative, and their union is of fixed size; in T1 positive and negative sets are of equal size.

Are procedural differences responsible for the success claimed for the Theios et al. model in T1 and T2, or are tests that are restricted to mean RTs too insensitive to reveal actual failures? Let us consider how well the variable-size buffer model fits the mean RT data reported in T1 and T2.

In T2 three aspects of the data are emphasized: (1) the effect of positive-set size on $\overline{RT}$, (2) the effect of stimulus presentation probability on $\overline{RT}$, and (3) the effect of prior stimulus sequence on $\overline{RT}$. Consider these effects in turn.

(1) Set-size Effect. In T1 the model fits well. However, in T2 Theios emphasizes the statistical significance of the deviation of the observed $\overline{RT}$ versus set size function (which is concave downward) from the best-fitting linear function. *A fortiori* the observed function must differ significantly from the function generated by the variable-size buffer model that best fits these data, since the latter function is concave upward.

(2) Stimulus-probability Effect. In T2 the model fits reasonably well. But in T1, data and best-fitting

model deviate systematically. Relative to the data, the fitted probability effects are too large for the larger positive sets and too small for the smaller. This systematic discrepancy is shown in Table 6, whose entries were obtained from Table 2 in T1. The observed effects are based on data from Exps. 1 and 2 in T1. The fitted effects are based on simulations of the variable-size buffer model. For each size, s, of positive set the tabulated "effect" is defined by the difference between mean RTs for the smallest and largest presentation probabilities used with that s-value.

Table 6

Observed versus Fitted Effects (in msec) of Stimulus Probability from Theios et al. (1973)

Size of Positive Set	2	3	4	5
Probability Range	.15-.35	.05-.30	.05-.20	.05-.20
Positive-set Stimuli				
Observed Effect	45	82	63	52
Fitted Effect	26	71	71	90
Difference	+19	+11	- 8	-38
Percentage of Observed	+42%	+13%	-13%	-73%
Negative-set Stimuli				
Observed Effect	31	88	37	47
Fitted Effect	19	47	44	71
Difference	+12	+41	- 7	-24
Percentage of Observed	+39%	+47%	- 19%	-51%

The interaction between stimulus probability and size of positive set is substantially larger in the best-fitting model than in the data.

(3) Sequence Effect. To my knowledge, sequential effects predicted from the best-fitting model have never been quantitatively compared to sequential effects in the data.

Despite these discrepancies, it was concluded in both T1 and T2 that the self-terminating search model fitted well. One possible reason may be that the overall goodness-of-fit measure is insensitive. An indication of this insensitivity is the fact that using this measure, Theios et al. were unable to reject any of six different models (including pairs of models describing three conceptually distinct processes) fitted to the data in T1. In this connection it is useful to note the number of free parameters used in fitting the favored model in T1 and T2. For each size of positive set, a different buffer size was assumed. If we take each buffer size to be a separate parameter (possibly a debatable assumption) then in T1 there were nine free parameters (24 data points fitted) and in T2 there were eight free parameters (ten data points fitted).

Finally, a comment should be made on two apparently arbitrary features of the favored model in T1 and T2. First, the assumed increase in size of the short-term memory buffer to conform to the size of the positive set is said (T1, p. 332) to result from the subject's attempt to "perform the task as instructed". Otherwise the subject could keep the buffer at its minimum size and perform more efficiently with the larger set sizes. We must thus assume that although the subject is told to perform as fast as he can, he chooses not to do so. The second feature is that the time to access the long-term memory (represented by the last position in the memory buffer) is assumed to decrease as the size of positive set decreases. If this latter assumption is not made, the variable-size buffer model fails at a gross level; if the varied buffer-size feature is dropped then we have already seen (Section A2.1) that for experiments with a fixed stimulus ensemble the self-terminating search model fails seriously; this failure also applies to the data reported in T2.

A8. BOUNDS ON RT-DISTRIBUTIONS FOR INDIVIDUAL ITEMS IN THE TWO-STATE CASE

Models in which the number of distinct states is limited to two have been of particular interest to students of choice behavior (e.g., Falmagne, 1965; Falmagne & Theios, 1969; Theios & Smith, 1972; Falmagne et al., 1973).

A8.1 Two States Plus Packing Assumption

Suppose we introduce the packing assumption as well as the restriction to two distinct states. Then, for a fixed total number of items, the number of items occupying each of the two states at any time must remain fixed across trials and conditions, while the identities of the items, and the matrix of state-occupation probabilities can vary. Furthermore, for any pair of conditions, a and b, we see that by expressing mG(t) as a sum, the invariance property (Sec. A1.2) asserts that

$$\sum_i G_{i,a}(t) = \sum_i G_{k,b}(t), \text{ (all } t), \qquad (23)$$

where $G_{i,k}(t)$ is the cdf for item i in condition k, and

$i = 1,2,\ldots,m$. What more can be said about the cdf's of RTs for individual items in the two-state case?

A8.2 Two States Without Packing Assumption

More can be said even without requiring the packing assumption. First, Falmagne (1968) has observed that the following fixed-point property must apply: If $G_{1,a}(t)$ and $G_{1,b}(t)$ touch or cross at any point $t = t_0$, so that $G_{1,a}(t_0) = G_{1,b}(t_0) = \alpha$, then for any third condition, c, $G_{1,c}(t_0) = \alpha$. That is, $G_{1,c}$ must cross or touch the other cdf's at the same fixed point. (Falmagne has shown that the same property must also hold for the density functions $g_{1,a}$, $g_{1,b}$, and $g_{1,c}$.) If the fixed-point property was satisfied for all items, but the invariance property (Eq. 23) was not, this would give support to a two-state model without the packing assumption. (Note, however, that for the fixed-point property to be tested requires that we observe at least three conditions, and that the cdf's or density functions from two of the conditions touch or cross.)

An additional consequence of limiting the number of distinct states to two is obtained if the approach of the present paper is applied to the cdf's for individual stimuli. Distributional bounds that depend on the mixing probabilities can then be derived. These bounds may prove useful for testing models in which the mixing probabilities are specified as functions of experimental conditions, or for estimating bounds on sets of unspecified mixing probabilities where a binary-mixture model is postulated.

Suppose two states. Then the cdf for any particular item, i, is a binary mixture of two distributions; call them F_1 and F_2. The only differences across items and conditions lie in the mixing probabilities. Let these probabilities be $\alpha_{i,1} = \alpha_i$ and $\alpha_{i,2} = 1 - \alpha_i$ for condition a, and $\beta_{i,1} = \beta_i$ and $\beta_{i,2} = 1 - \beta_i$ for condition b. Hence, for conditions a and b we have:

$$G_{i,a}(t) = \alpha_i F_1(t) + (1-\alpha_i)F_2(t), \text{ (all t, all i)}, \qquad (24)$$

and

$$G_{i,b}(t) = \beta_i F_1(t) + (1-\beta_i)F_2(t), \text{ (all t, all i)}, \qquad (25)$$

Assume that $\alpha \leq \beta$ (where the i subscript is suppressed). Multiplying Eq. 24 by α^{-1} and Eq. 25 by β^{-1}, and noting that $(1-\alpha)/\alpha \geq (1-\beta)/\beta$, it follows that $\alpha^{-1}G_a(t) \geq \beta^{-1}G_b(t)$. Similarly, multiplying Eq. 24 by $(1-\alpha)^{-1}$ and Eq. 25 by $(1-\beta)^{-1}$ it follows that $(1-\alpha)^{-1}G_a(t) \leq (1-\beta)^{-1}G_b(t)$. Combining and rearranging we have the bounds:

$$\frac{\alpha_i}{\beta_i} G_{i,b}(t) \leq G_{i,a}(t) \leq \frac{1-\alpha_i}{1-\beta_1} G_{i,b}(t), \text{ (all t, all i)}. \qquad (26)$$

(In all of the above, we have been assuming that a component distribution depends only on the state of an item, and not on the item itself. It should be pointed out that this assumption is not required for statements like Eq. 26 (above), and Eqs. 29-30 (below), which pertain to cdf's for the same item.) Note that if α and β are allowed to take on any values in the unit interval, then any pair of distributions (G_a, G_b) can be made to satisfy Eq. 26; the bounds are of interest only for testing a binary mixture model with specified α and β, or for specifying allowed ranges for α and β, given an observed pair of distributions.

Thus, if such a pair requires $\alpha/\beta \leq A$ and $(1-\alpha)/(1-\beta) \geq B$ to satisfy Eq. 27, α and β are constrained by

$$\alpha \leq \frac{A(1-B)}{A-B},\ \beta \geq \frac{1-B}{A-B}. \tag{27}$$

A simple graphical method exists for estimating or testing values of the multipliers A and B. If a P-P probability plot (Wilk & Gnanadesikan, 1969) is made in which $G_{i,a}(t)$ is plotted as a function of $G_{i,b}(t)$ (with t the parameter), then the resulting curve must fall between two lines through the origin, one of slope A and the other of slope B.

A8.3 Two Items Plus Packing Assumption

For this situation to be nontrivial there are two distinct states. Equation 23 can be written

$$[G_{1,a}(t)-G_{1,b}(t)] = -[G_{2,a}(t)-G_{2,b}(t)] \tag{28}$$

which says that a change in condition must produce equal and opposite changes in the cdf's for the two items.

Furthermore, on any trial in this situation, one of the two items is in each of the two states, so that for the two conditions, a and b, $\alpha_1 = 1 - \alpha_2 = \alpha$ and $\beta_1 = 1 - \beta_2 = \beta$. Equation 26 then becomes

$$\frac{\alpha}{\beta} G_{1,b} \leq G_{1,a} \leq \frac{1-\alpha}{1-\beta} G_{1,b} \tag{29}$$

and

$$\frac{1-\alpha}{1-\beta} G_{2,b} \leq G_{2,a} \leq \frac{\alpha}{\beta} G_{2,b}. \tag{30}$$

The bounds above (Eqs. 26, 29, 30) apply to the _same_ item in _different_ conditions. Arguments like those above also give bounds for the pair of distributions obtained from _different_ items within the _same_ condition:

$$\frac{\alpha}{1-\alpha} G_{1,a} \leq G_{2,a} \leq \frac{1-\alpha}{\alpha} G_{1,a}, \ (\alpha \leq .5) \tag{31}$$

and

$$\frac{\beta}{1-\beta} G_{1,b} \leq G_{2,b} \leq \frac{1-\beta}{\beta} G_{1,b}, \ (\beta \leq .5) \tag{32}$$

If conditions 29 and 30 are satisfied for values of α and β that are regarded as reasonable, but conditions 31 and 32 are not, this could be taken as evidence against the assumption that the component distributions F_1 and F_2 are independent of i, since conditions 29 and 30 do not depend on this assumption, whereas conditions 31 and 32 do.

REFERENCES

Baddeley, A. D., & Ecob, J. R. Reaction time and short-term memory: A trace strength alternative to the high-speed exhaustive scanning hypothesis. University of California San Diego Center for Human Information Processing, 1970. Report No. 12.

Baddeley, A. D., & Ecob, J. R. Reaction time and short-term memory: Implications of repetition effects for the high-speed exhaustive scan hypothesis. Quarterly Journal of Experimental Psychology, 1973, 25, 229-240.

Bamber, D. Reaction times and error rates for "same"-"different" judgments of multidimensional stimuli. Perception & Psychophysics, 1969, 6, 169-174.

Bradshaw, J. R., & Wallace, G. Models for the processing and identification of faces. Perception & Psychophysics, 1971, 9, 443-448.

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. A. Some variables influencing the rate of gain of information. Journal of Experimental Psychology, 1962, 63, 105-110.

Bricker, P. D. Information measurement and reaction time: A review. In H. Quastler (Ed.), Information theory in psychology. Glencoe, Illinois: The Free Press, 1955.

Donders, F. C. Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868-1869, Tweede reeks, II, 1868, 92-120. Translated by W. G. Koster in W. G. Koster (Ed.), Attention and performance II. Acta Psychologica, 1969, 30, 245-252.

Egeth, H. E. Parallel versus serial processes in multidimensional stimulus discrimination. Perception & Psychophysics, 1966, 1, 245-252.

Falmagne, J.-C. Stochastic models for choice reaction time with applications to experimental results. Journal of Mathematical Psychology, 1965, 2, 77-124.

Falmagne, J.-C. Note on a simple fixed-point property of binary mixtures. British Journal of Mathematical and Statistical Psychology, 1968, 21, 131-132.

Falmagne, J.-C., Cohen, S., Dwivedi, A., & Smith, P. G. Memory-ordered scanning in choice RT. Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1973.

Falmagne, J.-C., & Theios, J. On attention and memory in reaction time experiments. Acta Psychologica, 1969, 30, 316-323.

Harter, H. L. Expected values of normal order statistics. Biometrika, 1961, 48, 151-165.

Hick, W. E. On the rate of gain of information. Quarterly Journal of Experimental Psychology, 1952, 4, 11-26.

Jastrow, J. The time-relations of mental phenomena. Fact & theory papers No. VI. New York: N. D. C. Hodges, 1890.

Lively, B. L. Speed/accuracy trade off and practice as determinants of stage durations in a memory-search task. Journal of Experimental Psychology, 1972, 96, 97-103.

Lively, B. L., & Sanford, B. J. The use of category information in a memory search task. Journal of Experimental Psychology, 1972, 93, 379-385.

Meyer, D. E. Correlated operations in searching stored semantic categories. Journal of Experimental Psychology, 1973, 99, 124-133.

Meyer, D. E., & Schvaneveldt, R. W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 1971, 90, 227-234.

Naus, M. J., Glucksberg, S., & Ornstein, P. A. Taxonomic word categories and memory search. Cognitive Psychology, 1972, 3, 643-654.

Nickerson, R. S. Binary-classification reaction time. Psychonomic Monograph Supplement, 1972, 4, No. 17 (whole no. 65).

Sternberg, S. Retrieval from recent memory: Some reaction-time experiments and a search theory. Paper presented at the meeting of the Psychonomic Society, Bryn Mawr, August 1963.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Sternberg, S. Retrieval of contextual information from memory. Psychonomic Science, 1967, 8, 55-56. (a)

Sternberg, S. Two operations in character-recognition: Some evidence from reaction-time measurements. Perception & Psychophysics, 1967, 2, 45-53. (b)

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 1969, 4, 421-457. (a)

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), Acta Psychologica 30, Attention and performance II. Amsterdam: North-Holland Publishing Company, 1969. Pp. 276-315. (b)

Sternberg, S. Two invariances in retrieval of contextual information from memory. Paper presented at the annual meeting of the Eastern Psychological Association, Philadelphia, Pennsylvania, April 1969.

Taylor, D. H. Two dimensions for reaction-time distributions. Nature, 1965, 206, 219-220.

Theios, J. Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.), The psychology of learning and motivation, vol. VII. New York: Academic Press. In press.

Theios, J., & Smith, P. G. Can a two-state model account for two-choice reaction-time data? Psychological Review, 1972, 79, 172-177.

Theios, J., Smith, P. G., Haviland, S. E., Traupmann, J., & Moy, M. C. Memory scanning as a serial self-terminating process. Journal of Experimental Psychology, 1973, 97, 323-336.

Welford, A. T. What is the basis of choice reaction-time? Ergonomics, 1971, 14, 679-699.

Wilk, M. G., & Gnanadesikan, R. Probability plotting methods for the analysis of data. Biometrika, 1968, 55, 1-17.