HUMAN INFORMATION PROCESSING: TUTORIALS IN PERFORMANCE AND COGNITION.

Edited by B. H. Kantowitz (Purdue University), 1974. Hillsdale, NJ: Erlbaum

4

ISSUES AND MODELS CONCERNING THE PROCESSING OF A FINITE NUMBER OF INPUTS[1]

J. T. Townsend
Rockefeller University

INTRODUCTION

Eighty years ago or so the title would probably have been something like "An Attempt at Clarification of Proposed and Counter-Proposed Properties of Consciousness." The revolution in information, computer, and communication sciences (à la Shannon, Von Neumann, and Wiener) and the growing vapidity of behaviorism led in the 1950's and 1960's to a resurgent interest in centralism, that is, a renewed interest in the study of properties of the mind, as opposed to the exclusive peripheralistic study of stimulus-response correspondences. The terminology of the "new" psychology has been almost completely drawn from the above disciplines, but very recently investigators have become less self-conscious about employing mentalistic terms and concepts that would have brought applause from the structuralists and phenomenologists of yesteryear (see, e.g., Boring, 1950). Although skirting the interesting questions involving possible fruits and pitfalls of this trend, I might simply note that this chapter continues, after the introduction, in the recent tradition of use of information-processing argot.

In the present approach, we go inside the human *qua* information processor to as shallow or deep a level as we care to, and seize upon an abstract black box residing there in the processing chain. We shall think of our black box as engaging in the processing of some finite number of inputs (or elements). Words like "processing"

[1]I am grateful to W. K. Estes for the opportunity to spend the year 1972–1973 at Rockefeller University under Grant GM16735 from the National Institute for General Medical Sciences. During that time, much of the writing of this chapter was completed. I am indebted to W. K. Estes, B. Kantowitz, D. A. Taylor, D. Snuttjer, and L. Paul, as well as several helpful anonymous reviewers, for their comments on an earlier version of this manuscript. I benefited from conversations with W. K. Estes, R. N. Roos, D. A. Taylor, Joan G. Snodgrass, and D. Vorberg and from discussions with H. Schultz, J. Wandmacher, and D. Albert concerning the issues investigated herein.

and "elements" are purposely kept abstract to enhance the generality of our discussion. And the view we take is that we should endeavor to comprehend the possible workings and characteristics of fairly atomistic levels of processing, not necessarily before consideration of more complex levels (e.g., interaction of several black boxes), but at least concomitantly with the latter. The name we shall use for our black box is "system" albeit it may be a very small (sub) system relative to the total information-processing system in which it operates; the important matter is that we may conceive of it as a unit in its operations on the inputs. Our system could be as large as consciousness or as small as a set of hypothetical neural nets in the retina, depending on our level of analysis and the type of inputs under scrutiny.

Our concentration in this chapter will be on four dimensions or aspects of the system, posed, as they have been implicitly or explicitly in the literature, as dichotomies. It is convenient to introduce the issues in a different order than they will appear in the subsequent sections.

Independent vs Dependent Processing

The first of these is less molar than the others, but likely to be of interest as knowledge of processing in various kinds of experimental settings improves. It is the question of whether processing of the separate elements is statistically independent or not. To be more precise, assume that two elements are to be processed, and that the total time from when the processing system begins to work on the elements to the completion of each element is recorded. Then, if there is no interaction or dependence between the two total completion times, these times are said to be independent. If knowing that one of the elements took a long time to be completed makes it more likely that the other took a long time as well, a positive dependency exists. If the same knowledge increases the chances of a short total completion time by the other element, a negative dependency is said to occur.

Now, consider a situation where the processing system is allowed some fixed time t, through limiting the stimulus exposure duration for example. It may then be asked whether the knowledge that one element was completed during that time affects the likelihood that the other was completed. The answer to this question follows directly from the dependency relations, as discussed in the foregoing paragraph, between the unconstrained total completion times. Independence or dependence of the former implies independence or dependence of the latter. For example, assume that long unconstrained completion times of one of two elements are associated with long completion times of the other and that short times of the one are associated with short times of the other, that is, a positive dependency exists. Then, in the limited duration case, if it is known that one of the elements is completed the chances are increased that the other was also finished. Conversely, if it is known one of them did not complete processing, it is also likely the other did not. Recent examples of an experimental context wherein dependence has been investigated are found in Estes and Taylor (1966) and Wolford, Wessel, and Estes (1968). These studies used the detection paradigm, (Estes & Taylor, 1964) wherein the subject attempted to detect which of two target letters was embedded in a random array of noise letters on each trial. The exposure duration was limited, and it may be supposed the processor involved in comparing display with target letters also had only a limited time to perform its task. The variable of interest here was the manipulation of the number of repetitions of the target letter presented on a given trial. This technique permits a test of the independence assumption.

Self-terminating vs Exhaustive Processing

Another issue, pertinent to experimental trials where some proper subset of the inputs (that is, fewer elements than are input to the processor) or elements contain information sufficient to enable the subject to make the required response, is whether the system is capable of cessation of processing when that subset is completed (self-termination) rather than having to nevertheless do all the elements (exhaustiveness). One of the first mathematical processing models where self-termination was assumed is found in Estes and Taylor (1964), and attention has since been brought to bear on it in other recent papers (e.g., Atkinson, Holmgren, & Juola, 1969; Sternberg, 1966; Townsend & Roos, 1973). In the first of a series of papers employing the detection paradigm by Estes and his colleagues, the Estes and Taylor (1964) "serial processing model" was based on the assumptions of geometrically distributed durations of stimulus availability (e.g., an icon), serial (one-at-a-time) comparisons of target and display letters, and termination of processing when (and if) the target letter was located. Although this specific model has since met with severe difficulties, it contributed to elucidation of these and similar data at the time (see, e.g., Townsend, 1968).

Serial vs Parallel and Limited vs Unlimited Capacity Processing

The remaining two issues are probably the most important of the four and are considered together in our introduction; they are rather closely linked historically and have generated some mutual confusion even quite recently. The first is concerned with the question of capacity of a system—whether processing efficiency, in terms of speed or accuracy, is degraded or is unaffected when increased numbers of elements are input to the system.[2] This issue currently goes under the rubric "limited versus unlimited capacity." The second is whether the capacity of the system is applied to the inputs one at a time, or whether it is deployed simultaneously to all the inputs. For a large variety of contemporary experimental paradigms, the latter becomes the serial versus parallel processing issue.

The broadest and most critical arena of investigation and contention with regard to these two matters has always been consciousness itself. That is, can consciousness act on more than one element (thought, image, sensory impression, etc.) at a time (the serial-parallel issue), and when it has to deal with an increased amount of material, is it less effective (the capacity question)? The question of whether the mind (via

[2]The problem of effectiveness or efficiency immediately ramifies into questions as to what level of processing is under observation. For instance, are we concerned with effectiveness per unit of material or with total processing time for all the material, or some other level? These issues will be addressed in the section on capacity.

consciousness) is capable of engaging in more than one cognitive activity at a time can be traced back at least as far as Plato and Aristotle. But probably the first directly relevant experimental evidence came from the personal equation difficulties met by puzzled astronomers and later by the budding experimental psychology of the nineteenth century. The situation was simply this: A method of measuring positions and relative velocities of stellar bodies consisted of determining the amount of time necessary for a body to cross a reticle in an optical instrument. The time was calculated from (*1*) the number of complete tick-intervals from a nearby clock heard by the astronomer during the transit time plus (*2*) the perceived position of the body relative to the reticle, between the two final clicks. That is, the proportion of the interval between two marks corresponding to the body's position at the two clicks, that was in front of the reticle (e.g., if the movement was left-to-right, the proportion of the final interval to the left of the reticle was the required measurement), was the final proportion of the time interval added to the "whole" time intervals.

The dilemma was that different astronomers obtained different estimates of the times, and we are most of us acquainted with the famous story of an unfortunate assistant immortalized by being fired (his name was Kinnebrook) for producing transit times larger than his employer (who shall here remain nameless in retribution for this ignominious action). The difference between two observers subsequently came to be called the personal equation. Without getting involved in all the technicalities and subtleties of the matter, we may note that these astronomical observations involved the processing of information from two inputs, namely, the eyes and the ears. Computation of the "exact" transit time required knowledge of the temporal point between two ticks of the clock at which the stellar body crossed the reticle. If two observers differ in when they think the body crossed the reticle, we may infer that they differ in the time necessary to compare their visual with their auditory impressions. This suggests that in neither case are the comparisons simultaneous, and that the delay is simply greater for one observer than the other.

If it turned out that all the delay and differences in delay could be placed in the sensory channels themselves, then perhaps the delay and allied factors would not be of much psychological interest. However, it was found that predisposition and expectation affected the results, especially in the famous complication clock experiments of von Tchisch around 1885. When decision processes can asymmetrically influence tasks where it would be to the subject's advantage always to utilize information from the auditory and visual channels simultaneously, this suggests the possibility that in fact the cognitive processing of the information may not be able to proceed simultaneously (i.e., in parallel), or that if it does that there may occur a degradation in performance (the capacity question) as compared to only one channel being monitored cognitively at any given moment. Although an entirely satisfactory theory, from a modern point of view, has never been set down for these phenomena, we might note the recent reawakened interest in such topics (e.g., Stemberg & Knoll, 1972).

Another empirical problem, related in at least its systemic properties to the

preceding, was illustrated by the philosopher Hamilton (1859), who convinced himself by tossing variable numbers of dice that consciousness is capable of simultaneously apprehending a maximum of 4 to 6 objects. Thus, 6 came to be thought of as the range of capacity of visual apprehension (or attention as we would probably say now). Hamiton himself believed that when attention was focused on fewer objects, the apprehension was superior to when it was focused on more objects. This is an example of a parallel, limited capacity system, since it can work on more than one object, but resolution suffers to some extent when this is done.

In his later work at least, the prodigious Wundt appears to have accepted this alleged limited capacity of parallelism, as well as the ability of the mind to "focus up or down" on larger or smaller numbers of objects, respectively. On the other hand, certain other investigators, such as Hylan, (1903), offered evidence for the proposition that attention was limited in capacity not because of a parallel spreading of attention over the set of objects to be apperceived, but because there occurs an all-or-none fluctuation of attention from one object to another. The latter was, of course, a serial type of mechanism. The fact that two early experimental psychologists argued, one for a serial type of limited capacity system and the other for a parallel limited capacity type, already suggests the logical independence of the serial-parallel and the capacity issues. In fact, the gist of our argument on all the issues in the sections to follow is that since a processing system can be made up by selecting a stance on each of the issues independently, that these *are* logically separate and independent. For instance, we could build a system that is serial, unlimited capacity (in one of the precise senses to be developed below), self-terminating, and is characterized by independent processing of the separate inputs. We could then change any or all of the stances on each of the issues (e.g., change serial to parallel) without logical or mathematical violation. An entirely distinct question is whether the system we build (or the model of this system) is realistic, that is, whether we expect to encounter such systems in nature. The problem of "intuitiveness" and "reasonableness" has, it seems, caused confusion as to the logical distinctions involved.

To proceed with our brief longitudinal view of parallel and serial hypotheses about apprehension, chinks in our knowledge continued to be filled in over the years, especially in the counting of objects in brief visual displays (see, e.g., Hunter & Sigler, 1940). However, the actual possible processing mechanisms and their properties appear to have been largely neglected. In any case, the whole story was reopened by Sperling's (1960) memory-apprehension study, in which an important dissection of processing stages was suggested. Some of his later results, (Sperling, 1963), where he found that increasing the display time of several letters (with a preceding and postceding visual noise mask) led to a linear increase in the number correctly reported, were typically taken as supporting the notion of serial processing. Furthermore, a linear increase of RT in short memory search unfortunately has been assumed to falsify independent parallel models and to provide strong support for serial models (Sternberg, 1966). That these conclusions are wrong or at best misleading (Town-

send, 1969a, 1969b, 1971b, 1972a) will be shown in the section on capacity. At any rate, the concept of parallel, limited capacity systems seems to have been forgotten for a number of years.

Broadbent (1958) unveiled a theoretical structure that allowed collation of a substantial body of experimental literature and that was seminal in its influence on later developments. Although not put directly in terms of parallel vs. serial processing, the hypothetical filter responsible for selection of material for further processing appeared to function in a serial-like fashion when the input exceeded the capacity of the "bottleneck."

Around 1969, several investigators rediscovered limited capacity parallel systems. For example, Erikson and Spencer (1969) used the analogy of a zoom lens to illustrate a (parallel-limited capacity) system that processes everything simultaneously but with better resolution when focused on a smaller field. Too, there is an implicit acknowledgement of the possibilities of limited-capacity parallel systems making predictions similar to typical serial systems in A. Treisman's review on selective attention (1969, p. 292). Atkinson et al. (1969) pointed out a type of parallel model mathematically identical to a type of serial model, and the writer gave a paper that discussed some rather large classes of probabilistic parallel and serial models that are equivalent (Townsend, 1969a, 1969b, 1971b). Soon thereafter, Gardner (1970, 1972) contributed an innovative discussion of these and similar issues. In a similar vein, Corcoran (1971) discussed a nonprobabilistic limited capacity parallel model.

Christie and Luce (1956) appear to have been the first to approach the parallel vs. serial issue from a mathematical point of view. Thomas (1969) has suggested some nonparametric tests that may be utilized in certain cases. The writer's own work in this area (Townsend, 1969a, 1969b, 1971b, 1972a, 1972b, 1972, 1973) is briefly summarized and related to other efforts in Townsend (1974). This brings us into the seventies and completes our necessarily laconic historical groundwork.

Goals of the Present Approach

It is our purpose to formulate the foregoing issues in an informal but hopefully rigorous fashion in the main text. It is unfortunately true that confusions and incorrect conclusions, especially concerning tests between parallel and serial processing, are still to be found in the psychological literature. The overall goal is to facilitate logical and systemic distinctions among these four issues and to provide a framework conducive to further theory building and more accurate testing. Within this larger goal is the aforementioned aim of convincing the reader, if he is not presently of this opinion, that the four issues are indeed logically distinct. Without detailing the mathematics presented elsewhere, we also shall try to gain some insight into why many of the experiments designed to throw light on seriality vs. parallelism have not accomplished this purpose. In the process of our development we shall have occasion to construct sundry models and refer to a number of experimental applications. The applications will often cluster in the areas of visual and memory search, since these are of special interest to the writer, but the concepts will hopefully have more extended reference than to these alone.

THE SERIAL VS PARALLEL PROCESSING ISSUE

Definitions of Parallel, Serial, and Hybrid Systems

To begin with, we wish to distinguish between systems that exist in the real world and the models of those systems which are abstract fabrications.[3] In a very real sense, there is no difficulty at all in telling systems apart, or there would not be if all the workings were completely open to our inspection. If two systems work differently or are constructed differently, then they are just so and that is all one may say about the matter.

The difficulties arise, of course, when we wish to make inferences about the inner workings of the system for some reason but cannot actually inspect these, because it is too difficult or because society frowns on indiscriminate slicing into certain black boxes of interest to psychology. Then it may be that a model, hypothesized to describe these workings and employed to predict input-output behavior of the system, may be similar or even identical to a model that began by picturing the system in a drastically different way. That the models may be equivalent does not mean that the systems they refer to are equivalent, but only that they may act in an equivalent manner, when we can observe what happens at the input (stimulus) and output (response) stages. This will become more clear as we progress. First, two informal definitions of parallelism and seriality are given.

Definition I. A *Serial System* is a system that processes elements one at a time, completing one before beginning the next.

Definition II. A *Parallel System* is a system that begins processing elements simultaneously; processing proceeds simultaneously, but individual elements may be completed at different times.

Suppose processing begins in either type of system at time $t = 0$ on elements (a) and (b) and that the total time (z) spent by the system actually processing (a) if it is completed first is given by z_{a1} and by z_{a2} if it is completed second. Similarly z_{b2} is the total time spent actually processing (b) when it is completed second, and z_{b1} is the comparable time if it is completed first. Figure 1a shows a schematic of these quantities when (a) happens to be completed first.

There are two features of Fig. 1a that are indispensable in defining seriality vs. parallelism. First note that in a parallel system, up until one of the elements is completed, there can be no prior designation of the *order of completion* [in Fig. 1a, (a) is completed first], but there can be in the serial system, since on each trial one and only one of the elements is selected to be first. Another way of putting this is to observe that the *serial* system can select (a) to be processed first with some arbitrary probability p; (b) of course is processed first with probability $1 - p$. The variable or parameter p need bear no relation whatsoever to the amounts of time spent processing

[3] It is convenient here to use the term "system" to refer to actual, physical entities. One may occasionally find "system" used in an abstract sense, more in the way we employ "model" in the present treatment.

(a) and (b) (the z's in Fig. 1) once they are in the processor. However, the order of completion under *parallel* processing must depend on the relative rates at which (a) and (b) are being processed. These concepts will be made explicit below.

Notice also that in the parallel system z_{b2} is necessarily larger than z_{a1}, but not in the serial system. This, too, is an important distinction between a parallel and serial system: The amount of time spent by the parallel system processing the element that is completed last is always equal to the total amount of time from the beginning to the

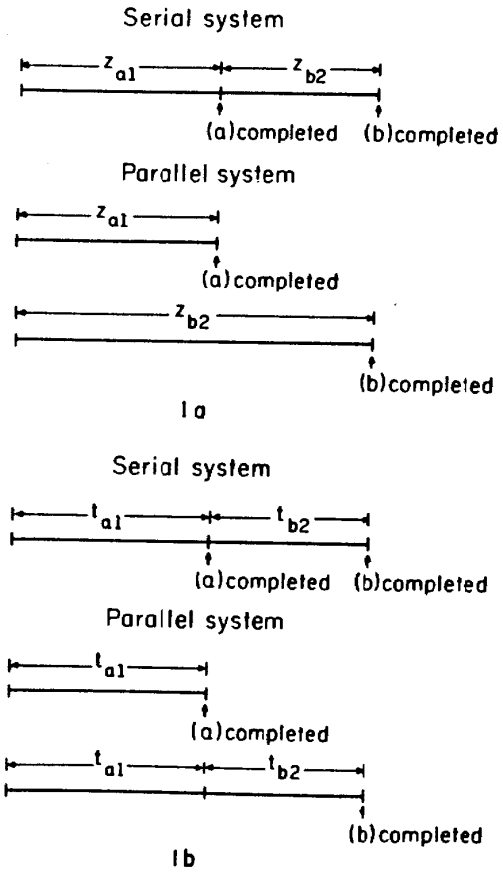

FIG. 1. 1a.—Schema showing total time spent processing (a) and (b) in a serial or parallel system when (a) is completed first. 1b.—Schema showing intercompletion times in processing (a) and (b) in a serial or parallel system when (a) is completed first.

completion of processing. But in a serial system the processing duration on the last element obviously does not include that spent on the earlier ones.

It often turns out to be helpful to discuss system performance or model predictions in terms of intercompletion times, that is, the time from the point at which one element is finished to when the next is completed. If (a) is done first (in either system), the first intercompletion time [from $t = 0$ to when (a) finishes] is called t_{a1}, and if (b) is first, the first intercompletion time is t_{b1}. Similarly, the time interval from the completion of (a) to the completion of (b) is just t_{b2} and alternatively, the duration from the instant when (b) is completed first to when (a) is completed is called t_{a2}. Figure 1b shows the events of Fig. 1a but in terms of intercompletion times. Hence, in a serial system, $z_{a1} = t_{a1}$, $z_{b2} = t_{b2}$, but in a parallel system, $z_{a1} = t_{a1}$ but $z_{b2} > t_{b2}$, in fact, $z_{b2} = t_{a1} + t_{b2}$. The crux of the difficulty in experimentally testing between parallel and serial systems is that in being confined to input-output relationships, all we actually observe, even under optimal conditions, are t_{a1}, t_{b2} or t_{b1}, t_{a2}. The relationship of the intercompletion times to the actual processing times must be inferred. When parallel and serial models produce identical predictions for the t's, they represent systems that are behaviorally indistinguishable.

We are emphasizing parallel and serial processing in this exposition and neglecting other types because of want of space, but a few cursory remarks may be pertinent to "hybrid" systems and models.

Definition III. A *Hybrid System* is a system that processes in neither a parallel nor a serial manner (Townsend, 1969a, 1971b, 1972a).

Hybrid models have been of limited current theoretical interest, probably due in part to the difficulty in testing them experimentally. As remarked earlier, the quite special case of seriality vs. parallelism is difficult enough to discriminate experimentally. Among such hybrid models are those that represent processing as being serial part of the time and partially parallel within trials. That is, during part of a trial the elements may be worked on simultaneously, and during another part of the trial they may be worked on one at a time, each being completed before the next is begun. On the other hand, another hybrid model might suppose that on some proportion of trials processing is entirely serial, and on the remainder, it is entirely parallel. Of more recent theoretical interest have been models that can be interpreted either as parallel or hybrid (Townsend, 1972a). Rumelhart's (1970) parallel model is a special case of this class of models. In its hybrid interpretation, such a model presumes time-sharing, in that one element is worked on at a time but that element may be abandoned while the processor works on another element for awhile, after which time the processor may return to the "first" element, and so on. Strict serial processing demands that each element be completed before processing is begun on the next. We may note that in certain experimental contexts, the distinction between time-sharing and seriality may be supererogatory. In a situation where a subject must attempt to monitor two different messages transmitted to separate ears, the elements can be conceived of as the separate (ear) channels. In time-sharing or serial processing with respect to the two channels, attention switches back and forth in an all-or-none manner between channels, but a parallel processor places attention on each channel simultaneously.

A last remark concerning terminology. We shall refer to stage 1 as the state of a system from the start of processing until the first element is completed. Similarly, stage 2 is the state of a system from the completion of the first until the completion of the second, and so on; stage i is the state of the system from the completion of the $(i-1)^{th}$ element to the completion of the i^{th}. At stage $n+1$ a system is once again at rest.

Practically our entire discussion will be based on the concept of *rates* of processing and *order* of completion of the elements being processed.

The concept of order of completion is, of course, natural and well-defined. Rate, however, is more fuzzy. We shall take the tack here of establishing rate as a variable expressed in specific cases as a positive number. In certain simple distributions on processing times, these rates will appear as actual parameters in the models. The rates are then related in a very natural way to mean intercompletion times and overt processing durations. However, "rate" may also be *defined* in terms of selected statistics.

For example, if the mean (expected) time to completion for (a) is $E_a(t)$, then we could *define* the rate of processing of (a) as $u_a = 1/E_a(t)$; in models based on exponentical intercompletion times, this relationship appears automatically.[4]

First, consider the serial models. Let u_{a1} = processing rate for element (a) when it is processed first, (i.e., during stage 1) and u_{b1} the processing rate for element (b) when it is processed first. Similarly, let u_{a2} be the processing rate for element (a) given that it is processed second (stage), after (b) and u_{b2} the rate for (b), when it is second, after (a). Let p be the probability that (a) is processed first. It is extremely important to note that p is, in general, independent of the processing rates in serial models. That is, the serial system may be viewed as making a selection among the possible processing orders on each trial according to some arbitrary but fixed probability distribution. In the case of two elements, the probability distribution is just p for order (a) first, (b) second, and $1-p$ for order (b) first, (a) second.

Turning to parallel models, let v_{a1}, v_{b1} be the processing rates for elements (a), (b) respectively, during stage 1. In probabilistic parallel systems, either (a) or (b) may by chance be completed during stage 1. If (a) has this distinction, then unless the system terminates processing for some reason, (b) will complete its processing at the rate v_{b2}. If (b) is finished first, (a) will be completed second at the rate v_{a2}. Unlike the serial models, we expect the order of processing to depend on the relative rates of (a) and (b) during stage 1, v_{a1}, v_{b1}. If v_{a1} is much larger than v_{b1} it may typically be predicted that (a) will finish first more often than will (b) and vice versa.

[4]Typically, quantitative statements concerning the u, v, and p variables are exact only in the case of models with exponential intercompletion times, that is, for models whose intercompletion times (the t's in Fig. 1) can be described by density functions of the form $u \cdot \exp(-ut)$. However, they may be approximations in other cases. One of the aims of this chapter is to indicate the existence of parallel and serial models that *are* mathematically identical and discuss the related intuitions that hold exactly for certain classes of models and approximately for others. Precise delineation of the mathematical detail is beyond the scope of this chapter. With regard to approximation, it is possible to make two models, one serial and the other parallel, give identical predictions at, say, the level of mean RT, although the full models are not equivalent on all possible statistics (see Townsend, 1972a).

Of the four issues raised in the introduction, that of parallel vs. serial processing is pivotal. The reason is that the interpretation of the other issues assumes different forms in parallel as distinguished from serial processing.

We do not have sufficient space to reiterate in depth previous results on parallel and serial models that are equivalent, and therefore indistinguishable by behavioral data. But, the resultant intuitions that relate to ideas of rates and order of processing will help elucidate the present information structure and the four issues in particular.

A Coin Tossing Example

Let us consider a coin tossing analogy. The analogy corresponds quite closely to the actual modeling situation, where parallel-serial equivalence can appear. The main difference is that in the coin case, we use discrete time but in the typical real modeling case, continuous time. One other important difference will be pointed out below.

Consider as the serial analog a coin tossing mechanism that tosses each of five coins, one at a time. It must obtain a head for each coin before going on to the next, and we assume each coin may be identified by some mark placed on it. For a given experiment, the mechanism sets up a probability distribution on the different processing (tossing) orders for the five coins. Arbitrarily assigning the letters a, b, c, d, and e as names for the coins, then any order of tossing, for example, <d, a, e, c, b>, will have some likelihood of being selected. It is assumed that there are no orders that never occur. Further, we assume that each coin may have its own, possibly unique, probability of coming up heads.

As the parallel analogue, simply assume that all five coins are tossed simultaneously, but each continues being tossed until it personally achieves a head. That is, all five coins are tossed, those coins attaining a head are removed, the remaining coins are again tossed simultaneously, and so on. The basic unit of time is that fixed interval required for one simultaneous toss.

Now, in this discrete case, the parallel system can never quite be exactly the same as the serial because on a given toss (the fixed duration of a toss defining a unit of time) more than one head can come up. In the serial case, only one head can come up during the fixed unit of time. Thus, the exact probabilistic description, and therefore the sampling distribution, cannot be the same. However, when the basic time unit or interval is squeezed down smaller and smaller, in the direction of attaining continuous time, this difference disappears for many interesting models.[5] In any case, we may compare the parallel and serial coin mechanisms in overall behavior to continuous-time parallel and serial systems.

First, it is clear that if the same coins were used first in the parallel tossing system

[5]We can not go into detail here but, for example, parallel models with intercompletion times that are independent across the elements at any given stage fulfil the condition that for very small intervals the probability of two or more completions becomes very small, thus making possible serial-like behavior. It should be noted that independence referred to in this footnote is *not* independence of *successive* intercompletion times or even necessarily of overall completion times (see "Independence vs. Dependence Issues").

and then in the serial system, the time in toss units until the first head appears will be significantly shorter in the parallel than in the serial case. Hence, to make this time roughly the same, the probabilities (of a head) of the parallel coins will have to be made smaller, or equivalently, the serial probabilities will have to be made larger. At each succeeding stage the total behavior of the parallel coins has to correspond to the behavior of the toss of a single coin in the serial system; either the serial coins, the parallel coins, or both must undergo some adjustment. The important point is that the summed activity of the parallel system must equal the individual activity of the serial system for each remaining number of parallel and serial coins.

The other critical aspect of the analog is the order of completions. Somehow, the probability distribution on these orders used by the serial system has to correspond to that obtained by chance in the parallel system. It is quite apparent that the order of finishing in the parallel system will depend on the various magnitudes of the head-probabilities of the coins. In the serial case, on the other hand, the actual order is entirely established by the probability distribution on the orders, independently of the coins' probabilities of turning up heads. Hence, we can try to make the serial system mimic this aspect of the parallel system's behavior by adjusting the order probabilities so that, say, the likelihood that coin "d" gets a head third equals the probability that it does so by chance in the parallel system. If its parallel probability of getting a head is intermediate among the others, this serial probability of being third would be relatively high. This probability could be easily arranged in the serial system. Conversely, the parallel head-probabilities could be adjusted to match the likelihoods associated with the order probability distribution in the serial system.

Parameter Mappings That Can Render Parallel and Serial Models Equivalent

Although we shall return from time-to-time with intuitions based on the coin example, let us turn to the continuous time case with our serial, u_{ai}, u_{bi}, and parallel, v_{ai}, v_{bi} rates ($i = 1, 2$, the stages) for elements (a) and (b); recall that p is the probability that (a) is chosen to be processed first in the serial system.

If equivalence is possible between serial and parallel *models*, then there must exist transformations or *mappings* that express the serial parameters in terms of the parallel parameters, and vice versa. Further, these mappings must ensure equivalent predictions of all possible statistics by the two models. The parameters of the serial model we will continue discussing are the rates u_{ai}, u_{bi}, and p. The parameters in the parallel model are likewise v_{ai}, v_{bi}. If such mappings can be found then, for example, typical values may be assigned to the parallel parameters and the mappings used to find the corresponding serial parameter values. This serial model then will make predictions identical to those of the parallel model.

In probabilistic continuous-time systems, the intercompletion times when more than one element is being processed in parallel will typically be shorter, assuming fixed processing rates, than if only one is being processed. In fact, we can represent this speed-up by summing the two parallel rates, $v_{a1} + v_{b1}$. Hence, to make the parallel and serial behavior alike up until the first element is completed, it must be the case that $u_{a1} = v_{a1} + v_{b1}$ when (a) is processed first in the serial system, and $u_{b1} = v_{a1} + v_{b1}$ when (b) is processed first in the serial system. That is, the sum of the processing rates for the two elements undergoing parallel processing must be the same as the individual serial rates during stage 1. But observe that *both* u_{a1} and u_{b1} must equal $v_{a1} + v_{b1}$ in order for the behavior of the parallel and serial systems to look the same during the first stage of processing. This result suggests what turns out to be an important fact concerning the present parallel and serial models: The serial models are more general because the processing rate at stage 1 can depend on which element is being processed first. It is comparable to the greater generality of the serial coin tosser: the average waiting time to the very first head depends on which coin is tossed first. This means that the subset of trials on which coin "b," say, is tossed first will in general have a different frequency distribution and mean than will that subset of trials on which coin "e" was tossed first. The parallel tosser, on the other hand, will have the same probability distribution on every trial, that representing a series of simultaneous tosses of coins with different probabilities attached. Similarly, the continuous-time parallel system and its model (or models) will have a description of the first intercompletion time (that is, waiting time to the first completion) that depends only on $v_{a1} + v_{b1}$ on every trial and hence cannot show the diversity of behavior for this event that the serial system and models can.

The actual situation is somewhat more complex than this, since the serial systems can make the processing rates dependent not only on the element processed but also on the complete order of processing of the various elements on a given trial. Thus, in the general case of n elements there are $n!$ different serial processing rates for stage one, but only one for the parallel model. However, as processing evolves during a given trial, the parallel system gains more and more freedom, since it can make its rates depend on the previously processed elements and their order (on the same trial). During the final stage of processing the parallel system attains freedom equal to that of the serial system. To see this, observe that there are n different elements that can be processed last and $(n - 1)!$ different orders in which the other $n - 1$ elements could have finished, so there are $n!$ different rates for the last stage in the parallel as in the serial system (the serial system can have $n!$ different rates at every stage). It is reasonable that a serial system can predetermine processing order and have the potential of making processing rates depend on these orders, but the order of processing evolves by chance in probabilistic parallel systems and, hence, the parallel rates at stage 1, for example, cannot depend on what element finishes fourth. A parallel system could only attain this aspect of generality processed by serial systems by the singular property of prescience.

It would appear possible to get around this constraint by taking probability mixtures of parallel systems. That is, if $n!$ parallel systems are available and one is chosen with some probability on each trial, then one could have $n!$ different rates during stage 1 as in the serial systems. However, not only cannot this "mixed" parallel model be equivalent to the serial model, in general, but the serial model cannot generally be equivalent to the parallel. The main proviso here is that if the serial distributions are of a certain class, for example exponential, gamma, etc., then those distributions determining processing time frequencies for each element processed in

parallel must be the same, and vice versa. Probability mixtures of members of a certain class of distributions are not generally of the same class as the members. To see that the mixed parallel model still cannot typically mimic the serial, observe that if we consider a single order, say (a) is first, of processing two elements, then the processing time for (a) will be given by a distribution depending on a single rate, but in a mixed parallel system, it would be given by a probability mixture of distinct-rate distributions. For similar reasons, the serial models cannot mimic mixed parallel models, except in degenerate cases or in cases where the underlying "component" distributions are assumed to be different in the parallel and serial models (see Townsend, 1972a, p. 178).

Returning to ordinary parallel systems, we see that by setting $u_{a1} = u_{b1} = v_{a1} + v_{b1}$, and $u_{a2} = v_{a2}$, $u_{b2} = v_{b2}$ the parallel and serial models look alike because the rates governing observable input-output behavior are alike. We are completely free to make $u_{a2} = v_{a2}$ and $u_{b2} = v_{b2}$ because the parallel models' freedom at the last stage, here stage 2, equals 2, the same as the serial case. However, there is one other constraint that must be present, one relating to order of processing. Roughly, the likelihood that (a) is completed first in the parallel system should be related to the magnitudes of the rates v_{a1} and v_{b1}, that is,

$$\text{Probability}\left[(a) \text{ is completed first}\right] = \frac{v_{a1}}{v_{a1} + v_{b1}}$$

But in the serial case this probability is just given by p. So for the orders to be comparable we require that $p = v_{a1}/(v_{a1} + v_{b1})$. By use of these parameter mappings, we can retrieve serial models that give behavior expected of parallel systems, and parallel models that yield behavioral predictions reminiscent of serial systems. For instance, the mapping above immediately tells what the serial parameters have to look like to mimic parallel behavior. The reverse mappings, parallel parameters written as functions of the serial parameters, are just $v_{a1} = p \times u_1, v_{b1} = (1-p) \times u_1, v_{a2} = u_{a2}, v_{b2} = u_{b2}$; note that we use u_1 to stand for the common values of u_{a1} and u_{b1}. It is important to notice though, that the direction of the mapping, parallel-to-serial or serial-to-parallel, does not by itself specify whether the models are describing activity *typically* associated with parallel as opposed to serial processing (or with either, for that matter). For example, the actual numbers for the u's in a given case may already be such as to predict behavior commonly thought of as parallel; in this case when the v's are obtained via a serial-to-parallel mapping, the resulting values will be those "typical" of parallel models and, hence, of parallel systems. However, if the numerical values of the u's are those expected of serial systems, then the v's will be specified parameters of a parallel model that yield serial-like predictions.

Can people process in parallel? Here, as in the applications section, we see that the firm establishment of one or the other way may be quite tricky. Further, the proper answer to the question will almost certainly be positive for some types of processes and negative for others. Surely the brain is capable of doing a number of complex things simultaneously. For instance, the driving of an automobile while engaged in sophisticated problem solving is obviously parallel processing, albeit by probably

more or less *functionally* distinct subsystems. Another aspect is a tie-in with organization or Gestalt properties of the stimulus or task. In watching a ballet, it appears considerably easier to "pay attention" to all the dancers simultaneously when they are involved in rhythmic movements with a high degree of similarity and symmetry across individuals and continuity in time than when the dancers are performing sequences less integrated into a holistic Gestalt.

THE SELF-TERMINATING VS EXHAUSTIVE PROCESSING ISSUE

Basic Concepts

As shall become increasingly apparent, the parallel-serial distinction, being directed to how the available processing resources are deployed to the inputs, serves as a basic descriptive agent and largely determines the approaches to the other issues.

There are many occasions in perceptual and memorial experiments where the information sufficient to make a correct response is embedded in only part of the total stimulus pattern presented to the subject. In this case, if the subject's information-processing system is capable of a cessation of processing when that information has been obtained, we say it is self-terminating. Otherwise it is exhaustive: It must always process all the information available. Of course, some trials by their very nature may require exhaustive processing, but this does not enable us to call the processor exhaustive. It is the incapability of self-termination that is defined as exhaustiveness. An experimental paradigm on memory search that involves trials where self-termination can occur, and others where processing must be exhaustive, with respect to the basic input units (in this case digits), is given in the Sternberg (1966) paper, and a visual span of apprehension analog to this design is given by Atkinson et al. (1969). The "detection paradigm" formulated by Estes and Taylor (1964) and mentioned in the introduction, is basically a two-alternative forced-choice design that allows for self-termination on all trials.

General Memory and Display Search Predictions for Two Elements

Let us imagine a Sternberg (1966) (or Atkinson et al., 1969) task where on every trial either one or two letters are presented to a subject for memorization, followed by presentation of a target letter. The subject responds positively if the target is in the memory and negatively if not. On one-half of the trials, the target is in the memory set; in the other half, it is not. Accuracy is high and RT is the main dependent variable. RT includes other components besides that of the processing times, of course, but we must ignore those here. Before giving serial and parallel predictions for self-terminating and exhaustive-processing times, we need to recall some simple relations between processing rates and mean processing times.

The situation is especially nice for serial models, since the average of the sum of individual processing times is the sum of the individual (element) averages. If we take the relation of the mean processing time to the rate as the aforementioned reciprocal relationship, then the expectation or mean for, say, element (a) at stage 1 is simply

$E(t_{a1}) = 1/u_{a1}$. When processing is exhaustive and we know that (a) is processed first, then the total mean processing time is just $E(t) = (1/u_{a1}) + 1/u_{b2}$.

However, turning to parallel models, we have connected the rates of intercompletion times, at any given stage, to the sum of the individual rates. The intuition for this, again, is that with many elements undergoing processing simultaneously, the chances are increased that one of the elements will be finished in a relatively early time, simply by chance. Although parallel models obviously do not permit simple additivity of processing times for the individual elements, we may still retain an additivity of intercompletion times, the overall sum of intercompletion times just being the overall processing time. Thus, we dissect the total processing time, be it self-terminating or exhaustive, into the separate intercompletion times and add them up. Now we have seen that regardless of which element, (a) or (b), finishes first in the present parallel systems, the rate of this event is just $v_{a1} + v_{b1}$; so according to our reciprocal rule relating mean processing times and rates, $E(t_{a1}) = E(t_{b1}) = (1/v_{a1} + v_{b1})$.[6] However, if (a) finishes first then (b) goes ahead and completes processing at the rate v_{b2}, and if (b) finishes first then (a) continues until its completion with rate v_{a2}. Hence, the average total exhaustive completion time *given* that (a) is completed first is $E(t|(a)\text{first}) = (1/v_{a1}+v_{b1}) + (1/v_{b2})$ and the other is $E(t|(b)\text{first}) = (1/v_{a1} + v_{b1}) + (1/v_{a2})$.

Relating these developments to the Sternberg paradigm when $n = 2$, and designating "processing" as being the comparison of the target letter with each of the memory letters, we observe that an "input" to the processor is a pair of letters, the target and one of the possible memory letters.

On a positive trial, that is, the target is also in memory, if processing is *self-terminating*, the serial predictions for the two possible cases of target placement are

$$E(t\,|\,\text{target is } (a)) = p\frac{1}{u_{a1}} + (1-p)\left(\frac{1}{u_{b1}} + \frac{1}{u_{a2}}\right) \tag{1}$$

and

$$E(t\,|\,\text{target is } (b)) = p\left(\frac{1}{u_{a1}} + \frac{1}{u_{b2}}\right) + (1-p)\frac{1}{u_{b1}}. \tag{2}$$

The prediction for negative trials and for all trials if processing is exhaustive is

$$E(t\,|\,\text{target absent or processing is exhaustive}) = p\left(\frac{1}{u_{a1}} + \frac{1}{u_{b2}}\right) + (1-p)\left(\frac{1}{u_{b1}} + \frac{1}{u_{a2}}\right). \tag{3}$$

[6]This property implies that the average first-stage processing time is independent of which element finishes first; it is peculiar to a class of processes of which the exponential is a special case. Investigation of a functional equation that is necessary and sufficient for parallel-serial equivalence with independent parallel models (Townsend, 1974), shows that in more general cases, the class of serial models may still be more general than the class of parallel models. The more general limitation on parallel models is closely related to this special case limitation.

We see that when the target is (a) that self-termination is reflected in the absence of the (b) component when (a) happens to be processed first, with probability p, and vice versa when the target is (b).

To arrive at the parallel predictions we could work directly from our remarks about additivity of intercompletion times, but we may alternatively use our previously established serial-to-parallel mappings to immediately acquire them. That is, let $p = (v_{a1}/v_{a1} + v_{b1})$, $u_{a1} = u_{b1} = u_1 = v_{a1} + v_{b1}$, $u_{a2} = v_{a2}$, $u_{b2} = v_{b2}$; which converts Eqs. (1), (2), and (3) to

$$E(t\,|\,\text{target is } (a)) = \frac{v_{a1}}{v_{a1}+v_{b1}} \times \frac{1}{v_{a1}+v_{b1}} + \frac{v_{b1}}{v_{a1}+v_{b1}} \times$$

$$\left[\frac{1}{v_{a1}+v_{b1}} + \frac{1}{v_{a2}}\right] = \frac{1}{v_{a1}+v_{b1}} + \frac{v_{b1}}{v_{a1}+v_{b1}} \times \frac{1}{v_{a2}}, \tag{4}$$

$$E(t\,|\,\text{target is } (b)) = \frac{1}{v_{a1}+v_{b1}} + \frac{v_{a1}}{v_{a1}+v_{b1}} \times \frac{1}{v_{b2}}, \tag{5}$$

$$E(t\,|\,\text{target absent or processing is exhaustive}) = \frac{1}{v_{a1}+v_{b1}} + \frac{v_{a1}}{v_{a1}+v_{b1}} \times \frac{1}{v_{b2}} + \frac{v_{b1}}{v_{a1}+v_{b1}} \times \frac{1}{v_{a2}}. \tag{6}$$

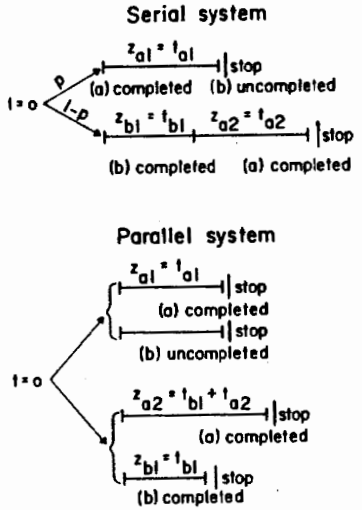

FIG. 2. Self-terminating serial and parallel systems when completion of element (a) terminates processing.

So far, so good. Figure 2 shows in schematic form the operations of self-terminating parallel and serial systems in the situation when (a) is the target.

The Standard Serial Model

The ony problem with application of the foregoing formulae to real data is that there are five parameters in the serial model (p, u_{a1}, u_{b1}, u_{a2}, u_{b2}) and four in the parallel (v_{a1}, v_{b1}, v_{a2}, v_{b2}). There will in general be $(n + 1)! - 1$ parameters for the serial model and

$$n! \sum_{k=0}^{n-1} \frac{1}{k!}$$

for the parallel, for each value of n. Now, a typical experiment and data reduction may result in only $n + 1$ degrees of freedom, at each value of n, corresponding to the n serial position mean RT's when the target is present and 1 for the target-absent mean RT. And this is without considering RT components from other parts of the information processing chain. Hence, it is of substantial interest to investigate models with further constraints on their parameters. A favorite serial model results from assuming that $u_{a1} = u_{b1} = u_{a2} = u_{b2} = u$, that is, the serial rate parameter is constant across the various elements and also through the different stages of processing. Further, this homogeneity of processing rate is extended across different numbers of input elements to be processed, so that if u_1 stands for the processing rate for $n = 1$, u_2 for the processing rate when $n = 2$, etc., then the latter assumption is expressed as $u_1 = u_2 = u_3 = \ldots = u_{n^*}$ where n^* is the largest number of inputs conceived of for an experiment. This last assumption basically involves capacity notions as we shall see later.

A known fact of statistics tells us that the mean of a sum of randomly distributed times is equal to the sum of the means of the individual times. Thus, any set of distributions yielding means that are equal on the different elements, independent of order and magnitude of n, will be referred to as a standard serial model. Specific assumptions about the underlying probability distributions will, of course, yield particular members of the class of standard serial models.

The preceding assumptions yield the classical straight line RT functions as n is varied. Inserting these constrained values of the u_i in expressions (1), (2), and (3) we obtain,

$$E(t \,|\, \text{target is } (a)) = \frac{1}{u}\ (2-p), \tag{7}$$

$$E(t \,|\, \text{target is } (b)) = \frac{1}{u}\ (1+p), \tag{8}$$

$$E(t \,|\, \text{target absent or processing is exhaustive}) = \frac{2}{u}. \tag{9}$$

The resulting processing time on a negative (target absent) trial then $n = 2$ is obviously just twice what it is when $n = 1$. When processing is self-terminating, it can be seen from Eqs. (7) and (8) that the processing time on a positive (target present) trial will depend on which element is the target.

It is typical to presume that the comparison times for the different items or elements (e.g., different letters of the alphabet) used in a Sternberg task are about the same; certainly the different items, usually letters or numerals, are randomly assigned positions in the displays. We therefore use (a) and (b) to designate different serial positions. Thus, (a) is taken as serial position 1 and (b) as serial position 2. If the letters are displayed visually, serial position 1 refers to the leftmost element, and so on until serial position n is the last toward the right. If they are presented acoustically, serial position 1 refers to that element presented first, serial position 2 to that presented second, etc.

Hence, it happens that processing time will differ according to serial position on positive trials when $n = 2$ *if* self-termination is possible *and* p is not equal to 1/2. [This can be concluded by inspection of Eqs. (7) and (8).] The latter means, of course, that the subject does *not* process (a) first just as often as he processes (b) first. The next thing to note is that the processing time averaged over serial positions, again assuming random placement of the target, is given by

$$E(t \,|\, \text{target present}) = \frac{3}{2u}. \tag{10}$$

The general formulae that give the straight lines as functions of n, again with the constrained u's, are

$$E(t \,|\, \text{target absent}) = \frac{n}{u}, \tag{11}$$

$$E(t \,|\, \text{target present}) = \frac{n}{2u} + \frac{1}{2u}, \tag{12}$$

where Eq. (11) clearly is the result of summing n average processing times, and Eq. (12) may be thought of as proceeding from the fact that on the average positive trial the self-terminating processor will go half-way through the n elements in order to find the target. Expressions (11) and (12) yield the famous, but rarely encountered, 2-to-1 slope ratio of negative to positive processing times, $(n/u) \div (n/2u) = 2$. We omit the expressions that correspond, in the case of arbitrary n, to Eqs. (7) and (8), since these can depend in quite complex ways on the probability distribution on processing orders of the n elements. The results up to here lead us to remark on a dilemma that has faced investigators of the Sternberg and similar paradigms: It is possible to obtain serial position effects, which suggest self-termination, and to simultaneously obtain equal-sloped positive and negative RT functions, which suggests exhaustive processing (with, for example, slopes of n/u). We shall pursue this dilemma in our section on applications to real data. We shall delay consideration of a typical self-terminating parallel model until the next section, since it ties in well with the independence-dependence issue. However, we close the present discussion by noting a parallel model that gives equivalent predictions to the present equal-u serial model. Let $v_{a1} = pu$, $v_{b1} = (1 - p)u$, $v_{a2} = v_{b2} = u$. That it does give equivalent predictions may be confirmed by plugging these values of the parallel parameters into expressions (4),

(5), and (6). More general serial-to-parallel mappings for larger values of n have been worked out elsewhere (Townsend, 1972a, 1972b).

THE INDEPENDENCE VS DEPENDENCE ISSUE

Independence of Total Completion Times

In some respects this issue is difficult to discuss. One reason is that there are many kinds of independence. The second is that the type we wish to approach has as yet been studied empirically only a little. This, as mentioned in the Introduction, is independence of total completion times of the various elements. We take especial note here that in all subsequent sections, when we speak of independent parallel processing or models, we are referring to independence of total completion times. By total completion time, we refer to the total amount of time the whole system is in operation before the element whose total completion time we are recording is finished. Putting it another way, the total completion time in a serial system is the processing time of the element whose time we are recording *plus* the sum of the processing times of the other elements whose processing precedes this designated element. In a parallel system, on the other hand, the total completion time *is* just the processing time of the element itself. Viewing Fig. 1a again, we see that our total completion time for (b) for the particular case shown, is $z_{b2} = t_{a1} + t_{b2}$ in the case of the parallel system, and $z_{a1} + z_{b2} = t_{a1} + t_{b2}$ in the case of the serial system. Neither the actual processing duration notation, given by the z's, nor the intercompletion time notation, given by the t's, is that needed for total completion time, so let us use x to refer to this new variable. Let x_a and x_b refer to the total completion times of elements (a) and (b) respectively. Then, we wish to learn something about how and/or why x_a affects or is interdependent with x_b.

Independent parallel models. Simply on the face of it, we might suspect that since these variables are just the processing times for the parallel models, independence would be easier or more natural to come by. Indeed, this is the case, as we may by fiat declare it so by selecting processing time probability distributions for x_a and x_b, so that they are independent; that is so that $f(x_a, x_b) = f(x_a) \times f(x_b)$ where $f(x_a, x_b)$ is the joint probability density function for x_a and x_b, and $f(x_a)$ and $f(x_b)$ are the respective marginal density functions.[7] In particular, the mathematical formulation for those simple parallel models possessing exponential intercompletion times, which corresponds to the intuitive parallel descriptions used throughout the other sections, can be easily made to do this. Let $v_{a1} = v_{a2}$, $v_{b1} = v_{b2}$, and this results in an independent parallel model. The rationale is that since the rate of processing (a) and of (b) proceeds unchanged across stages, there is no effect in completion time of one can have on the other.

Aside from the independence itself, we can observe that the average total times for

[7] Roughly, $f(x_a, x_b)$ as the joint probability density function gives the likelihood that the values x_a, x_b occur. The marginal probability density functions $f(x_a)$, $f(x_b)$, give the likelihoods of x_a, x_b, respectively, when one has averaged over x_b, x_a, respectively. For a more precise treatment see Parzen (1960).

(a) and (b), *without* the independence assumptions, are just Eqs. (4) and (5), $E(x_a) =$ (4) and $E(x_b) =$ (5), in other words, the expressions for self-terminating means when the target is (a) or (b), respectively. However, under our independence assumptions, Eqs. (4) and (5) reduce to

$$E(x_a) = \frac{1}{v_a}, \tag{4'}$$

$$E(x_b) = \frac{1}{v_b}. \tag{5'}$$

This is the parallel independent model that will be applied to RT data from a search task in the section on applications. Its model yields accuracy predictions of the sort associated with the parallel, independent sampling model suggested by Wolford et al. (1968). However, there were no processing time assumptions attached to their model apart from that of independence. Since their main interest was accuracy results with repetitions of the target element, specific processing time distributions weren't necessary. Any model that predicts $f(x_a, x_b) = f(x_a) \times f(x_b)$, and the obvious generalization to larger values of n, will predict independence of sampling for a fixed allotted time allowed the processor. That is, if total completion times are independent then, given a fixed time t that processing is going on, the probability that (a) is completed during that time is independent of whether or not (b) is finished during the same time interval. In symbols,

$$P\left[(b) \text{ is completed during time } t \,|\, (a) \text{ is completed during time } t\right] = P\left[(b) \text{ is completed during time } t\right]$$

or equivalently,

$$P\left[(a) \text{ and } (b) \text{ are both completed during time } t\right] = P\left[(a) \text{ is completed during time } t\right] \times P\left[(b) \text{ is completed during time } t\right].$$

Turning from the question of *independence* for a moment, we inquire if there are manipulations in the rates can cause *dependence* in parallel models. The answer is yes. Thus, a mathematical specification of parallel models gives a negative correlation when $v_{a2} < v_{a1}$ and $v_{b2} < v_{b1}$. The reasoning is that if the second stage rates are slower than the first stage rates, knowing that, say, (b) was completed during that time increases the likelihood that part of (a)'s processing time was governed by the slower rate. Thus, the chances are increased that a longer processing time was required for (a) and, hence, reduces the probability that it was completed during the fixed time t. We write this

$$P\left[(b) \text{ is completed during time } t \,|\, (a) \text{ is completed during time } t\right] < P\left[(b) \text{ is completed during time } t\right]$$

$$P\left[(a) \text{ is completed during time } t \,|\, (b) \text{ is completed during time } t\right] < P\left[(a) \text{ is completed during time } t\right].$$

Taking the opposite tack and setting the parallel rates so that $v_{a2} > v_{a1}$ and $v_{b2} > v_{b1}$ ensures that if one of the elements is completed sometime during a fixed time t, it is more likely that the other was also completed during that interval. That is,

$$P\left[(b) \text{ is completed during time } t \,|\, (a) \text{ is completed during time } t\right] > P\left[(b) \text{ is completed during time } t\right]$$

$$P\left[(a) \text{ is completed during time } t \,|\, (b) \text{ is completed during time } t\right] > P\left[(a) \text{ is completed during time } t\right].$$

It may be apparent that by trying intermediate alternatives, such as $v_{a2} > v_{a1}$, but $v_{b2} < v_{b1}$, that intermediate cases may be attained. Here, for example, the occurrence of (a)'s completion would lower the chance of (b)'s, but the completion of (b) would enhance the likelihood that (a) was also finished.

Independent serial models. Where do serial models stand on the question of independence of total completion times? Again, hammering home the point of the logical distinctiveness of the four issues, we find that serial models may yield independence or negative or positive dependencies, these three possibilities appearing as the relationships among the rate parameters are changed. In fact, it should now occur to us that we have only to employ our previous parallel-to-serial mappings to create serial models that evidence any of the three possibilities. This type of interplay crops up again and again in exploring processing structure of parallel and serial models: It is quite frequent that one investigates either the serial or parallel model, depending on which is most "natural" in some sense, simplest to work with, or both, and then employs the transformations of the variables (parameter mappings) that give equivalent results in terms of the other type of model.

Thus, to achieve an independent serial model, set $p = (v_a | v_a + v_b)$, $u_{a1} = u_{b1} = u_1 = v_a + v_b$, $u_{a2} = v_a$, $u_{b2} = v_b$. Note that the independence constraints are already contained in the parallel parameters and these are now incorporated into the serial structure via the mappings. To illustrate this, we can write the average self-terminating processing times in the serial version and then use the mappings to show the equivalence. Of course, equivalence of the average processing times says nothing about independence in general, but it will give us a hint about what is going on in averaged performance without deeper involvement in the mathematics. So, we rewrite Eq. (1):

$$\begin{aligned} E(t \,|\, \text{target is } (a)) &= p\,\frac{1}{u_{a1}} + (1-p)\left(\frac{1}{u_{b1}} + \frac{1}{u_{a2}}\right) \\ &= \frac{v_a}{v_a+v_b}\,\frac{1}{v_a+v_b} + \frac{v_b}{v_a+v_b}\left(\frac{1}{v_a+v_b} + \frac{1}{v_a}\right) \\ &= \left(\frac{v_a}{v_a+v_b} + \frac{v_b}{v_a+v_b}\right)\frac{1}{v_a+v_b} + \frac{v_b}{v_a(v_a+v_b)} \\ &= \frac{1}{v_a+v_b}\left(1 + \frac{v_b}{v_a}\right) \\ &= \frac{1}{v_a}. \end{aligned}$$

Similar operations show Eq. (2) to be:

$$\begin{aligned} E(t \,|\, \text{target is } (b)) &= p\left(\frac{1}{u_{a1}} + \frac{1}{u_{b2}}\right) + (1-p)\,\frac{1}{u_{b1}} \\ &= \frac{1}{v_b}. \end{aligned}$$

It is not at all difficult to arrive at serial models that possess positive or negative correlations in total processing times by way of our parallel-serial equivalences. Given our limited space, however, we choose to take a brief look at a standard serial model with specific assumptions about the probability distribution on the intercompletion times. This is the model with a processing rate that is constant over stages, elements, and different number of input elements. It yields the average processing time predictions of the standard serial model presented in the previous section. When these assumptions are embedded in a model with exponential intercompletion times and when $p = 1$ or $p = 0$, the model becomes a Poisson process or alternatively and equivalently, the total completion times are Gamma distributed. For any value of p, it can be shown that there is a positive correlation between, say, x_a and x_b when $n = 2$. Thus, when it is known that (a) has completed processing during time t, the likelihood is increased (over what it would be if there were no knowledge of (a)) that (b) has also been finished during that same time period. This particular serial model corresponds to one of the parallel models, as far as the independence problem goes, that assumes an *increase* in processing rates across stages, with the relative values of the v's at each stage mimicking the serial probabilities of the various processing orders.

Figure 3 shows how parallel and serial rates act in cases of positive dependence and Fig. 4 shows independence when (a) is finished first. Note that to achieve independence, it is not enough simply for the serial rates at stage 2, u_{a2}, u_{b2}, to be less than the stage 1 rate, u_1; they must be less than pu_1 and $(1-p)u_1$, respectively.

Other Types of Independence

Before leaving the topic of independence, it may be helpful to our perspective to review some of the types of independence other than that of total completion times.

Independence of number of elements. For the first, it is common to use the term independence to refer to a functional independence of the number of elements to

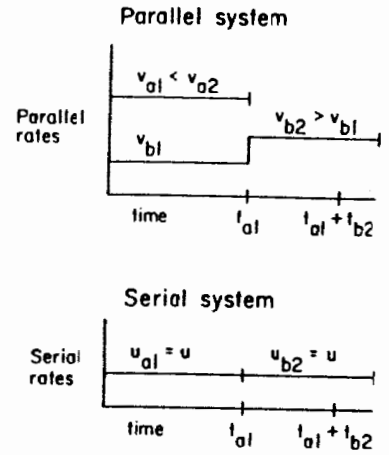

FIG. 3. Positively dependent systems shown when (*a*) is completed first.

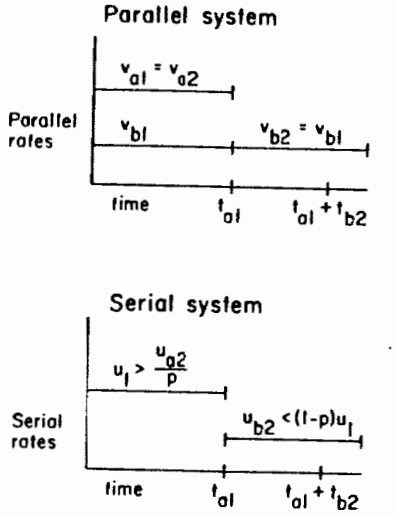

FIG. 4. Independent systems shown when (*a*) is completed first.

be processed, which has no necessary logical connections with statistical independence. If it is found that RT curves as n is varied are flat, it is said that processing is independent of n. This is okay, but it must be kept in mind that the question of how processing time and therefore RT varies as n varies is really one of capacity, as noted in the Introduction.

Donderian independence. The term "stage" is often used to denote successive

subsystems in the human information-processing chain. Used in this more macroscopic sense, if two successive stages are additive in that the processing time of the second begins immediately when the first is completed and not before, and if they are non-interactive in the sense that variables that affect the magnitude of the processing time of one do not affect that of the other, then here also a separate type of "independence" occurs. An hypothesis that this kind of independence characterizes successive stages is termed an "assumption of selective influence" by Stemberg (1969). Two successive stages will be *dependent* in this sense if something that affects the overall average magnitude of one affects the overall average magnitude of the other (see also Pachella's discussion of this type of independence, this volume). This type of independence is statistical only in the sense that statistical analyses, as linear regression or analysis-of-variance, may be used to test it.

Independence of successive intercompletion times with element identity. Yet another kind of independence, or dependence, is statistical in a fundamental sense. Some serial models, associated with this kind of independence, which give means that are additive in the mean processing times of the individual elements (i.e., giving those serial formulae presented earlier), posit independence of successive intercompletion times. That is, the fact that (*a*) finishes first in exactly t msec has absolutely no bearing on how long it now takes (*b*) to be processed, and vice versa if (*b*) is completed first. Although this sense of independence at first may appear to be intimately related to that of total completion times, models possessing independence of successive intercompletion times can predict either independence of the type (i.e., of total completion times) we considered in detail earlier. Parallel models can also possess independence of successive intercompletion times with element identity. In fact, the models on which we base our discussions in this chapter assume this type of independence.

Independence of successive intercompletion times without element identity. A slightly grosser level at which the question of independence may be asked is with respect to whether the *overall* processing time of the second element is independent of the processing time of the first. This is not quite the same thing as that in the preceding paragraph. There, not only the *duration* of processing of the first is given, but also the *identity* of the element processed first. In the level referred to here, the identity is not known, only the duration of the first element's processing is known, with an averaging over the identity of the elements. It can be shown that the presence of independence of successive intercompletion times with element identification does not imply independence of successive intercompletion times when such knowledge is absent; nor does it imply dependence.

One other remark is pertinent here. The general idea of independence of intercompletion times, or successive stages in the microscopic sense used in this paper, may be applied to macroscopic successive stages. That is, it may be asked whether, for example, stimulus encoding time is statistically independent of short-term memory storage time. This question is itself "independent" or orthogonal to the question of the Donderian type of independence. Thus, an experimental manipulation that affects the overall rate of processing in the stimulus encoding stage may or may not affect the

overall rate of processing in the storage stage. Either may occur without in any way relating to whether or not the magnitude of time taken by the encoding stage on a given trial affects that taken by the storage time on the same trial. For instance, the latter may show a correlation via a trial-to-trial interdependence of the two stages' deviations about their relative means. On the other hand, we could discuss models in which these two types of dependence are tied together, but since they are not *necessarily* connected they should always be treated in a logically separate manner.

In concluding this section, it is germane, as usual, to emphasize the limitations on interpretations of experimental results. A finding of independence of total completion times, for instance, although perhaps more intuitively associated with parallel models, can be predicted by serial models. Hence, under conditions where there might be some slowing down of processing by a serial mechanism in succeeding stages, it may be reasonable to postulate a serial, independent system. Conversely, there may exist processing tasks where processing is parallel but a warm-up effect occurs, inducing a speed-up of processing rates; such a system can yield positive correlations of the sort associated with serial systems.

THE LIMITED VS. UNLIMITED CAPACITY ISSUE

Basic Concepts

The capacity issue refers naturally enough to the distinction between channels or processors that take more time or make more errors as the number of inputs increases —as opposed to those that do not. This theoretical dichotomy is frequently referred to as limited- vs. unlimited-capacity processing systems.

To be sure, no real system, biological, electronic, or mechanical, ever really has unlimited capacity. All physical or biological devices or systems have an upper limit on their capacity. In fact, the term "attention" would be unnecessary in psychology if the mind *really* had unlimited capacity. However, it is quite helpful to entertain the possibility that within certain limits a system might perform in a constant manner (to be made more precise below) regardless of the number of inputs. For example, the possibility that some central agent responsible for monitoring incoming signals from our various modalities had unlimited capacity relative to, say, hearing, seeing, and touch was suggested to be false early in the history of experimental psychology, and the prior-entry hypothesis (Boring, 1950) (essentially a serial explanation) was used to explain how the central agent reacted to more or less simultaneous inputs over separate modalities. A current matter of interest is whether for certain types of symbol-detection tasks the visual system is capable of acting as an unlimited capacity processor (Estes & Taylor, 1964; Erikson & Spencer, 1969; E. E. Smith, this volume).

Any complete theory of processing has to include structure explaining how errors may arise. Within the present theoretical treatment, which emphasizes the possible stochastic nature of processing of individual inputs, a natural way of introducing error is through time limitations. For example, speed-accuracy tradeoff predictions can be obtained by assuming the subject establishes a criterion duration at the end of which

he makes a response based on the information he has been gathering up to that time. Clearly, error rate will be inversely proportional to the value of the criterion. Similarly, experimental manipulations of the duration that the stimulus inputs are available for processing, as with the use of brief displays and backward masking, would be expected to affect error probability.

Another possibility in explaining speed-accuracy trade-offs, at least at a rudimentary level, would be to assume the subject sets a "state" criterion rather than a time criterion; that is, he might decide to quit processing after some given number of elements have been processed. More subtle versions of time-dependent decisions abound in the modern sensory and information-processing literature. A recent example of interest to which the reader is referred is the comparison of predictions by an average "waiting-time" criterion with those from a "count" criterion in psychoacoustic experiments (Green & Luce, 1972).

We will confine our attention here primarily to how variation in number of inputs affects overall processing times, assuming, as we have throughout the chapter, essentially unlimited processing time is available, relative to the processing rates. Predictions of probability correct, etc., involving truncated processing times, can be retrieved by employing the same probabilistic descriptions of the models from which processing time predictions are derived. Any assumption of error-free processing is, to be sure, only an approximation under the best of conditions. Under certain circumstances, it may be that errors reflect aberrations in mechanisms other than the processes under scrutiny (e.g., the fast-guess model, Yellott, 1971), but this would have to be empirically demonstrated. We cannot go further into these issues here, but it is becoming increasingly apparent that a thorough treatment of the error problem is overdue in memory and perceptual-scanning studies. An approach that promises to be fruitful within certain experimental contests is offered by Theios, Smith, Haviland, Traupman, and Moy (1973). Their serial model employs some specific assumptions about order of processing in short-term memory search and derives predictions about errors, reaction time, and sequential effects.[8]

Capacity at Different Levels

As noted above, the terms "unlimited capacity" and "limited capacity" are not very accurate ones. However, rather than resort to neologisms, we shall try to make clear how we believe the usual terms should be employed. Apart from the misleading semantic connotations, however, there is a more substantive problem, namely, the capacity issue may be raised with respect to any number of different levels of

[8]The same questions could be posed for time and noise as are posed here for n. That is, systems that have unlimited capacity with respect to time limitations or with respect to variations in the amount of noise present, at least up to a point, are in principle conceivable. For example, one could devise a system which up to certain limits could provide a variable amount of filtering action in order to maintain a constant processing time and/or constant accuracy rate. The present theoretical apparatus will have to be imbued at some point with structure for direct representation of noise effects before latency predictions concerning them can be dealt with. There is not much evidence for human information-processing systems that have unlimited capacity with respect to time, in other than the trivial sense of perhaps having much more time than they need under some conditions.

processing. Thus, unlimited capacity at the level of individual elements means that when an element is actually undergoing processing, the amount of time it requires to be processed is unchanged as the total number of input elements, n, is varied. The processing durations referred to are, of course, simply the z's in Fig. 1a. Limited capacity at this level would mean that some measure of the individual processing times, most likely the mean, would increase as n increased. At the opposite pole in the hierarchy of levels is exhaustive processing. Unlimited capacity at this level has to imply that the overall total time to complete all the elements, again probably a measure of central tendency like the mean, remains constant as n is increased. Limited capacity points to an increase in this measure. We shall use mean processing time as the desired measure.[9]

Two other related aspects must be mentioned in preparing to make specific statements. One is that fixing a type of capacity at one level may involve inexorable consequences in capacity at the other levels. Thus, we shall show that an assumption of unlimited capacity at one level may simultaneously force a limitation in capacity at higher levels, and something "better" than unlimited capacity at lower levels. "Higher" here refers to more elements, so the exhaustive level is higher than the individual element level. "Better" in the above sentence brings up the second aspect and suggests the concept of "super capacity," in that processing may actually speed up on the average at a given level as n increases; we shall use the term "super capacity" to that effect in the following.

Unfortunately, the situation is even more complicated than we have so far admitted. The additional complexity arises because the rates, which specify speed of processing of individual elements, may be related to the individual mean processing time in an intricate manner. This happens, for example, when the rates are free to change across stages during a single trial. The outcome for modeling is that a parallel and serial model may make identical predictions everywhere, including capacity effects, but the ways the predictions are evidenced via the rates of processing will typically be completely different for the two. We shall have occasion to remark on this circumstance again below.

The Standard Serial Model

We begin with our typical or standard serial model, the one with equal u's over everything. The first point to be made here is that this model has unlimited capacity at the level of individual elements, since the rates and individual mean processing times are unchanged as n varies. As was emphasized earlier, this assumption that the serial rate is constant over differing values of n is in no way a necessary concomitant of the concept of seriality: one-at-a-time processing. In the present serial model, processing is of limited capacity both at the self-terminating level and at the exhaustive level of processing, since mean processing time increases (linearly as it happens) with n for

[9]The reader may wish to try his hand at deriving predictions for a level we shall have to neglect here: that of minimum processing time. Egeth (1966) discusses some of these.

TABLE 1

Capacity as a Function of Model and Level

	Level		
	Individual element	Self-terminating	Exhaustive
Standard serial model with equal constant u's	UC	LC	LC
Standard parallel UC deterministic models	UC	UC	UC
Probabilistic nonindependent model that is UC at exhaustive level	SC	SC	UC
Standard parallel probabilistic independent model with equal constant rates	UC	UC	LC

Note.—UC = unlimited capacity; LC = limited capacity; SC = super capacity.

both of these levels. These features are represented in the first row of Table 1. Another feature of this serial model that is interesting to contrast with parallel models is that the limitations in capacity for these higher levels do not depend on any particular type of probability distribution for processing times. The fact that the mean processing time is just the sum of the means of the individual elements implies this generality. Indeed, it should be clear that a deterministic serial model that assumes a processing time of exactly $1/u$ for every element gives descriptions of capacity identical to probabilistic models that assume that $1/u$ is the mean processing time for the individual elements.

Standard Parallel Models

It provides an interesting contrast to next discuss standard parallel models that also have unlimited capacity with respect to *individual* elements. Here, it *does* matter whether we are discussing deterministic or probabilistic models and systems. If processing is deterministic and processing rates for individual elements are unchanged as n increases, then the exhaustive and self-terminating processing times will also be unchanged and therefore of unlimited capacity. On the other hand, if processing is probabilistic, then whether exhaustive processing is of unlimited capacity will depend on whether correlations in total completion time exist (see the section on the independence-dependence issue) and whether the correlations are

positive or negative. The deterministic parallel capacity predictions are shown in row 2 of Table 1.[10]

To produce flat exhaustive curves in nondeterministic cases it must happen that processing somehow speeds up as n increases. A fairly reasonable way to produce this is to assume that as each element finishes, the remaining elements speed up proportionately. This assumption, along with the assumption that at the *start* of processing on any given trial, each element has the same rate as if there were only one element to be processed, is sufficient to produce a flat, mean exhaustive-processing time curve as a function of n. Thus, if v is the processing rate for a single element when only one is input, then that is also the processing rate for each of n elements at the first stage of processing when n are input.

At the end of stage 1, which is itself governed by total speed or rate nv (that is, the sum of the n individual rates), assume that the new rate for each of the $n - 1$ remaining elements is $(n/n - 1)v$, that it is $(n/n - 2)$ at the third stage, and so on. The time required, on the average, for stage 1 is naturally the reciprocal of the total rate, namely, $1/nv$. Since the rate during stage 2 is $(n/n - 1)v$ for each of the remaining $n - 1$ elements, the total speed or rate is $(n - 1) \times (n/n - 1)v = nv$, and by the same principle as before, the average time for stage 2 is just $1/nv$. Following this line of reasoning we see that the overall mean exhaustive processing time will then be the sum of the times over all n stages, $1/nv + 1/nv + \ldots + 1/nv = n \times 1/nv = 1/v$, the same duration as that taken when $n = 1$, thus yielding exhaustive unlimited capacity. However, this model does not give unlimited capacity at the self-terminating or individual levels. On the contrary, it predicts *super capacity* performance on both these. This is because the average completion time for a given element is found by averaging over all the possible stages in which the given element would be completed. If it is not finished until last, it takes time $1/v$. Any other stage of completion results in less time, so the overall average is itself less than $1/v$, the time required for an element when $n = 1$. Thus individual and self-terminating processing time is super capacity, that is, decreases with n (this may be shown precisely by carrying out the averaging process and observing the resulting expressions' behavior as a function of n). Row 3 of Table 1 summarizes this model. Another characteristic of the model is that the total completion times are positively correlated.[11]

[10]A slight intricacy in parallel deterministic processing should be noted here. If the elements to be processed are heterogeneous with regard to how fast they can be processed, then increasing n will mean that, on the average, exhaustive times will increase, even though the rates of the individual elements do not increase as n increases. The reason is that since the longest time determines the exhaustive time, the more elements there are to be processed, the greater the likelihood that a very slow element will be included on any given trial. However, if self-terminating processing requires that only a single critical element be processed, then heterogeneity will not affect this processing time, since the probability of a slow or fast element being the critical element will be unchanged as n increases (under suitable stimulus randomization).

[11]Yet another possibility is that while the elements' total completion times are parallel and independent, the individual times consist of a series of steps that are negatively correlated with one another; when one step takes a long time, the succeeding step takes only a little time, and so on. The negative correlations in these random times could make the overall result appear deterministic.

There is a fair amount known about the case of independence of parallel total completion times and the influence of varying n on exhaustive processing times, especially when the processing time distributions are the same on all the elements. The reason is that the latter problem is really just that of computing the maximum value of a random sample of n identically distributed random variables. The maximum value corresponds, naturally enough, to the longest processing time of the n elements. The exact distribution of the maximum is usually not trivial to find. However, the case of n independently sampled, identically distributed values (e.g., equal constant v's), is perhaps the most classic of modern statistics, so it is not too surprising that something has been achieved on the mean or expected maximum as n varies (see e.g., Gumbel, 1958, p. 91). An important finding for most reasonable probability distributions on processing time for a single element is that the *largest possible* mean or expected exhaustive processing time is $E(t \mid \text{exhaustive processing of } n \text{ elements}) = E(t \mid \text{one element}) + \frac{\sigma(n-1)}{\sqrt{2n-1}}$, where σ = standard deviation for $n = 1$. It follows easily from this expression that this largest possible mean exhaustive time grows at a rate slower than $\sqrt{n \mid 2}$. The curve of such a function is increasing but at an ever slower rate. This does not imply that all exhaustive-mean predicted curves of unlimited-capacity (individual level) independent, parallel models must be positive, and negatively accelerated as the above is. However it, along with the observation that several typical distributions do just that, suggests this may not be uncommon behavior. Given an initial mean and standard deviation, it does give bounds on how much increase is possible for any larger value of n. This is apparently the rationale behind Sternberg's (1966) parallel-bound curve.

In any case, this type of system or model is typically limited capacity at the exhaustive level by virtue of the increment in processing times associated with increases in n. It is also the set of models falsified by Sternberg's (1966) data. Nevertheless, the implication in that paper that the fairly steep straight-line RT curves obtained by varying the number of elements in memory cannot be predicted by independent parallel models is wrong, as we shall show later in the present section. On the other hand, it is true, by virtue of the negative acceleration pointed out in the preceding paragraph, that typical independent parallel models that are *unlimited capacity* at the level of *individual elements* do not predict *linear* increasing RT functions.

To gain some insight into these negatively accelerated exhaustive-processing functions, assume that each of a set of memory elements is itself made up of more elementary features. Assume further that the memory elements are compared in parallel and independently with a target element but that within each memory element the features of which it is composed are serially matched against those of the target element. Now, since we are considering exhaustive processing, we ask about the finishing time of that last element to be completed. But, on the average, when all the features of the first element to be finished are done, some of the features of the others will also have been completed. Finally, when $n - 1$ of the n memory elements are

done, most of the features of the last one will already be finished so only a little more time will be required to complete it. Hence, it can be seen that for each added memory element, a smaller increment to the exhaustive processing time will be added, this increment corresponding to the smaller number of unfinished features left after the next to last element is completed.

A special case of this class of models is obtained when each memory element is compared holistically with the target letter, or equivalently, only a single feature is required for comparison. In this case, we let v represent the processing rate for each element, and since this model is unlimited capacity at the level of individual elements, the rate v is constant across values of n. This model, assuming as it does exponential intercompletion times, yields the expected negatively accelerated exhaustive processing curves.

The idea here is that when processing begins at stage 1 for arbitrary n, each element rate is v (for any value of n) and by our former remarks on intercompletion times for such parallel models, the first mean intercompletion time will be $1/nv$. At the second stage, with one element completed, the total sum of parallel rates is $(n-1)v$ so the next intercompletion time average is $1/[(n-1)v]$ and so on until the last intercompletion time is $1/v$, yielding the exhaustive mean processing time of $1/v + 1/2v + \ldots + 1/nv = \frac{1}{v}\sum_{i=1}^{n}\frac{1}{i}$. A note on this sum that is often useful is that its curvilinearity is well approximated by $\log(n)$ and thus conforms to our previous statement about parallel independent, individual unlimited capacity, with identically distributed times having increasing negatively accelerated curves.

What happens at the self-terminating level in these models? Since processing is independent on the individual elements' processing times, no element affects any other, and so the self-terminating time is invariant as n changes. Table 1, row 4 summarizes the capacity properties of these parallel models.

Mimicking Parallel and Serial Capacity Predictions

Parallel models that mimic the standard serial model. First, consider the standard serial model (Table 1, row 1). We already possess the type of mapping strategy that produces parallel models that are equivalent to serial models when $n = 2$. This same strategy may be expanded for larger n. Complete derivations are given elsewhere (e.g., Townsend, 1972a, 1974). Here we confine ourselves to seeing what the mappings look like when all processing orders are equally likely so that all the serial p values are equal. In that case, writing the equal v values as functions of n and stage, $v(i,n)$ we have $v(i,n) = [v(1)/(n-i+1)]$, ($v(1)$ = processing rate when $n = 1$). Since at stage i, there are $n-i+1$ elements remaining, all being processed at rate $v(i,n)$, it follows that the total rate at each stage is just $(n-i+1)\times v(i,n) = v(1)$, so that the average total time for exhaustive processing is $nv(1)$. The average time for self-terminating processing is, of course, $(n + 1/2)v(1)$, so letting $v(1) = u$ produces equivalent predictions to the standard serial model.

On several occasions in this section, a within-trial speed-up of *individual* parallel rates has been employed to some purpose; in the present case, to help mimic standard serial behavior. A possible rationale for such a speed-up is the reallocation of capacity that was devoted to a completed element to one or more of the uncompleted elements.

The preceding parallel model does not possess independence of total completion times of its elements, so it was not one of the parallel models claimed to be falsified by Sternberg (1966). This study involved the memorization of letter lists of varying (but small) size and a consequent presentation of a target letter which the subject reported as present or absent in his memorized list. However, it was then implied that the class of parallel independent models was falsified by the resulting relatively steep straight-line RT functions of n. Since the target-absent and target-present observed curves were coincident, therefore having equal slopes, it was proposed that the appropriate model was (what we are calling) the standard serial model with exhaustive processing even on target-present trials.

We now present an independent parallel exhaustive model that predicts straight line RT curves. We may not let the rates change with stage (e.g., no speed-up within trials is assumed) because this typically produces nonindependent processing. However, we can let v vary with n, writing $v(n)$. Specifically, for $n = 2$ let, $v(1) = 1/k$, for $n = 2$, $v(2) = (1 + \frac{1}{2})/2k$, and in general,

$$v(n) = \frac{\sum_{i=1}^{n}\frac{1}{i}}{kn}.$$

We can easily prove the conjecture that this model predicts straight-line curves with arbitrary slope k. Recalling that the average exhaustive processing time in the case of independent parallel processing is simply the sum of the average times for each successive stage, and noting that since the rates of each individual element remain constant across stages, the total speed decreases in proportion to the number of elements completed, produces:

$$E(t) = \frac{1}{nv} + \frac{1}{(n-1)v} + \frac{1}{(n-2)v} + \ldots + \frac{1}{v} = \frac{1}{v}\sum_{i=1}^{n}\frac{1}{i}$$

Substitute in $v = \dfrac{\sum_{i=1}^{n}\frac{1}{i}}{kn}$

and verify the conjecture that $E(t) = kn$. Observe that $v(n)$ decreases as n increases, so that this model has limited capacity at the individual element level. This particular independent model also predicts no serial position effects, as was apparently the case in Sternberg's data. Independent parallel models can predict serial position effects by assuming unequal rates at the different positions. In the next section, we show that an independent parallel model can handle straight-line RT data with serial position effects. Now, it is not claimed that this particular model is necessarily the "true" model for these data, only that the data did not actually rule out parallel independent (limited-capacity) models.

Serial models that minic parallel, deterministic, unlimited capacity models. A task that is interesting to consider in the present connection was used by Neisser (1963). In these experiments, a subject holds one or more items (e.g., letters) in

memory and looks through lists to detect matches with the memory list. When they have had a great deal of practice, it is not uncommon to find that subjects can perform scans of lists of letters for memory set sizes of 1 to 10 letters, with no appreciable increase in average scan rate per display-list letter. To the extent that this can occur without a corresponding increase in error rate, some form of unlimited capacity model at the level of exhaustive processing is supported. The idea is that although when the processor comes upon a target letter, the comparisons can be self-terminating, that the bulk of the time is taken up with negative comparisons with nontarget letters. This is because most of the list letters are nontarget letters. All the latter comparisons must be exhaustive with respect to the memory list. The deterministic parallel model with unlimited capacity at the individual level makes appropriate predictions for these data (Table 1, row 2). The probabilistic parallel models that assume super capacity at the individual element level cannot be ruled out either, since there was no opportunity in those experiments to test the prediction of these models that self-terminating times should decrease with n (Table 1, row 3).

It is very critical that we be clear on the prediction of serial models with respect to flat RT functions as n varies. Serial models *can* indeed predict such functions, therefore preserving the logical and mathematical independence of capacity and parallel-serial assumptions that we have argued for. This is not to say that the serial explanations are intuitive or natural. In fact, this is one of the few situations where there is a strong intuitive argument for one of the types of processing over the other. If the curve of RT as n increases went up at all, then it would be easy to find serial models *or* parallel models that could explain the data in an intuitive fashion. But to predict completely flat functions, the serial models have to say that processing is super capacity on individual elements, that is, a speed-up takes place as n increases. And this super capacity cannot happen in a reasonable way as it could in the case of parallel processing. The parallel processing rates could be the same (as when $n = 1$) at stage 1 and then increase across stages by (perhaps) picking up excess processing capacity left over from the completed elements. But in serial processing, since the total sum of n individual times must equal that for $n = 1$, there can be no such capacity sharing. The type of speed-up across stages that works for a parallel system, for example, will not work for a serial system. Consider a serial system that takes the preceding rate and adds it to the rate for the next element. This was mentioned as a rationale for parallel models that have unlimited capacity at the individual level. Although not perhaps as reasonable as for a parallel model, such a speed-up within trials might occur due to a within-trial priming or warm-up as processing gets under way. In any event, the total exhaustive processing time would be $1/u + 1/2u + 1/3u + \ldots + 1/nu$. This, of course, is not equal to $1/u$ as it would have to be. A (unintuitive) type of speed-up that *would* work is to simply let the (supercapacity) $u(n) = nu$; that is, the rate of processing grows proportionately to n.

Serial models that mimic standard probabilistic, independent parallel models that have unlimited capacity at the level of individual elements. What kind of serial model mimics the negatively accelerated exhaustive curves as n varies, as well as the unlimited capacity aspects of the standard probabilistic parallel models

(Table 1, row 4)? The serial models making these predictions assume that overall there is an increase in processing rate as n increases but that there is a decrease in rate as successive elements are completed within a trial. Specifically, let us mimic the standard parallel model given on p. 164 (the equal-v model).

The serial model that is equivalent to this model assumes that for any value of n, $u(1,n)=nv$, $u(2,n)=(n-1)v \ldots, u(n,n)=v$, *where* $u(i,n)$ is the serial rate for any element that is processed at stage i when a total of n elements are being input into the system. Also, the probability that any given element is processed i^{th} is just $1/n$. It is fairly obvious that the increasing rate with n mimics the addition of more elements in the parallel case and that the within-trial decrease in rate mimics the slow-down in overall parallel production (i.e., the more elements that are finished, the longer the intercompletion times) as more and more elements are completed. Perhaps the reader will wish to prove further to himself that the predicted self-terminating time is indeed $1/u(1,1) = 1/v$. The serial, parallel-mimicking system that conforms to this model can be viewed as gearing up its stage 1 processing rate more and more as the number of inputs increases, but "getting tired" within a trial, slowing down to its old rate $(u(n,n) = v = u(1))$ for the last (n^{th}) element to be processed. Further, the equal parallel rates force the equal p's on the serial model (i.e., equally likely processing orders), so that both would predict a lack of serial position effects.

Serial position and other order effects have been neglected in this section because some studies either don't find them or preclude their appearance. The presentation would also have been substantially complicated had they been included. However, conclusions concerning predictions and rate effects will continue to hold on the average when order related phenomena are treated.

In the following section, we review the modeling and conclusions of a recent experiment and offer a general discussion of models for search experiments.

AN EXPERIMENTAL APPLICATION AND CONSIDERATION OF GENERAL MODELS OF SHORT-TERM MEMORY AND BRIEF VISUAL DISPLAY SEARCH

The Experiment

Mathematical characterization of cognitive processing and issues is all well and good but of little interest to the experimentalist unless it can be shown to aid him in the penetration of data. In this section, we take a small but representative portion of data obtained in some recent multisymbol search experiments (Townsend & Roos, 1973) and show how some of the concepts developed in the preceding sections can be applied to data. There have been few mathematically specified serial models actually fit to data, and with the exception of Rumelhart (1970) almost no mathematical parallel applications. The investigation involved behavior in a short-term memory experiment of the Sternberg (1966) variety, as well as in a brief visual display search-task previously employed by Atkinson et al. (1969). Part of the study was directed toward an attempt to formulate a qualitative model of an overall system,

subsystems of which were assumed to interact in producing performance in the two tasks. Several broad characteristics of the data were very similar in the two situations, with some mildly unusual facets appearing in the memory part of the task. Two quantitative models were fit to the data of both experiments. Some general mathematical models were examined that were supported or falsified by these and like data of other experimenters. Because the visual search data were more typical of those obtained by others than the memory data, we shall take a portion of the former for detailed consideration.

To get to particulars, the data we take up were acquired in a task in which the subject had to indicate by pressing one of two buttons whether or not a letter that had been given to him visually 3 seconds earlier was in the later multiletter visual display. The first display appeared for 2 sec, and the later display appeared for 400 msec. The second multiletter display was centered horizontally in the subject's fixation area, and its size (n) varied randomly from 1 to 5 letters. On ½ of the trials the second display did contain the "target" letter, and if present, the target occupied any of the n display (serial) positions with probability $1/n$, (n = 1, 2,...5). This type of paradigm, run under high accuracy conditions, as it was here, was apparently first employed by Atkinson et al. (1969), although it has generic precedents in the designs of Estes and Taylor (1964) and Sternberg (1966).

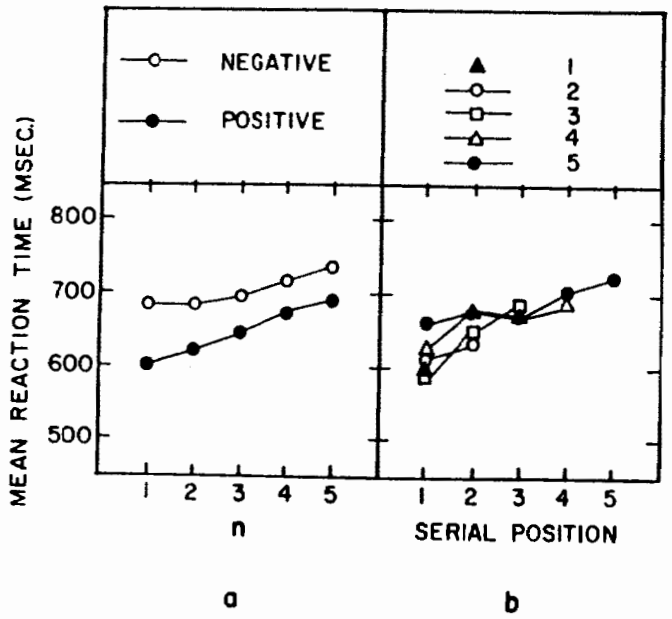

FIG. 5. 5a.—Target present (positive) and target absent (negative) mean reaction time as a function of number of letters in display. 5b.—Mean reaction time as a function of serial position (target location); the different curves are for n = 1,2,3,4,5.

Reaction times were recorded, and it was duly noted that errors were small and inconsistently related to n. We pause again to note that a full treatment of such experiments will ultimately have to include error prediction. But as suggested in earlier papers (e.g., Townsend, 1971b, 1972a) the problems of equivalence between models is portentous, even omitting the error dimension, and the other issues are sufficiently complex to command some sustained attention.

The three subjects employed produced results rather similar to one another including similar trends in serial position effects. It will be recalled that serial position effects are the RTs for trials when the target was present (+ trials) as functions of the position of the target. Serial position effects are obviously confined to + trials since target absent (−) trials provide no differential position RTs. Figure 5a shows the overall mean RTs as functions of n (our numerical illustration shall consider only n = 1, 2, and 3, although the figures show all 5 points obtained in the study), and Fig. 5b shows the serial position functions, where each separate curve is for a different n and both graphs represent group averages. The average standard error for a single point in Fig. 1a was between 5 to 10 msec, and there are 750 trials underlying each point.

A strong linear trend is obvious in the overall means, indicating limited capacity at exhaustive and possibly self-terminating levels.[12] The serial position curves indicate a tendency to respond faster, the more left the target appeared in the display.[13]

Serial and Parallel Models: Fits to Target-Present Data

We shall take a couple of "typical" models, one serial, the other parallel, and fit them to the RT^+ (reaction time) data and then see how well they predict the RT^- data. We shall then discuss the implications these results have for parallel and serial models in general.

The serial model. The serial model we are going to use is self-terminating and assumes that the processing rate is constant over serial positions and stages of processing but allows the rate possibly to change as a function of n [i.e., we use $u(n)$]. Hence, this model assumes equal u's for *fixed* n, but the capacity question is left

[12] A detailed discourse on conceivable explanations of the limited capacity nature of these and similar results cannot be taken up here. However, recent studies have shown that identification of letters may be affected by interletter similarity and by lateral masking effects in multiletter situations (Estes, 1972; Estes & Wolford, 1971; Townsend, 1971a; Townsend, Taylor, & Brown, 1971; see also the discussion by E. Smith, this volume).

It may be observed in this context, that the present definitions concerning capacity purposely do not distinguish between structural or bottleneck as opposed to energy-source types of limitations (see e.g., Kahneman, 1973). It does not seem entirely clear at present, whether a structural vs. energy nature of capacity dichotomy will turn out to be fruitful. When a parallel distribution of attention is reduced to a single element, seriality and the appearance of a "bottleneck" is obtained. Possibly, a criterion for distinguishing structural from other kinds of capacity-limiting mechanisms should be the relative lability of the distribution of capacity.

[13] In this study, the *S*s were instructed to attempt to scan from left to right in order to see if serial position effects could be elicited. As they were weak or nonexistent in the Atkinson et al. (1969) experiment, it was of interest to optimize the chances that they could occur. Since they did, manipulability of this aspect of processing behavior is suggested.

open. In order to obtain a workable theory of the distribution on processing paths, it is assumed that with each position in the processing order i, there is a probability [i.e., $(p_i(n)]$ that the letter in the corresponding serial position i will be processed then (i.e., i^{th}). If [with probability $1 - p_i(n)$]the i^{th} serial position is not processed i^{th}, any of the other $n - 1$ possible serial positions is processed i^{th} with probability $1/(n - 1)$, that is, the i^{th} processing position then has an equal probability of containing any other serial position's letter in the processing order. Note that this automatically assures that the total probability that the i^{th} processing position contains *some* serial position is equal to 1. However, it also the case that the total probability that each serial position gets processed *somewhere* in one of the n processing positions must be 1. Since this constraint is not forced by the assumptions given so far, it was assumed for purposes of parameter estimation that for any serial position k ($k = 1, 2, \ldots n$), the probability that k is in the n^{th} processing position is just 1 minus the sum of the probabilities that the k^{th} serial position is contained in one of the other processing positions (that is, in one that is less than n).

In addition to these "processing" parameters, there is a parameter, t_0, required to predict a nonzero intercept. The parameter t_0, which we call the residual latency parameter, presumably includes averaged sensory, motor, and other time contributions that are not directly related to comparing the display letters with the target. We make the usual assumption that t_0 is additive with the comparison processing times as n is varied. These assumptions about processing lead to the following formulae for serial-position mean reaction times, $t_{n,i}(i = 1, 2, \ldots n; n = 1, 2, 3)$, where i = serial position and n = number of elements in the display:

$$t_{1,1} = t_0 + \frac{1}{u(1)}, \quad n = 1$$

$$\left.\begin{aligned} t_{2,1} &= t_0 + p_1(2) \times \frac{1}{u(2)} + [1 - p_2(2)] \times \frac{2}{u(2)}. \\ t_{2,2} &= t_0 + [1 - p_1(2)] \times \frac{1}{u(2)} + p_2(2) \frac{2}{u(2)}. \end{aligned}\right\} \quad n = 2$$

$$\left.\begin{aligned} t_{3,1} &= t_0 + p_1(3) \times \frac{1}{u(3)} + \frac{[1 - p_2(3)]}{2} \times \frac{2}{u(3)} + \frac{[1 - p_3(3)]}{2} \times \frac{3}{u(3)}. \\ t_{3,2} &= t_0 + \frac{[1 - p_1(3)]}{2} \times \frac{1}{u(3)} + p_2(3) \times \frac{2}{u(3)} + \frac{[1 - p_3(3)]}{2} \times \frac{3}{u(3)}. \\ t_{3,3} &= t_0 + \frac{[1 - p_1(3)]}{2} \times \frac{1}{u(3)} + \frac{[1 - p_2(3)]}{2} \times \frac{2}{u(3)} + p_3(3) \times \frac{3}{u(3)}. \end{aligned}\right\} \quad n = 3$$

Note that the RT for serial position is composed of contributions from occasions when it is processed at other locations in the processing order as well as the i^{th} location. It would not be correct to write, say,

$$t_{3,1} = t_0 + p_1(3) \times \frac{1}{u(3)} + \frac{[1 - p_1(3)]}{2} \times \frac{2}{u(3)} + \frac{[1 - p_1(3)]}{2} \times \frac{3}{u(3)}$$

The latter expression would result if we viewed the system as taking each serial position in turn and selecting a processing position for it. It seems more natural to assume that at each stage of processing, a serial position must be selected whose letter will be compared with the target.

The simplifying assumptions in the present serial model, namely, that if serial position i is not processed i^{th} (with probability $(1 - p_i(n)$ then any of the other serial positions are processed i^{th} with probability $1/(n - 1)$ should be viewed as an expedient to assess seriality and the magnitude of serial position effects (the latter via the values of the $p_i(n)$). There are two reasons for this. The first is that if we took the above assumption seriously, exceedingly strong consequences ensue. For instance, when $n = 3$, it follows from assuming that $\frac{1 - p_3(n)}{2} = 1 - p_1(3) - \frac{1 - p_2(3)}{2}$, that $p_1(3) = p_2(3) = p_3(3)$; that is, no serial position effects could be predicted. By allowing the probability of processing the k^{th} serial position n^{th} to be 1 minus the sum that it was processed earlier, *without* regard to $\frac{1 - p_n(n)}{n - 1}$ (e.g., $\frac{1 - p_3(3)}{2}$), we avoided this untoward result. The second reason for viewing the assumption as a convenient fit-device is that there seems to be no processing interpretation that yields nondegenerate values of the $p_i(n)$ (e.g., values such that $p_i(n) \neq p_j(n)$, $i \neq j$).

Estimation of t_0 was performed by subtracting the slope of the best fitting linear function (given for these data by $RT^+_n = 578.3 + 22.7n$) from the intercept. That is, the fit of the straight line gave an overall estimate of n which was then employed to estimate t_0. The intercept is *not* equal to t_0 for the + RT's since, as we have seen, the slope of a self-terminating serial model with $u(n) = u$ is $1/u \cdot 1/2$ rather than $1/u$ and, hence, the intercept is really $t_0 + 1/2u$. This would not be quite so simple if the RT function had not turned out so straight. Our policy here is to estimate t_0 based on the assumption of a constant u, then use this to estimate the p's and reestimate the u's for each n, thus providing a test of u's constancy over n.

Once t_0 is estimated in this fashion ($t_0 = 555.6$), the average serial position processing times are computed from $t_{n,i} - t_0 = x_{n,i}$ as are the average processing times for each n, $x_n = RT_n - t_0$. The latter is used to estimate $u(n)$ as we illustrate for $n = 1, 2$.

When $n = 1$

$x_1 = x_{1,1} = RT_1 - t_0 = t_{1,1} - t_0 = 1/u(1) = 601.61 - 555.6)$ msec $= 46.01$ msec and hence, $u(1) = .022$ letters/msec.

Similarly when $n = 2$

$x_2 = RT_2 - t_0 = 3/2u(2) = (621.56 - 555.6)$ msec $= 65.96$ msec so that $u(2) = .023$ letters/msec

and when $n = 3$

$u(3) = .022$ letters/msec.

Obviously the estimates of u are quite close to one another and thus the standard serial model with $u(n) = u$ really applies to this part of the data.

Recall next that imposed constraints on the p's require that

$p_2(2) = p_1(2)$

and

$$p_3(3) = 1 - \frac{1-p_1(3)}{2} - \frac{1-p_2(3)}{2} = \frac{p_1(3) + p_2(3)}{2}.$$

The value for $p_1(1)$ is just 1, of course, but for $n = 2$, we have now, using the above constraints,

$$x_{2,1} = p_1(2) \times \frac{1}{u(2)} + \left[1-p_1(2)\right] \times \frac{2}{u(2)},$$

$$x_{2,2} = \left[1-p_1(2)\right] \times \frac{1}{u(2)} + p_1(2) \times \frac{2}{u(2)},$$

which are two simultaneous equations, linear in $p_1(2)$ and $1/u(2)$. But, we have already estimated $u(2) = .023$ letters/msec and, hence, can solve for $p_1(2)$, which turns out to give $p_1(2) = .82$. That is, the estimated probability that the letter in serial position 1 is processed first, and that in serial position 2 is processed second, is quite high (.82).

When $n = 3$, the simultaneous equations become (where we denote the estimated u's appropriately)

$$x_{3,1} = p_1(3) \times \frac{1}{u(3)} + \left[1-p_2(3)\right] \times \frac{2}{u(3)} + \left[1-p_1(3) - \frac{(1-p_2(3))}{2}\right] \times \frac{3}{u(3)},$$

$$x_{3,2} = \frac{\left[1-p_1(3)\right]}{2} \times \frac{1}{u(3)} + p_2(3) \times \frac{2}{u(3)} + \left[1-p_2(3) - \frac{(1-p_1(3))}{2}\right] \times \frac{3}{u(3)},$$

$$x_{3,3} = \frac{\left[1-p_1(3)\right]}{2} \times \frac{1}{u(3)} + \frac{\left[1-p_2(3)\right]}{2} \times \frac{2}{u(3)} + \frac{\left[p_1(3) + p_2(3)\right]}{2} \times \frac{3}{u(3)}.$$

Substituting in $u(3) = .022$ and then solving for $p_1(3)$ and $p_2(3)$ yields $p_1(3) = .99$, $p_2(3) = .87$, and hence, $p_3(3) = .93$, thus illustrating within the context of this model the high likelihood that a subject processed from left to right with occasional switches of serial positions 2 and 3.

The reader may verify by substitutions the estimated p's and u's and t_0 into the original formula that this model fits the + RT data extremely well. All predictions were within one standard error (10 msec) of the observed serial position means. Even

though the number of parameters is greater by one than the degrees of freedom in the data, when 5_0 is included, a perfect fit is not guaranteed (for proof of this fact, see Townsend & Roos, 1973, Appendix A). Roughly, the reason is that the serial position means cannot be too different from one another without causing p and/or u to be less than zero.

The parallel model. The parallel model developed for application to the same data assumed (*1*) independent processing (the rates, v, constant over stages), but (*2*) with different processing speeds assumed for different serial positions $v_i \neq v_j$ and (*3*) with an explicit dependence of rate on n, $v_i(n)$, thus permitting a limited capacity self-terminating predictor. These assumptions resulted in a limited-capacity parallel model with the same number of parameters, when a t_0 is estimated for the residual RT effects, as for the preceding serial model. Unlike the serial model, however, this parallel model's structure is guaranteed to fit the present data since

$$x_{n,i} = \frac{1}{v_i(n)} = t_{n,i} - t_0$$

and $v_i(n)$ can be any positive number at all. Note that the expressions for $x_{n,i}$ just follow from the independent parallel assumptions giving self-terminating means of $1/v_i(n)$, when the target is placed in serial position i.

The overall mean RT here is just

$$\mathrm{RT}_n + = t_0 + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{v_i(n)},$$

the arithmetic average of the serial position means. We will not pursue the obvious fit technique here, but the reader may note that it can be done directly from the data.

At first glance, it might seem as if the obvious linearity in the data reduces the number of parameters more in the serial model than in the parallel model since there would be only one u instead of 3 (for $n = 1, 2, 3$). However, the linearity may be similarly employed in the parallel context.

Thus, linearity implies that the average self-terminating prediction (i.e., the means on target-present trials) must be proportional to n, that is,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{v_i(n)} = cn,$$

where c = positive constant. But this expression implies that one of the $v_i(n)$ may be written as a function of the others. Although we haven't space to show it here, this operation reduces the number of parallel parameters to that of the present serial model with a single value of u assumed.

Let's take a brief digression to discuss a couple of aspects of theorizing related to the above model testing.

One of the arguments sometimes advanced for employing serial models would seem to be that predictions of mean processing times can be derived without detailed particular probability distributions on the element completion times. This is because

of the additivity of the serial successive completion times. However, more detailed assumptions must be made if variances, for example, are to be predicted. In any case, this type of argument still appears weak, to the extent that as psychologists, we are interested in finding out how things actually work in the black box, not simply what is an adequate summarizing description. It may be feasible (but it is by no means guaranteed) to develop the mathematics and concomitant experimentation that will allow specification of ever narrowing classes of models or theories that detail the kinds of mechanisms and functions that must be present to yield behavioral data. In recent years, investigators in finite automata theory have worked out means of determining canonical sets of sequential machines that can predict observed input-output data (see e.g., Booth, 1967). Undoubtedly any type of analogous development in psychology will be an Herculean task in comparison and will likely be quite different in form. One detrimental factor is that often we must rely on summary statistics within a given experiment, rather than an entire scope of input-output sequences. Also, the sheer complexity of the human black box is almost dismaying. A somewhat ameliorative aspect is that we have an extremely rich spectrum of data, across conceivable experiments, to work with, that may provide an ever mounting set of converging operations.

Predictions of Target-Absent Data

At this point the parallel and serial models have done equally well with the RT^+ data. We could proceed to similarly fit the RT^- data, but to do so would not teach us much. Instead, we employ the parameters estimated above to predict the RT^- data, assuming the exhaustive serial and parallel models, since processing must be exhaustive when the target is absent from the second display.

The exhaustive serial prediction for a given n is just

$$RT_n^- = t_0^- + \frac{n}{u(n)}$$

which effectively becomes $RT_n^- = t_0^- + \frac{n}{u}$ because of the constancy of $u(n)$ in the RT^+ data. The corresponding prediction for the parallel model is quite hideous (the reader may refer to Townsend and Roos for the complete expression for arbitrary n). However, it was found that the exhaustive expression that is associated with rates that are constant over serial position gave quite similar predictions for our data, and hence, the following expression was used to predict the exhaustive processing times:

$$RT^- = t_0^- + \frac{1}{v(n)} \sum_{i=1}^{n} \frac{1}{i}.$$

This latter expression is found by using the techniques shown in the section on capacity (in particular see "Standard Parallel Models"). It can be seen that it was necessary to grant a different residual latency parameter to the exhaustive models, $_o > t_o^+$.

The results of these predictions for the RT^- data may be viewed in Fig. 6, with mild distaste if not outright revulsion. So much for typical serial *or* parallel models being able to handle *both* RT^+ and RT^- data, assuming self-termination on + trials! The large number of parameters estimated from the RT^+ data make the RT^- falsifying conclusions even stronger. For scientific purposes, the bigger the class of falsified models the better, and the smaller the class of supported models the better. If processing is really exhaustive on both + and − trials, then there is no discrepancy for the overall mean RT functions, but then where do serial position effects (the quite strong ones found here, for example) fit in? This outcome leads us to discuss more reasonable exhaustive alternatives below, as well as self-terminating models that can predict the present kinds of data. An alternative explanation is that serial position effects arise outside the comparison processing box, but there is as yet little or no direct evidence for this.

Bamber (1969), in discussion of same-different data, has suggested an interesting type of dual processing mechanism, one for target, the other for nontarget comparisons, that can predict equal-sloped + and − RT functions. One, the identity reporter, is fast and emits signals for "same" responses only. The other, a serial processor, is slow and emits signals for both "same" and "different" responses. These two

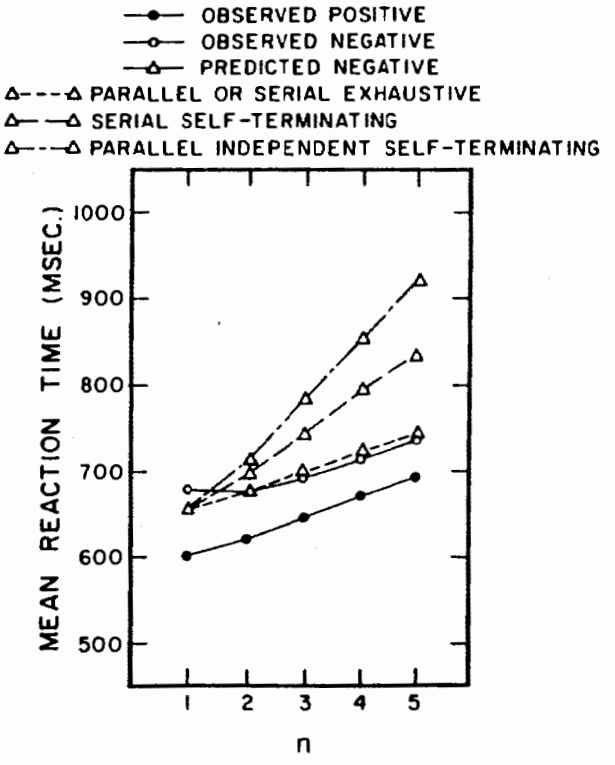

FIG. 6. Predictions of negative reaction times by self-terminating and exhaustive models based on parameter estimates from positive reaction times.

mechanisms appear to function more or less autonomously. They were able to predict the form of Bamber's results. It is of interest to consider alternative conceptualizations that do not require two separate processors, but yet allow for more latitude in thinking about processing + as contrasted with − comparisons. Specifically, Townsend and Roos proposed the hypothesis that − comparisons may take more time and/or their rates may change differently, as n varies, than do the + rates (that is, for example, $u^+(n) \neq u^-(n)$, where the superscripts refer to a match or mismatch respectively) because a unitary (single) processor is engaged in radically distinct tasks in the two cases. Measuring the fit of two things (letters) that match perfectly might require a different amount of time to achieve a certain degree of accuracy than measuring the fit of two things that mismatch to a greater or lesser extent. This notion can be discussed within either a template or feature-testing type of scheme, but is more general than either.

A Large Class of Falsified Serial Models

Not all models that allow for different rates on + and − comparisons are capable of explaining typical search results; quite the contrary. There is a large class of serial self-terminating models that cannot predict equal-sloped + and − RT curves as functions of n (Townsend & Roos, 1973, pp. 325–326 and Appendix B). This class of models is large in the sense that processing rates can vary with processing order and individual element, processing order may vary from trial to trial, − rates can differ from + rates, and the rates may vary as a function of n subject to the following constraint. The models are limited by what we have referred to as an "equal average time-increment assumption," which limits the *overall* effect that changes of rates with n can have (a capacity assumption). Basically the constraint means firstly that as n increases, the time to process a + comparison, averaged across element and processing order, remains constant. It is critical to perceive that the average time to process a + element alone (z, see Fig. 1) is being discussed, rather than the average total time to complete the positive trial. The latter duration is the self-terminating time, includes − comparisons, and may increase. Secondly, the constraint says that as n increases, the average time to process a − element remains invariant, again averaged over processing order and element. This is just like the + constraint, except that since nothing but − comparisons are processed on target-absent trials, multiplying the average − processing time by n gives the average exhaustive-processing time for that value of n. These two constraints represent a substantial generalization of the usual assumption that u is constant over everything. Aside from the point that many models are consequently falsified by most of the extant search data, it should be evident that the common feature of the falsified models is one of capacity changes in the model or system, nothing else.

There is, of course, nothing mathematically sacrosanct about equal sloped curves. If most investigators found curves with four-to-one slope ratios, then similar theorems could be proved about models that are falsified by these results. It *is* true that theoretically predicted equal-sloped curves and the class of models that is falsified by these both seem singularly simple and natural.

Three Classes of Models that Predict Equal Slopes and Serial Position Effects

With regard to variable and changing-rates kinds of models that *can* predict the dyad, serial position effects, and equal sloped + and − curves, there are two models that appear interesting to discuss; one is self-terminating, the other exhaustive. A third model employs overlapping comparison and response selection stages to achieve proper predictions. To lessen confusion, we shall intersperse this latter model between the discussion of the two models that put all the explanatory burden on the comparison process.

Within models that posit differential changes of + and − comparison rates as functions of n, either parallel or serial processing may be assumed. Further, the typically reported intercept difference between and − curves can be put in mechanisms external to comparison processing (whose effects are then clustered in t_0) or assumed to result from + and − rate differences that are present when $n = 1$ [for example, in $u^+(1) = u^-(1)$]. Because of limitations in space, we shall confine the present treatment to mentioning only the most central ideas and possible intuitions connected with the models.[14]

The self-terminating, changing-rates class of models. The first to be considered, the self-terminating changing-rates of comparison models, permit the + comparison rates to become smaller relative to the − rate, which effectively allows the ascent of the + RT function to increase and be equal to the − RT curve. For example, consider an independent parallel model of the sort fit to the data of this section, but assume in addition that although $v^-(1) = v^+(1) = v(1)$, in general, $v^+(n) \neq v^-(n)$, $n > 1$. For purposes of explication, we assume that rates are constant across element (or serial position) and processing order; these restrictions are absolutely not necessary but lead to simplification of the mathematics. The inclusion of variable rates over serial positions for the purpose of predicting serial position effects only lends added generality to the present models. Now, it turns out that if $v^-(n)$ decreases in less than inverse proportion to n, and $v^+(n)$ decreases in inverse proportion to n, that exactly equal + and − RT functions (except for the possible t_0 difference) can result. More precisely, the rates as functions of n that yield equal slopes are

$$v^-(n) = \frac{v(1)}{n} \sum_{i=1}^{n} \frac{1}{i}, \quad v^+(n) = \frac{v(1)}{n},$$

where $v(1)$ is the + and − rate at $n = 1$, and it can be seen that $v^+(n)$ decreases at a faster rate than $v^-(n)$. The assumption that $v^+(1) = v^-(1)$ is not necessary, but is convenient for illustration. Although we have given the illustration in terms of simple exponential intercompletion times, the qualitative result is more general than this. As we saw, independent parallel models that assume *unlimited capacity* at the individual

[14] The first model to be discussed, the self-terminating changing-rates model, is presented in Townsend and Roos (1973). All three models, the two changing-rates models and the nonDonderian model were recently given a general mathematical characterization (Townsend, 1973). It should be pointed out that D. A. Taylor has made progress in applying a particular form of nonDonderian model to visual search data (Taylor, 1973).

level and equal rates for + and − comparisons (e.g., $v^+(n) = v^-(n) = v$, a constant) predict flat self-terminating + functions and increasing negatively accelerated − functions (see "Standard Parallel Models"). The upshot of this state of affairs is that whatever curvature the equal-sloped exhaustive − and self-terminating + curves take, the average individual comparison times for − comparisons will never evidence greater increases (that is, a slow-down in rate) than will those for + comparisons.

Turning to serial models, we find a somewhat more complex situation but a similar outcome. The treatment is too detailed to be given here, but the method and a fairly general statement of results can be found in Townsend and Roos (1973, pp. 326–327). Under all reasonable conditions that the writer has been able to devise, equal slopes of + and − RT functions of n are associated with + comparison rates that slow down more, or speed up less, than − comparison rates.

What might be responsible for such changes in rates in typical search experiments? One possibility that comes to mind is closely associated with speed-accuracy tradeoff. In the usual experiment, the target and nontarget elements are drawn at random, but without replacement, from a fixed-sized alphabet with varying degrees of similarity among the members of the alphabet. In the present type of search task, the number of nontarget elements is increased to enlarge n, but there is always only one target element. Now, assume for the purpose of discussion that errors originate from confusions of nontargets with targets or vice versa, rather than from processes outside the comparison mechanism (e.g., the fast-guess model is of the latter type). Then the fact that in the present kind of search experiment there is always at most one target element, but the number of nontarget elements changes in order to vary n, may radically affect the relative average processing rates of + as opposed to − comparisons. Consider the fairly general case where either nontarget or target may contain both target and nontarget features on a given trial; nontargets may contain target features simply due to similarity with the target, and the target may have some nontarget features because of noise. Suppose further that feature matching between target and display (or memory) elements is all-or-none and is always performed correctly (these assumptions do not appear to be critical to the general outcome). Finally, suppose that the subject rejects an element in a self-terminating fashion as a nontarget if some criterial number of negative feature matches is attained and accepts an element as the target if some criterial number of positive feature matches is attained.[15]

[15] An aside here of some importance in the theory of feature testing relates to the number of matches or mismatches required for acceptance or rejection of the element, or pattern, in a pattern recognition context. It is sometimes assumed that exactly one mismatch is sufficient to reject the element. Obviously, if the situation is noisy, perfect accuracy cannot be so obtained; false negative responses (i.e., misses) would occur. Another assumption rather often found in the literature is that on individual elements, matching of + features must be exhaustive on target-target comparisons, the rationale being that less than complete matching by all the features may lead to false positives (i.e., false alarms). However, in a nonnoisy situation, where it never happens that features get added to the elements in the display or memory set, there will exist a subset of target features, the combination of which is unique to the target. Hence, if the subject is not merely capable of counting feature matches but also of taking note of what features match, then he can terminate when one of these subsets is completed. In many cases, there will be one or more single features that is unique to the target, and the efficient subject will be able, in principle, to terminate when he correctly matches one of these to its correspondent in the target.

In any case, the probability that any given criterial number of nontarget features is found in the target-target comparison will not increase with n, but the probability that the analogous criterial number of target features will be found in one of the $n - 1$ nontargets increases as n increases. Hence, if the subject interprets his instructions as meaning that he should maintain a low and constant error rate, he will have to change his feature criteria. Specifically, he may increase his criterion for the number of + feature matches required to accept an element as target, and this means a slow-down in the apparent + processing rate $[u^+(n)$ or $v^+(n)$ relative to the − rate $(u^-(n)$ or $v^-(n)]$. The difficulty with this explanation is that one would suppose that unless the − criterion is also raised, more false negatives would be found for larger n; the reason being that the − criterion would be reached more often on target-present trials due to the higher + criterion. Although the number of errors in the earlier discussed experiment was too low to permit a definitive error analysis (about 1.8%), it appeared that only one of the three subjects evidenced a real increase in false-negative errors across n. In any case, to obtain precise predictions, the rates should be applied at the feature-comparison level rather than the overall element-comparison level. At this point we must move on to the next set of models; the treatment will have to be even more laconic than the immediately preceding one.

A nonDonderian class of models. In the second of the three classes of models, we entertain the possibility that the equal slopes of the + and − RT functions come from two separate mechanisms, or stages in the overall processing chain, as opposed to being entirely in the comparison link. In particular, Donders' (1868) assumptions are purposely violated; namely, the stipulation that separate stages in processing do not overlap temporally but are rather distinct and additive, and, secondly, that these additive steps are independent. Two side comments are pertinent here. One is that *within* the comparison stage alone, serial processing is in accord with the additive assumption of Donders. However, the usual assumptionof independence, in this case among individual element processing stages, may not hold if there is a correlation among the intercompletion times (which in the serial instance are also the individual processing times, the z's; see "The Independence vs Dependence Issue"). Conversely, parallel processing is certainly not of a Donderian nature but can clearly yield results that manifest an appearance of additivity and independence. Thus, the non-Donderian quality of the present model is placed in the relationship between the comparison stage and *another* stage, specifically the *response selection* stage.

The present models represent the selection process as beginning during the comparison stage and as possibly biasing toward or preparing to make one or the other of the two responses, depending on the specific experimental design. In the case of search for single targets, it is plausible that before any element is processed the subject is equally disposed to making an affirmative or negative response, since the probability of a target-present trial is typically 1/2. However, as more and more comparisons are completed, the more likely it is that the set of elements on that trial does not contain the target. Thus, acting in an efficient manner, the subject's response selection mechanism may be readying to perform the "target absent" response until such time as either the target is located and the "target present" response ensues or all

n elements are processed as − matches and the − or "target absent" response takes place.

General but convenient mathematical models that contain these ideas (Townsend, 1973) may be constructed. This is accomplished by a decomposition of the total processing time (ignoring t_0, which now does not contain response selection but still includes the actual motor response time) into the comparison time plus the amount of time that is required to select the proper response after the comparison process is completed. The latter duration is shortened on − response trials by an amount depending monotonically on how many negative elements have been processed; this number is always n on − trials. The response selection duration will be lengthened by an amount depending on how many − elements were processed before the target was located on + trials. The interesting and critical aspect of this type of model, as far as working out predictions goes, is that the total response selection time, including that part overlapping with the comparison stage, may be written as a sum of 2 (random) variables. The first is the time required if no preselection biasing or asymmetry in readiness occurs, plus or minus that amount of time saved or wasted during the comparison stage. Thus, if no preselection biasing occurs, then the random variable Y represents the amount of time necessary to choose the response. However, on a − trial, the quantity $g(n-1)$ may be subtracted, where $g(n-1)$ is a random function whose mean may be assumed to increase as n increases. The reason that $n-1$ is employed rather than n is that we view the subject as going ahead with the response selection as soon as the *nth* element is completed, so the results of the *nth* comparison cannot further bias the selection mechanism. So, on a given trial the total time consumed by response selection is given by $Y - g(n-1)$.[16] On + trials, the biasing that occurs actually hurts or slows down the time to choose and then make the + response. If the target is processed first, then no biasing or predisposition will have a chance to occur; otherwise, the selection time will be $Y + g(k)$, where k is the number of − comparisons completed before the target is found. Again, we think that g will be larger on the average when k is larger. The + response selection process thus has to compensate for the increasing predisposition toward making a − response.

How does this relate to parallel + and − RT functions of n? The effect of the predisposition towards negative responses within trials is to increasingly slow down + RT's relative to − ones as n grows larger; obviously the larger n becomes, the more of a chance the bias process has to operate. We have already seen that decreasing the speed of + times relative to − times can equalize the slopes of the RT functions. In fact, it is possible to determine the relationship between the comparison process and the selection bias process that results in parallel or equal sloped + and − RT functions of n (Townsend, 1973). As in all three of the classes of models we are presently discussing, it doesn't matter at all whether comparison processing is assumed to be parallel or serial, although the exact relationship between g and the comparison rates will depend on this aspect of the model. Similarly in the present

[16]Observe the difference $Y - g(n-1)$. Since $Y - g(n-1)$ cannot be less than zero it is necessary to assume in an exact account that Y has some lower bound, say Y^*, and the $g(n-1)$ is always less than or equal to Y^*.

class, self-termination and exhaustiveness are both compatible with its nonDonderian characteristics.

In closing our discussion of this class of models, we might make a remark about the general epistemology of the mind and brain. Although employing Donderian types of assumptions (see, e.g., Sternberg, 1969; Pachella, this volume) has been quite fruitful and will undoubtedly aid us substantially in the future, it appears likely that much of the brain's work must be carried on in parallel and in an integrated, highly correlated fashion. When we are able to go beyond relatively simple (but to be sure very complex nevertheless) processing chains we shall probably find more and more activity proceeding simultaneously. And, the fact that flat RT functions as n is varied are sometimes obtained in cognitive tasks provides fairly strong support, as noted earlier, for the proposition that consciousness can sometimes be directed to more than one thing at a time (see, e.g., Egeth, Jonides, & Wall, 1972). Finally, from a physiological point of view, some of the scanning or comparison times reported seem to be faster than known neural structure is capable of producing in a serial fashion (see, e.g., Anderson, 1973).

A class of exhaustive models that predict serial position effects. One of the most entrenched assumptions in theorization about information processing and cognitive behavior concerns the meaning of spatial and temporal effects associated with experimental placement of target stimuli. Specifically, it is generally accepted that if serial position (or comparable) effects are to be explained within a processor assumed responsible for a given task, that these serial position effects imply *self-termination*. This is false. It *is* true that self-termination, along with a non-uniform distribution of attention (parallel) or preferred processing sequences (serial) can predict serial position effects. It is also true, contrary to the usual supposition, that *exhaustive* systems can produce serial position and like effects. This section is devoted to a brief discussion of such systems and models.

As usual, we can mathematically develop both parallel or serial models that make these predictions, but we confine our present discussion to serial models. All the exhaustive serial models that predict serial position effects *and* equal sloped + and − RT functions of n have the following two properties: (*1*) a difference in the positive processing rate and the negative processing rate at a given serial position, to yield serial position effects, and (2) the difference in the average individual + processing times from n to $n+1$ elements (n greater than or equal to 1) is equal to the difference in the average individual − processing times from n to $n+1$ elements, to yield equal sloped + and − RT functions of n. We mention two special cases that appear of interest. We concentrate on how the serial position effects occur since this aspect is probably of most import for exhaustive models, and the techniques of investigating slope equality have been given earlier in the chapter.

In the first special case, it is assumed that the − rates are constant for a given value of n but that the + rate differs across the serial positions. The serial position effects will naturally directly reflect the + speed of processing on the various positions. In contrast, the second special case produces serial position effects that are inversely related to processing rates. This type of model assumes that the + rate is constant over

serial positions for fixed n but that the − rate is not. Since it seems unlikely that the + rate is much slower than the average − rate, assume that in fact the + rate is faster than the − rate. Then a novel and intriguing aspect of this model is that the serial position effects are given by virtue of the property that when the target is in a given position, the − rate for that position does not occur (since a mismatch between target and an item in that position does not occur). Therefore, the serial position effects, revealed in the data, are actually in the *opposite* order of the − processing rates. For example, suppose that the − rates are slowest for the first serial position and graded up to being fastest at the last serial position; that is, there is a strong recency effect and no primacy effect. Then these models say that the serial position effects will show fastest times on target-present trials when the target is in the first serial position and slowest times when it is in the last position. Our conclusions about − processing rates must then be opposite of what we observe in the data! There are other interesting exhaustive models, many of them more general than the two included here, but perhaps the latter will serve to illustrate some of the attendant flavor and intuition.

The three classes of models discussed in this section seem to the writer to contain some aspects of parsimony and plausibility. They are all capable of quantitatively explaining reported search phenomena from basically divergent assumptions. They are geared around a kind of rough approximation to a minimal systems approach, in the sense that only few mechanisms are involved and they appear to function in a fairly reasonable way. Other hypotheses, such as repeated scans through a list, may be viable in certain circumstances. Also, although the kinds of ideas discussed here seem simple from the viewpoint of what the overall system or subsystem is doing, the way what it is doing is evidenced in the mathematics may be complex at times. When there is a high degree of complexity, there is often an increased problem of parameter estimation and thus an implicit challenge to select experimental conditions and stimulus materials in such a way as to reduce the number of parameters and test the basic assumptions of the models. Experimental testing even of only those three classes of models presented here seem formidable. Including predictions for errors may help if it turns out that the small error rates reported in brief visual-display search and/or short-term memory search are related to comparison processing or, in the case of the nonDonderian model, to response selection. In any case, error predictions should certainly aid model testing in the case of long-term memory search.

We should finally mention that critical experiments are needed to see how short-term memory search differs from visual search, if at all. The learning curve for memory search RT appears to be much more extended than that for visual display search (Townsend & Roos, 1973, pp. 323–324), but evidence concerning potential differences in actual comparison processes should also be helpful. There is also the possibility that we have, most of us, overemphasized the processing of both types of comparisons, + and −, in any search experiment. It may be efficient for the subject, in experiments involving at most the presence of a single target, for example, to look actively for a + match and to disregard elements with few or no matching characteristics (e.g., features). "Target absent" responses might then be made after some criterion time has elapsed with no target detected. In support of this possibility is the

fact that the subjects often tell of difficulty in recalling elements that were present in a list but did not match the target. To some extent, subjects may be able to tailor their processing strategies according to specific task demands.

SUMMARY

The goal of this chapter has been first to present a coherent and empirically relevant development of the issues of serial vs. parallel processing, independent vs. dependent processing, self-terminating vs. exhaustive processing, and unlimited vs. limited capacity, all of them appearing to be critical in the modeling of information-processing systems. The second goal, closely bound to the first, has been to attempt to make a strong case for the proposition that these issues are logically, mathematically, and systemically independent. Within these broad aims, many models were examined and problems of model equivalence were treated in detail. In the section on applications, an independent parallel and a typical serial model, each self-terminating, were fit to some recent search RT data, and it was found that both could predict linear target present (+) RT's as n (number of elements) was varied. However, if it was assumed that the same rates were applicable on target absent trials, then neither model could predict the − RT's. It was shown that a large class of serial self-terminating models is falsified by contemporary data, and then several models that are capable of predicting these experimental results were discussed.

REFERENCES

Anderson, J. A. A theory for the recognition of items from short memorized lists. *Psychological Review*, 1973, **80**, 417–438.

Atkinson, R. C., Holmgren, J. E., & Juola, J. F. Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, 1969, **6**, 321 – 326.

Bamber, H. K. Reaction times and error rates for "same-different" judgments of multi-dimensional stimuli. *Perception & Psychophysics*, 1969, **6**, 196 – 174.

Booth, T. L. *Sequential machines and automata theory*. New York: Wiley, 1967.

Boring, E. G. *A history of experimental psychology*. New York: Appleton-Century-Crofts, 1950.

Broadbent, D. E. *Perception and communication*. New York: Pergamon Press, 1958.

Christie, L. S. & Luce, R. D. Decision structure and time relations in simple choice behavior. *Bulletin of Mathematical Biophysics*, 1956, **18**, 89 – 112.

Corcoran, D. W. J. *Pattern recognition*. Middlesex, Pa.: Penguin, 1971.

Donders, F. C. Die Schnelligkeit psychischer Processe. *Archiv für Anatomie und wissenschaftliche Medizin*, 1868, 657–681.

Egeth, H. E. Parallel vs. serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1966, **1**, 245 – 252.

Egeth, H. E., Jonides, J. & Wall, S. Parallel processing of multi-element displays. *Cognitive Psychology*, 1972, **3**, 674–698.

Erikson, C. W., & Spencer, T. Rate of information processing in visual perception. Some results and methodological considerations. *Journal of Experimental Psychology Monographs*, 1969, **79** (2, Pt. 2).

Estes, W. K. Interactions of signal and background variables in visual processing. *Perception & Psychophysics*, 1972, **12**, 278 – 286.

Estes, W. K., & Taylor, H. A. A detection method and probabilistic model for assessing information from brief visual displays. *Proceedings of the National Academy of Sciences*, 1964 **52**, 446 – 454.

Estes, W. K., & Taylor, H. A. Visual detection in relation to display size and redundancy of critical elements. *Perception & Psychophysics*, 1966, **1**, 9 – 16.

Estes, W. K., & Wolford, G. L. Effects of spaces on report from tachistoscopically presented letter strings. *Psychonomic Science*, 1971, **25**, 77 – 80.

Gardner, G. T. Spatial processing characteristics in the perception of brief visual arrays. Unpublished doctoral dissertation, University of Michigan, 1970.

Gardner, G. T. Evidence for independent parallel channels in tachistoscopic perception. *Cognitive Psychology*, 1973, **4**, 130 – 155.

Green, D. M., & Luce, R. D. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, **79**, 14–57.

Gumbel, E. J. *Statistics of extremes*. New York: Columbia University Press, 1958.

Hamilton, W. *Lectures on metaphysics and logic*. Vol. 1. Edinburgh: Blackwood, 1859.

Hunter, W. S., & Sigler, M. The span of visual discrimination as a function of time and intensity of stimulation. *Journal of Experimental Psychology*, 1940, **26**, 160 – 179.

Hylan, J. P. The distribution of attention I. *Psychological Review*, 1903, **10**, 373 – 403.

Kahneman, D. K. *Attention and effort*. Englewood Cliffs: Prentice-Hall, 1973.

Neisser, U. Decision time without reaction time: Experiments in visual scanning. *American Journal of Psychology*, 1963, **76**, 376 – 385.

Parzen, E. *Modern probability theory*. New York: Wiley, 1960.

Rumelhart, D. E. A multi-component theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, **7**, 191 – 218.

Sperling, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74**(11, Whole No. 498).

Sperling, G. A model for visual memory tasks. *Human Factors*, 1963, **5**, 19 – 31.

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

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

Sternberg, S., & Knoll, R. L. The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1972.

Taylor, D. A. Questioning the assumption of selective influence in the analysis of search and matching tasks. Paper presented at the sixth annual Mathematical Psychology Meeting, Montreal, August, 1973.

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

Thomas, E. A. C. Alternative models for information processing: Constructing nonparametric tests. *British Journal of Mathematical and Statistical Psychology*, 1969, **22**, 105 – 113.

Townsend, J. T. Binocular information summation and the serial processing model. *Perception & Psychophysics*, 1968, **4**, 125 – 128.

Townsend, J. T. Stochastic representations of parallel and serial processes. Paper presented at Mathematical Psychologists Conference, Bloomington, Indiana, 1969. (a)

Townsend, J. T. Mock parallel and serial models and experimental detection of these. *Proceedings of the Purdue Symposium on Information Processing*, 1969, **2**, 617–628. (b)

Townsend, J. T. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 1971, **9**, 40 – 50. (a)

Townsend, J. T. A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, 1971, **10**, 161 – 163. (b)

Townsend, J. T. Some results on the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 1972, **25**, 168 – 199. (a)

Townsend, J. T. Identifiable parallel and serial models for same-different matching experiments and a paradigm for testability at the level of mean processing times. Paper presented at the fifth annual Mathematical Psychology Meeting, La Jolla, California, August, 1972. (b)

Townsend, J. T. Solutions to a functional equation necessary for parallel-serial equivalence. Unpublished manuscript. Purdue University, 1972. (c)

Townsend, J. T. Three models for search reaction-time performance. Paper presented at the sixth annual Mathematical Psychology Meeting, Montreal, August, 1973.

Townsend, J. T. A stochastic theory of matching processes. Unpublished manuscript, Rockefeller University, 1974.

Townsend, J. T., & Roos, R. N. Search reaction time for single targets in multiletter stimuli with brief visual displays. *Memory & Cognition*, 1973, **1**, 319–332.

Townsend, J. T., Taylor, S. G., & Brown, D. R. Lateral masking for letters with unlimited viewing time. *Perception & Psychophysics*, 1971, **10**, 375 – 378.

Treisman, A. M. Strategies and models of selective attention, *Psychological Review*, 1969, **76**, 282–299.

Wolford, G. L., Wessel, D. L., & Estes, W. K. Further evidence concerning scanning and sampling assumptions of visual detection models. *Perception & Psychophysics*, 1968, **3**, 439 – 444.

Yellott, J. I. Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, 1971, **8**, 159 – 199.