

SOME RESULTS CONCERNING THE IDENTIFIABILITY OF PARALLEL AND SERIAL PROCESSES

By J. T. TOWNSEND
Purdue University

A mathematical characterization of serial, parallel and hybrid processes is given, and this characterization is related to several current experimental paradigms. Non-identifiability (mimicking) between two systems (i.e. models of systems) is defined as equivalence of probability distributions on element completion times for the two systems, where n elements are available for processing by each. Results are then presented for a class of systems with exponential processing times, and it is seen that several interesting cases of parallel and serial systems are equivalent to systems of the opposite type. Evidence that will allow accurate discrimination between parallel and serial processing for this and other classes of systems either requires more complete and precise information about the actual probability distributions of the systems or more specialized sets of converging operations than is usually obtained in psychological experimentation. For example, it is noted that at the level of first moments (means), even a parallel independent system can predict results usually associated with a serial system (an overall increasing linear mean reaction time curve as a function of the number of elements to be processed). Next, a functional equation is developed that must hold in order for mimicking to occur between parallel and serial systems within the same general family of probability distributions, and three special cases are examined. A parallel system with gamma-distributed processing times for element completion is then investigated, and it is shown that a strictly serial system cannot mimic it, but an interesting hybrid system can. This is followed by discussion of two kinds of partial identifiability, mimicking at the level of means and possible predicted differences at higher levels, and mimicking by approximation. Some qualitative considerations that may enter into conclusions as to parallelity or seriality of processing are then introduced. Last, it is suggested that in a broad sense questions related to parallel and serial systems concern fundamental aspects of information-processing structure and distribution of processing energy and hence merit further mathematical investigation.

1. INTRODUCTION

Behavioural psychology can be characterized as developing laws that relate input variables to output variables. Whether the intervening black box that is the organism is relatively empty or contains rich hypothetical structure, the theorist is mainly interested in functional relationships. Unfortunately, as every theorist knows, a set of existing data can usually be predicted or described by more than one theoretical structure. The social sciences seem especially susceptible to models or theories with strongly differing assumptions, yet which are capable of predicting similar results for a number of dependent variables. This problem is especially endemic in the realm of highly verbalized, loosely

structured theories, which may appear to be distinct because of connotations and differing intuitions but, to the extent that they can be formalized, may be logically similar. However, even the more rigorous approaches to psychology are less insulated from such difficulties than is sometimes recognized. We will suggest here that these problems exist in a strong sense for parallel and serial models based on opposing axioms, and that consideration of such models that make identical predictions, and yet differ on a vital assumption, can aid our understanding of what this assumption means behaviourally and perhaps facilitate model testing.

Since the models which belong to one class never belong to the other, the problems of testability are not simply due to semantic difficulties. In addition, as a substantial amount of current experimental work is being directed to the question of whether certain psychological processes are serial or parallel, it is pertinent to attempt an answer to the question of what we can and cannot conclude from experimental results about the parallel or serial nature of the systems under scrutiny.†

Within the area of probability theory and its applications, these difficulties are referred to as problems in identifiability. Unfortunately, results are known only for some special cases (for an example in psychology, see Greeno & Steiner, 1964). Furthermore, many mathematical results for stochastic systems have only been obtained for systems which are in operation long enough to achieve stability or equilibrium, with fairly gross measures such as the pooled output of all the component processes of a parallel system, or the overall counting process of a serial system, or for particular features such as the mean and coefficient of variation (see Saaty, 1961). We shall here be working with systems containing a finite number, probably small in practice, of entities, hereafter referred to as elements (e.g. letters, words, values on one or more dimensions, etc.), which are to be completely processed, and we shall be interested in the mathematical description of the systems. By mathematical description we mean the specification of the probability distribution of the processing of the elements sufficient to compute any possible statistic associated with the parameters of the system.

Although investigators seem to know what they mean when using the terms 'parallel' and 'serial' with respect to underlying mechanisms, it will make our task easier if we attempt to give a more formal description of them.

2. CHARACTERIZATION OF PARALLEL AND SERIAL PROCESSES

We first express the 'thing' that may be operated on in serial or parallel as a super-vector, i.e. a vector of n vectors, each constituent vector of which constitutes one of the elements. Call the super-vector V , the component vectors X_i , and the components of X_i , a_j so that $V \equiv \langle X_1, X_2, \dots, X_{n-1}, X_n \rangle$ and $X_i \equiv \langle a_1, a_2, \dots, a_{n_i} \rangle$. To provide a milieu for our intuitions, we treat V as an input vector operated on in parallel or serial fashion by a black box (the processor 'G')

† A qualitative summary and discussion of application of these and related results may be found in Townsend (1971c).

which yields a corresponding output vector, $G(V; t) \equiv G(X_1, X_2, X_3, \dots, X_n; t)$, which gives the state of processing for each of the elements at time t . For example, if processing takes place on an all-or-none basis, we could let 0 represent incompleteness and 1 completion; thus $G(X_1, X_2, X_3; t) = \langle 0, 1, 1 \rangle_t$ shows that elements X_2 and X_3 are completed by time t , but X_1 is not.

Interpretations for some psychological experiments will now be given. The following specifications are not the only ones possible under the present characterization, but may suffice to illustrate the ideas.

2.1. *Multidimensional Display Perception*

In this kind of experiment, one typically arranges for the stimuli to vary either in the number of displayed dimensions or in the number of dimensions relevant to some task. Then predictions are made for some dependent variables, such as reaction time, in terms of how they should vary as a function of the number of dimensions (e.g. colour, size, shape and so on), given one or more serial and parallel models. An example of such an experiment is given in Egeth (1966), in which the author requires a same-different judgement on pairs of stimuli, and manipulates the number of dimensions on which they differ, recording the reaction times and comparing them with theoretical predictions. If colour and size were the relevant dimensions, each with two possible values, the possible stimuli and their description would be:

$$\begin{aligned} V_1 &= \langle X_1, Y_1 \rangle = \langle \langle a_1 \rangle, \langle b_1 \rangle \rangle, \\ V_2 &= \langle X_1, Y_2 \rangle = \langle \langle a_1 \rangle, \langle b_2 \rangle \rangle, \\ V_3 &= \langle X_2, Y_1 \rangle = \langle \langle a_2 \rangle, \langle b_1 \rangle \rangle, \\ V_4 &= \langle X_2, Y_2 \rangle = \langle \langle a_2 \rangle, \langle b_2 \rangle \rangle, \end{aligned}$$

where the V_i are the four stimuli, X_i and Y_i refer to colour and size, respectively, and a_i, b_i ($i = 1, 2$) are the different values for the two dimensions. In this particular example, X_i and Y_i are one-dimensional vectors with each dimension taking on one of two values.

2.2. *Multisymbol Processing Experiments*

An example of this kind of experiment is provided by Sternberg (1966). The subject is presented with a small number of symbols, say letters, which he is capable of memorizing rapidly. He is next given a test letter and must then respond whether it was or was not among those in the memorized set. The number of letters in the memorized set is varied, and predictions about reaction time are tested. Here, the subscript on X_i could refer to order of presentation in a temporal or spatial sense (e.g. in a left to right order, X_i is the i th letter). At the contemporary level of theorizing for this type of experiment, specification of the features making up a letter in position i has not been considered necessary. The present specification is sufficient to consider spatial and temporal position effects.

A logical precursor of this type of design was the 'detection' method originated by Estes & Taylor (1964) to study the visual span of apprehension. In this

paradigm, the multisymbol array (typically a matrix of letters) is presented after one of two test letters is shown, and the subject must report which test letter was in the array. The major dependent variables have been probability correct and reaction time. The appropriate specification for this experiment is similar to the preceding one with n equal to the number of letters in the array. One could specify the elements by two subscripts to indicate position in the visual array. A derivative of the Estes & Taylor and Sternberg methods employs linear arrays and reaction time with essentially perfect accuracy (Atkinson *et al.*, 1969). In this method, the test letter is given *prior* to a briefly displayed multi-letter array, and the subject answers whether or not the test letter appeared in the display.

2.3. Character Recognition Experiments

The type of character recognition experiment of most interest is perhaps that involving complete identification, where for every stimulus presented there is exactly one response and vice versa. An example is recognition of English alphabetic letters (e.g. Townsend, 1971*a, b*) with possible dependent variables of reaction time, confusion probabilities, or both. On each trial, let $V_i = \langle X_i \rangle = \langle \langle a_{i1}, a_{i2}, \dots, a_{in_i} \rangle \rangle$, where a_{ij} is the j th feature of character X_i and n_i the number of such features in X_i . An especially simple example would be where $n_i = n_j = n$, and a_{ij} specifies the presence or absence of a particular feature with a 1 or 0 for character X_i , and hence X_i is a vector of 0's and 1's. This type of description could serve as a basis for designing alphabets with the use of hypothetical features and/or for the theoretical analysis using feature-testing models of recognition. It is interesting to note that within the usual simplistic interpretation of template matching as involving an all-or-none match of the input pattern to the memory template, the mathematical specification is identical to that for single-featured characters with all-or-none testing. In that (mathematical) sense, template matching is a special case of feature testing and hence easier to falsify.

In examples (1), (2), $G(V; t)$ represents the state of processing of the elements X_i at time t , and the specification of a probability distribution, $P[G(V; t)]$, for every $t > 0$ and for all defined X_i is sufficient to obtain a complete description of the system. Similarly, in example (3) a complete description of $P[G(V; t)]$ will give the probability distribution on the possible state of processing of the a_{ij} 's for every $t > 0$.

The following will discuss parallel and serial processing relative to the X_i but exactly analogous statements hold for processing of the components of the X_i . In fact, we make the assumption that our level of analysis is sufficiently fine that we need only consider all-or-none processing. Thus if we wish to study partial processing of complex entities, we may study all-or-none processing of the constituents that make up the complex entities. Such an approach is theoretically unjustified if, for example, instead of X_i we have $X(r)$, a continuous function of $0 < r < K$, where r and K are real numbers. Nevertheless, by picking points sufficiently close in r , we can generate reasonable approximations for realistic cases.

A parallel process will mean that the X_i are operated on simultaneously so that processing begins and proceeds simultaneously on all the n elements but may finish on different elements at different times. In contrast, in a serial model, processes act on one element, and thus on one X_i , at a time, completing one before moving on to the next.

Fig. 1 is a representation of a parallel process for V for three elements, where G is given in terms of its three components functions. The parallel system is shown as having three channels but with the connections of the channels providing for possible interactions or correlations.

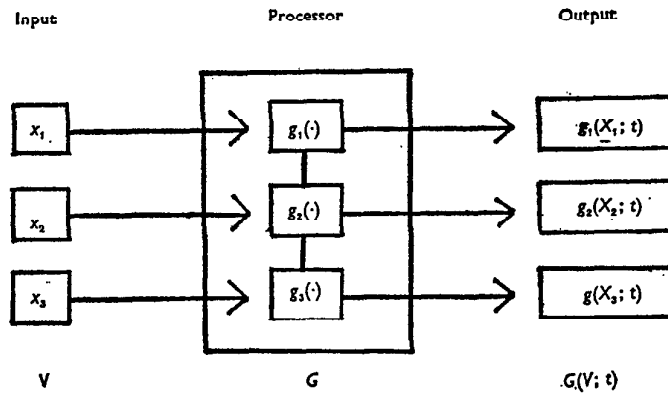


FIG. 1.—Schematic for a parallel processor.

Similarly, Fig. 2 shows a serial process, where we picture the components of V as having queued up to be processed one at a time, and with outputs emerging in the order in which they were processed.

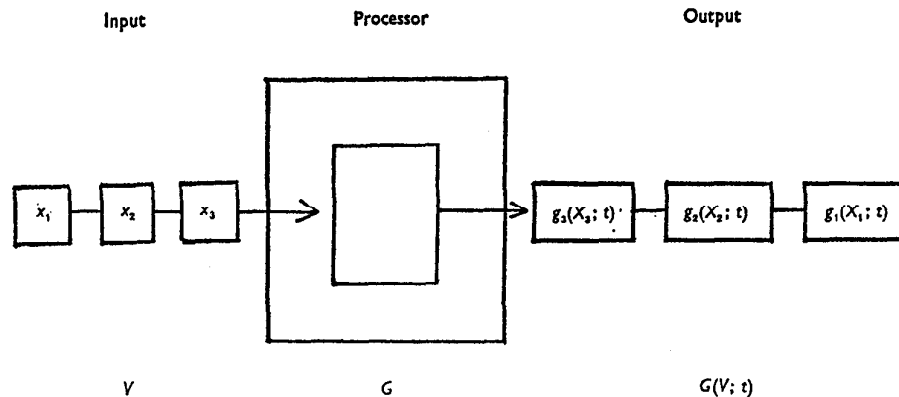


FIG. 2.—Schematic for a serial processor.

We can now give a more precise definition of identifiability as used to obtain the present results. When two classes of systems are specified, one class serial and the other parallel, given a member system of one class, and there exist no members of the other which predict the same joint probability distribution of element completion times, that system is said to be identifiable (i.e. it cannot be

mimicked within the specified class of the other type); otherwise it is non-identifiable. 'Identifiability' is typically defined on a single class of systems (Koopmans, 1950; Greeno & Steiner, 1964), the question being whether or not members of the class can be distinguished from one another. Hence the present use is an extension to questions of uniqueness between classes.

We assume that one is unable to observe directly the processing and thereby ascertain its spatio-temporal nature. Thus when identification is possible, it depends on a particular observation (e.g. minimum completion time or joint intercompletion time distributions), but when non-identifiability holds, it holds *a fortiori* for all conceivable experiments for the present class of processes.

The parallel systems on which we will concentrate mathematically will include the assumption that given any particular set of elements already processed and their order of completion, the set of processing rates for the remaining elements will be constant. This means that the parameters must be considered stable, and not to change adventitiously (e.g. depend on the future). This notion will be made more rigorous in our special cases. We will also consider probability mixtures of parallel systems. The mixture will fulfil our definition of a parallel system, since on any one trial only one of the member processes would be in operation (with some fixed probability). Each member process can possess a distinct set of processing rates within the above constraint. Similarly, a mixed serial system will allow probability mixtures of different processing orders with different rates, for the separate elements and separate processing permutations.

It is possible to pose hybrid models, such as a time-sharing process that works on one element, then goes to another without necessarily completing the first, and so on. Similarly, one might wish to consider models that are parallel with respect to the separate X_i , but serial with respect to the units making up the X_i vectors, or vice versa. The investigation of equivalence for these more complex models may prove difficult, but it appears that it will not be unusual for two models to make identical predictions on at least some of their statistics. For example, it will be shown below that there are time-sharing models, in the sense given above, with individual element processing identical to those of some parallel models.

In relating $G(V; t)$ to the processing of symbolic information in an actual experiment, we must specify the processes that precede and follow $G(V; t)$ so that their effects may be distinguished from those of $G(V; t)$ itself. For example, in the multisymbol visual perception experiments, it may be necessary to take note of the possible dependence of V upon time (Townsend & Fial, 1968). That is, if the legibility of the symbols changes for small values of t , as when photochemical reaction is not in equilibrium, then either we must write V itself as a function of time, $V(t)$, or must use display durations for which this effect is small. Similarly, Averbach & Sperling (1961) showed that special techniques were necessary (the partial report methods) to eliminate the lower bounds imposed by what they interpreted as subsequent memory capacity. This was done in order to investigate the size of the vector V (i.e. how many X_i 's were

available for the processor to work on). Thus we may relate such observables as reaction time to waiting time distributions for particular values of $G(V; t)$; the probability of being correct on, say, a symbol in a certain location for a given display duration, T , to the marginal expected value at that location, $E\{g(X_i; T)\}$, and so on.

The actual cases for which our first results are obtained are simple, being concerned with homogeneous, but not necessarily stationary, exponential processing: it is also possible to extend our results to certain non-homogeneous exponential completion times. (The reader is referred to standard works on stochastic processes for definition of these and related terms.)

There are four primary reasons for being interested in exponential processing. Much recent psychological theorizing emphasizes processes made up of exponential completion times (see, for example, McGill, 1963); some more complex distributions are themselves based on a series of exponential times, such as a special Erlangian (gamma) distribution (Cox, 1962; McGill & Gibbon, 1965), and non-identifiability at the exponential level implies non-identifiability at the more complex level; the mathematical properties of the exponential distribution are relatively simple; and our recent investigations suggest that at least some distributions may be more easily identified than the exponential.† The last reason is related to the fact that current experimental techniques do not allow us to obtain complete descriptions of the underlying probability distributions; in fact, means are frequently the only statistic employed. Since non-identifiability seems to be a particularly serious problem with exponential distributions, these should provide a conservative view of the possibilities for testing between serial and parallel processes.

We desire a means of investigating the complete equivalence of parallel and serial systems; this requires a specification of the joint probability distribution on the various elements to be processed. The approach employed for the exponential systems will be to consider the joint characteristic function of the times between the completion of processing of the individual elements. Although, unless otherwise noted, the results will be valid for an arbitrary number of elements, we shall here work with three element inputs, enough to illustrate the structure of the systems but much less cumbersome in notation than the general case. (For a method of proof for the case of n elements using cumulative distribution functions, see Townsend, 1969.) First, we characterize our special classes of parallel and serial processes.

3. GENERAL HOMOGENEOUS UNMIXED EXPONENTIAL PARALLEL AND MIXED SERIAL SYSTEMS

Consistent with our earlier definitions, we assume that all elements begin their processing simultaneously in the (unmixed) parallel system and that they finish

† Thus the density function for the 'gamma' distribution with parameters k , λ is

$$f_k(t) = \frac{(\lambda t)^{k-1} \lambda \exp(-\lambda t)}{\Gamma(k)} \quad (k > 0; \lambda > 0).$$

at times with exponential distributions. We allow the individual elements to possess different processing rates at different stages; i.e. each element can have a unique parameter depending on how many elements have already been processed, and depending on which elements have so far been processed. However, we conceptualize realistic parallel processes as constraining the parameters by the past. By this we mean that each time the system is run through n elements the processing rates for each element until the first is finished are the same as at any other time (although they may be distinct across elements). Similarly, given that a particular element finishes first, any other element must always have the same processing rate irrespective of which element happens to finish second, and so on. Thus, in an unmixed parallel system, processing rate may depend on the past but not on the future as the state of processing evolves. This formulation results in a total number of $n! \sum_{k=0}^{n-1} (1/k!)$ parameters.

As noted earlier, our characterization of serial systems will allow for the possibility of different orders of processing which may occur on different 'trials' according to some probability distribution. In general, for the unmixed serial system, it is reasonable to allow different processing rates of each element for each possible processing order. Note that since we conceive of a mixed serial system as possessing the ability to select a processing path, or permutation, before processing begins, the serial system is not constrained by the past in the same sense as the unmixed parallel system is. This allows for a total number of $(n+1)! - 1$ parameters made up of $n(n!)$ processing rates and $(n! - 1)$ probabilities. Since as n becomes arbitrarily large, the ratio of the number of parallel parameters to the number of serial parameters approaches $\{1/(n+1)\}e$ (where e is the base of natural logarithm), and this term in turn approaches zero, the possibility is suggested that the present type of serial system is more general than the parallel system, and that there exist serial systems k for which no unmixed parallel systems are equivalent. We now show why this is true. Our strategy will be to exhibit the most general mapping theorem possible for an unmixed parallel and mixed serial system, and to use the implied constraints on the systems to discuss possibilities of identification. We will then look briefly at some special systems of interest.

Let the various possible processing orders for the set of elements $\{a, b, c\}$ (which we employ for convenience rather than the more cumbersome $\{X_a, X_b, X_c\}$) be denoted by

$$\begin{aligned} o_1 &= \langle a, b, c \rangle, & o_2 &= \langle a, c, b \rangle, & o_3 &= \langle b, a, c \rangle, \\ o_4 &= \langle b, c, a \rangle, & o_5 &= \langle c, a, b \rangle, & o_6 &= \langle c, b, a \rangle. \end{aligned}$$

Using ν to denote processing rate in a parallel system, we specify the parameters for stage 1.

Stage 1. (No elements processed.)

$$a: \nu_a^{(1)}, \quad b: \nu_b^{(1)}, \quad c: \nu_c^{(1)},$$

where the superscripts represent the stages.

At stage 2, the remaining two elements must have a constant (but not necessarily equal) value independently of which finishes next. Thus for stage 2 the processing orders, together with their associated parameters, are:

$$\begin{array}{lll} o_1 \text{ or } o_2 & b: & \nu_{bo_{1,2}}^{(2)}, \quad c: \nu_{co_{1,2}}^{(2)}, \\ o_3 \text{ or } o_4 & a: & \nu_{ao_{3,4}}^{(2)}, \quad c: \nu_{co_{3,4}}^{(2)}, \\ o_5 \text{ or } o_6 & a: & \nu_{ao_{5,6}}^{(2)}, \quad c: \nu_{co_{5,6}}^{(2)}. \end{array}$$

Finally, at stage 3, since each of the $(n-1)! = 2!$ orders is unique for each possible remaining element, the parameters may be unique:

$$\begin{array}{ll} o_1 & c: \nu_{co_1}^{(3)}, \\ o_2 & b: \nu_{bo_2}^{(3)}, \\ o_3 & c: \nu_{co_3}^{(3)}, \\ o_4 & a: \nu_{ao_4}^{(3)}, \\ o_5 & b: \nu_{bo_5}^{(3)}, \\ o_6 & a: \nu_{ao_6}^{(3)}. \end{array}$$

Proceeding in an analogous fashion for the serial system, we can first arbitrarily specify five probabilities for the six different processing orders and then a set of parameters:

$$\{\lambda_{ao_i}, \lambda_{bo_i}, \lambda_{co_i}\} \quad \text{for } i = 1, 2, \dots, 6,$$

which give the processing rates for the three elements of the six possible orders.

Let t_i = time from $i-1$ th element completion to the i th completion, and let q be a dummy variable for element a , r for b , and s for c . Then the characteristic function for the joint intercompletion times is, say, for order o_4 , given by the expectation:

$$\begin{aligned} & E[\exp(irt_1), \exp(ist_2), \exp(iqt_3); o_4] \\ &= \int_{t_1=0}^{\infty} \int_{t_2=0}^{\infty} \int_{t_3=0}^{\infty} [\nu_b^{(1)} \exp(-\nu_b^{(1)} t_1) \exp(iqt_1)] \\ & \quad \times [\exp(-\nu_c^{(1)} t_1) \nu_{co_{3,4}}^{(2)} \exp(-\nu_{co_{3,4}}^{(2)} t_2) \exp(irt_2)] \\ & \quad \times [\exp(-\nu_a^{(1)} t_1) \exp(-\nu_{ao_{3,4}}^{(2)} t_2) \nu_{ao_4}^{(3)} \exp(-\nu_{ao_4}^{(3)} t_3) \exp(ist_3)] dt_3 dt_2 dt_1 \\ &= \frac{\nu_b^{(1)}}{\nu_b^{(1)} + \nu_c^{(1)} + \nu_a^{(1)} - ir} \frac{\nu_{co_{3,4}}^{(2)}}{\nu_{co_{3,4}}^{(2)} + \nu_{ao_{3,4}}^{(2)} - is} \frac{\nu_{ao_4}^{(3)}}{\nu_{ao_4}^{(3)} - iq}. \end{aligned}$$

Similarly, for the serial system,

$$\begin{aligned} & E[\exp(irt_1), \exp(ist_2), \exp(iqt_3); o_4] \\ &= P(o_4) \int_{t_1=0}^{\infty} \int_{t_2=0}^{\infty} \int_{t_3=0}^{\infty} [\lambda_{ao_4} \exp(-\lambda_{ao_4} t_3) \exp(iqt_3)] \\ & \quad \times [\lambda_{bo_4} \exp(-\lambda_{bo_4} t_1) \exp(irt_1)] \\ & \quad \times [\lambda_{co_4} \exp(-\lambda_{co_4} t_2) \exp(ist_2)] dt_3 dt_2 dt_1 \\ &= P(o_4) \frac{\lambda_{bo_4}}{\lambda_{bo_4} - ir} \frac{\lambda_{co_4}}{\lambda_{co_4} - is} \frac{\lambda_{ao_4}}{\lambda_{ao_4} - iq}. \end{aligned}$$

The (most general possible) appropriate mapping for this characteristic function and a paradigm for the others, which yields non-identifiability, is:

$$P(o_4) = \frac{\nu_b^{(1)}}{\nu_b^{(1)} + \nu_c^{(1)} + \nu_a^{(1)}} \frac{\nu_{co_{3,4}}^{(2)}}{\nu_{co_{3,4}}^{(2)} + \nu_{ao_{3,4}}^{(2)}} \frac{\nu_{ao_4}^{(3)}}{\nu_{ao_4}^{(3)}}, \quad (1)$$

$$\lambda_{bo_4} = \nu_b^{(1)} + \nu_c^{(1)} + \nu_a^{(1)}, \quad \lambda_{co_4} = \nu_{co_{3,4}}^{(2)} + \nu_{ao_{3,4}}^{(2)}, \quad \lambda_{ao_4} = \nu_{ao_4}^{(3)}. \quad (2)$$

The expression for $P(o_4)$ can be interpreted as a product of conditional probabilities $P(b) \cdot P(c|b) \cdot P(a|b, c)$ in the order given by o_4 .

Note, however, that $\nu_b^{(1)} + \nu_c^{(1)} + \nu_a^{(1)} = \text{constant}$, since $\nu_b^{(1)}, \nu_c^{(1)}, \nu_a^{(1)}$, being rates for a single unmixed parallel system, are the same from trial to trial, and that λ for the first element processed must equal this sum, i.e.

$$\lambda_{ao_1} = \lambda_{ao_2} = \lambda_{bo_3} = \lambda_{bo_4} = \lambda_{co_5} = \lambda_{co_6},$$

thus reducing the freedom of the system to exhibit a variety of first-completion times. Similarly, at the second stage, owing again to constraints on the parallel system, the serial parameters are confined to the values:

$$\lambda_{bo_1} = \lambda_{co_2}, \quad \lambda_{ao_3} = \lambda_{co_4}, \quad \lambda_{ao_5} = \lambda_{bo_6}.$$

At stage 3, there is complete freedom of parameter values for both parallel and serial systems. That is, we can set

$$\nu_{co_1}^{(3)} = \lambda_{co_1}, \quad \nu_{bo_2}^{(3)} = \lambda_{bo_2}, \quad \nu_{co_3}^{(3)} = \lambda_{co_3}, \quad \nu_{ao_4}^{(3)} = \lambda_{ao_4}, \quad \nu_{bo_5}^{(3)} = \lambda_{bo_5}, \quad \nu_{ao_6}^{(3)} = \lambda_{ao_6}.$$

This reduces the number of serial free parameters to that of the parallel system and renders them equivalent. It may readily be ascertained that $\sum_{i=1}^6 P(o_i) = 1$, when written in terms of the above mapping. A converse function may similarly be provided from the special, reduced, serial system to the parallel system. This is the most general mimicking that can take place between unmixed parallel and mixed serial systems for the present class of distributions. All other non-identifiability mappings will be subcases of those presented above.

The above limitations of the serial system are forced by the assumptions that make an unmixed parallel system act in the same way from trial to trial when the system is in the same internal state.

We see, then, that an unmixed parallel system is incapable of mimicking the most general mixed serial systems. If an experimenter wishes to test between

these two types of processes, and he is able to measure, for instance, minimum completion time for the elements, then he should employ stimuli that encourage different processing rates for the different elements. The distribution time to first completion averaging over order will be exponential for an unmixed parallel system but a probability mixture of exponentials for a mixed serial system. A similar point was made by Thomas (1969) in a discussion of non-parametric tests for three alternative models of processing.

The natural consequent question of whether a mixed parallel system can be equivalent to a mixed serial system can be answered in the negative. We demonstrate this for $n = 2$. Let ν_{a_i}, ν_{b_i} ($i = 1, 2$) and $\nu_{a_i'}, \nu_{b_i}'$ be the two sets (for each of two systems, say I and II) of parallel rate parameters for elements a, b where i refers to stage of processing, and let q be the probability of selecting system I, $1 - q$ the probability of selecting system II. Similarly, let $\lambda_{a_i}, \lambda_{b_i}$ ($i = 1, 2$) and p be the parameters for the mixed serial system. Then

$$E[\exp(irt_1), \exp(ist_2); \langle a, b \rangle] = p \frac{\lambda_{a_1}}{\lambda_{a_1} - ir} \frac{\lambda_{b_2}}{\lambda_{b_2} - is}$$

for the mixed serial system and

$$E[\exp(irt_1), \exp(ist_2); \langle a, b \rangle] = q \frac{\nu_{a_1}}{\nu_{a_1} + \nu_{b_1} - ir} \frac{\nu_{b_2}}{\nu_{b_2} - is} \\ + (1 - q) \frac{\nu_{a_1}'}{\nu_{a_1}' + \nu_{b_1}' - ir} \frac{\nu_{b_2}'}{\nu_{b_2}' - is}$$

for the mixed parallel system. The characteristic functions for the order $\langle b, a \rangle$ are found in a similar fashion. It follows from setting these two expressions equal to one another (and similarly for $\langle b, a \rangle$), and determining the parameter constraints for equivalence, that

$$\nu_{a_1} + \nu_{b_1} = \nu_{a_1}' + \nu_{b_1}' = \lambda_{a_1} = \lambda_{b_1} \quad \text{and} \quad \nu_{a_2} = \nu_{a_2}' = \lambda_{a_2}, \quad \nu_{b_2} = \nu_{b_2}' = \lambda_{b_2}.$$

That is, the non-identifiable models are subclasses of their respective class.

Intuitively, if we conditionalize on order, for example $\langle a, b \rangle$, then the minimum processing time for the mixed serial system will be exponentially distributed but the same quantity for the mixed (non-degenerate) parallel system is a probability mixture of exponentially distributed random variables with different processing rates. This provides one possible technique for testing between mixed serial and mixed parallel (exponential) systems.

Thus the class of unmixed parallel systems is less general and can be mimicked by special cases of the class of mixed serial systems. On the other hand, the class of mixed parallel systems, though being in a sense more complex than the class of mixed serial systems, does not include the latter class and hence cannot mimic it.

Another important point is that if only one order is ever taken by a serial system (i.e. $P(o_i) = 1$ for some i), then the parallel system is unable to mimic it perfectly. Although by appropriate adjustment of the parameters the parallel system can arbitrarily closely approximate such a serial system (since the

characteristic function is a continuous function of the rate parameters at the appropriate points), it is in fact acting very much like a serial system. This is because in such a case all the processing rates, except for the one associated with the element processed in a given position, must be near to zero at the appropriate stages for the approximation to hold. Hence a second distinct strategy would be to use instructions or materials conducive to processing elements in a fixed order (serial system) or, equivalently, ordering the magnitude of processing rates, so that emphasis is given to a particular element at each stage (parallel system). Good fits of such models would indicate that the system is capable of acting in a serial manner. However, as will be shown below, at the level of the first moment, even otherwise identifiable models can predict identical results under certain conditions. Therefore successful application of these conclusions depends on being able to extract higher level information about the processing distribution and/or designing experiments in a manner that will specifically exploit the present results.

4. MIXED-ORDER SERIAL SYSTEMS WITH A SINGLE PROCESSING RATE AND EQUIVALENT PARALLEL SYSTEMS

Using the structure presented above, we may consider some special cases of interest. The first is derived by making all the λ 's equal in a serial system for the different elements, but allowing an arbitrary distribution of probabilities over the possible processing orders. This results in a system with $n!$ parameters. Such a system might be expected when all the elements to be processed are homogeneous with respect to the system (i.e. equally difficult to process), but some orders of processing are necessary or preferred to others.

When just one element is presented to the system the mean completion time is $1/\lambda_1$ and when two must be processed the overall mean completion time is $2/\lambda_2$ when processing is exhaustive, and $3/2\lambda_2$ when processing is self-terminating (i.e. when only one of the two elements presented must be processed and Probability(a must be processed) = $\frac{1}{2}$). The mean completion times for elements a, b , respectively, are

$$E(t_a) = p \frac{1}{\lambda_2} + (1-p) \frac{2}{\lambda_2} \quad \text{and} \quad E(t_b) = p \frac{2}{\lambda_2} + (1-p) \frac{1}{\lambda_2},$$

when two elements are presented. When $\lambda_1 = \lambda_2 = \dots = \lambda_n$ (i.e. the number of elements to be processed is $1, 2, \dots, n$, respectively, and λ is the same in all cases), the system produces the usual 'serial' results of linear overall mean completion times, independent of the probability distribution over processing orders.

Since the number of parameters for a homogeneous parallel system is $n! \sum_{i=0}^{n-1} 1/i! > n!$, we expect that constraints must be imposed for parallel-to-serial mimicking to occur here. The general processing rates for $n = 2$ in the parallel system are $\nu_{a_1}, \nu_{b_1}, \nu_{a_2}, \nu_{b_2}$ for elements a, b in stages 1 and 2, respectively. From the previous section, we obtain that $\nu_{a_1} = p\lambda_2, \nu_{b_1} = (1-p)\lambda_2, \nu_{a_2} = \nu_{b_2} = \nu_2 = \lambda_2$ in order for non-identifiability to hold. Notice that $\nu_{a_1} + \nu_{b_1} = \nu_2 = \lambda_2$ for $n = 2$ and, in general, the sum of the ν 's at each stage in the parallel system must equal

λ_n . This means that when an element is completed, the parallel system allocates the attention or energy that was previously devoted to that element to the remaining unprocessed elements. If, in addition, $\lambda_n = \lambda$ for $n = 1, 2, \dots$, then the parallel system must be of limited capacity; in this case, the total amount of processing energy† is constant and must be spread over the available elements that require processing (see Gardner, 1970, for a general qualitative review of unlimited vs. limited capacity systems and related notions). For example, if $\lambda = 4$, then $\nu_{a_1} + \nu_{b_1} = 4$ and $\nu_{a_2} = \nu_{b_2} = 4$ when two elements are presented, and $\nu_a = \nu_b = 4$ when only one element is presented. In the serial system, the total processing energy, denoted by $\lambda = 4$, is applied entirely to one element at a time. In the mimicking parallel system the total processing energy, given by the sum of the ν 's, is spread out over all elements that must be processed. Non-uniformities in dependent variables related to completion times for individual elements, as reaction time to detect a test letter for different positions in an array, are handled by the probabilities for different processing orders in serial systems (e.g. $p \neq \frac{1}{2}$ for $n = 2$) and by unequal ν 's (e.g. $\nu_{a_1} \neq \nu_{b_1}$) in parallel systems.

Since the parallel and serial processes stated immediately above are equivalent, any statistic of interest will be identical for the two. In particular, given some fixed time for processing, T , the conditional probability of processing a given b completes processing may be computed. If a and b are the only two elements to be processed, this conditional probability can be written

$$P_T(a|b) = \frac{1 - \exp(-\lambda T) - \lambda T \exp(-\lambda T)}{p[1 - \exp(-\lambda T) - \lambda T \exp(-\lambda T)] + (1-p)[1 - \exp(-\lambda T)]},$$

where the variables are as defined previously, and this can be compared with the marginal probability that a completes processing,

$$P_T(a) = p[1 - \exp(-\lambda T)] + (1-p)[1 - \exp(-\lambda T) - \lambda T \exp(-\lambda T)].$$

With a development that is tedious but straightforward, it can be shown that $P_T(a|b) > P_T(a)$ for all $0 \leq p \leq 1$ and for positive $T < +\infty$ and that they are equal for $T = 0$ and $T \rightarrow +\infty$. The mimicking parallel model will also have this property and both may be contrasted with models that predict independence or that $P_T(a|b) < P_T(a)$. The latter, for example, is predicted by fixed sample size models (Estes & Taylor, 1964).

The above paragraphs referred to complete non-identifiability. In addition, a strictly independent parallel process can be used to mimic the means of the present serial systems even though they refer to distinct underlying probability distributions. As an example, consider an experiment in which processing need occur on only one element out of the total n on at least some proportion of the trials. Some of the experiments involving search for a 'test' element discussed earlier (e.g. Estes & Taylor, 1964; Atkinson *et al.*, 1969) include such a task. In such cases it is of interest to ask whether processing is self-terminating (stops

† We employ 'processing energy' as equivalent to 'processing capacity' and the former is not meant to connote specific neurobiological processes.

when the test element is processed) or exhaustive (stops only when all the elements have been processed). High accuracy implies that the test (often called 'critical' or 'target') element has been processed, and we can compare the mean reaction times associated with different test elements or different positions in a presentation. Differences in these means are usually taken to reflect self-terminating processing.

In a two-element task, with the elements designated as a , b , we can simply make

$$\nu_a = \left[p \frac{1}{\lambda_2} + (1-p) \frac{2}{\lambda_2} \right]^{-1}, \quad \nu_b = \left[p \frac{2}{\lambda_2} + (1-p) \frac{1}{\lambda_2} \right]^{-1},$$

or, mapping in the other direction,

$$\lambda_2 = \frac{3\nu_a\nu_b}{\nu_a + \nu_b}, \quad p = \frac{2\nu_a - \nu_b}{\nu_a + \nu_b}.$$

This yields identical mean processing times for a , b as well as for the overall self-terminating mean, $E_{st}(t) = \frac{1}{2}E(t_a) + \frac{1}{2}E(t_b)$ (when the test element is a 50 per cent of the time and b the remainder of the time) in the independent parallel and mixed serial process. However, once the constraint is imposed that the self-terminating means are equal on the set of trials when a test element was present, the predicted means for trials when no test element is present (and hence when processing must be exhaustive) will be different for the parallel independent and serial models. For instance, when $n = 2$ the exhaustive mean for the serial model is

$$E_{ex}(t) = 2/\lambda_2$$

and for the parallel independent model that mimics the serial model on the self-terminating means the exhaustive mean is

$$E_{ex}(t) = \frac{7-p+p^2}{3\lambda_2} > \frac{2}{\lambda_2}.$$

Unfortunately, it appears that, in at least some tasks, the parameters even within the same model cannot be the same on test-element-present trials as on test-element-absent trials (see Bamber, 1969; Townsend & Roos, in preparation), thus vitiating this particular distinguishing characteristic.

Nevertheless, within the same type of trials, say test-element-present, when the parallel and serial models are made to be equal by fitting $E(t_a)$ and $E(t_b)$, then they will not predict equal variances. Little experimental work appears to have been done employing anything beyond mean reaction times. Also, although the parallel independent parameters can fit any $E(t_a)$, $E(t_b)$ predicted by the serial model, the serial model cannot predict all values for these means that derive from the parallel independent model. For example, when $\nu_b = 2\nu_a$, unacceptable values for p are obtained in the serial model. This point will be elaborated in a more general discussion below. Finally, when the serial model predicts linear mean processing times as a function of n (as noted above,

$\lambda_1 = \lambda_2 = \dots = \lambda_n$), this can be predicted by a parallel independent model by carrying out the type of mapping described above for each n , as can the self-terminating mean processing times (with linear visual arrays, often referred to as serial position effects) as well.

Alternatively, we can provide the mapping sufficient to make the exhaustive means equivalent. For instance, since

$$E_{\text{ex}}(t) = \frac{1}{v_n} \sum_{j=1}^n \left(\frac{1}{j} \right)$$

for the parallel model with homogeneous rate parameters for the various elements or element positions, for this mean to be a linear function of n we require that

$$\frac{1}{v_n} \sum_{j=1}^n \left(\frac{1}{j} \right) = kn + a,$$

with k and a (typically) positive constants. Hence

$$v_n = \frac{1}{kn + a} \sum_{j=1}^n \left(\frac{1}{j} \right)$$

will ensure this equivalence, and since the last expression increases approximately as

$$v_n \simeq \frac{1}{kn + a} \log n,$$

it can be seen that the overall processing energy (which we take as the sum of the rates with which each element is processed), nv_n , increases approximately as $\log n$. The constant $-\lambda$ serial model's total processing energy is, of course, linear in n , $n\lambda$. This result is in contradistinction to the apparent claim by Sternberg (1966) that parallel independent models are incapable of such predictions, at least for his data. It is true that an unlimited capacity parallel independent model could not predict straight lines (e.g. $v_n = v$ and therefore $E_{\text{ex}}(t)$ is approximately logarithmic), but, as has been shown here, a limited capacity parallel independent model can. The overall rate of increase in processing time, and hence in reaction time, is thus seen to depend primarily on the available supply of processing energy rather than on whether processing is independent or not, or whether processing is serial or not.

This brings us to our next special case with complete non-identifiability, that of independent parallel processing.

5. PARALLEL SYSTEMS WITH CONSTANT PROCESSING RATES AND EQUIVALENT SERIAL SYSTEMS

In this type of parallel system we allow different processing rates for the different elements but assume that they are constant throughout processing, i.e. they are not allowed to change with stage. A 'typical' n -parameter parallel system with independent processing times for the n elements is the result.

For example, for $n = 2$, $v_{a_1} = v_{a_2} = v_a$ and $v_{b_1} = v_{b_2} = v_b$, but it is not necessary that $v_a = v_b$. The mean completion times for elements a , b , respectively, are

$1/\nu_a$ and $1/\nu_b$. The expression for the overall mean processing time for both elements is not as obvious as it was in the serial case. However, since the class of probability distributions for mixed serial systems includes those for unmixed parallel systems, we can write the mean completion time expression for the serial case and then use our mapping theorem in order to obtain the expression for the parallel case. The mean (exhaustive) completion time for the two-element serial system is

$$E_{\text{ex}}(t) = p \left[\frac{1}{\lambda_a^{(1)}} + \frac{1}{\lambda_b^{(2)}} \right] + (1-p) \left[\frac{1}{\lambda_b^{(1)}} + \frac{1}{\lambda_a^{(2)}} \right],$$

where $\lambda_a^{(i)}$, $\lambda_b^{(i)}$ give the processing rates for elements a , b when they are processed i th ($i = 1, 2$) and p is the probability of order $\langle a, b \rangle$. The earlier mapping theorem when used with the above constraints on the parallel system's processing rates (the ν 's) yields

$$E_{\text{ex}}(t) = \frac{\nu_a}{\nu_a + \nu_b} \left[\frac{1}{\nu_a + \nu_b} + \frac{1}{\nu_b} \right] + \frac{\nu_b}{\nu_a + \nu_b} \left[\frac{1}{\nu_a + \nu_b} + \frac{1}{\nu_a} \right]$$

as the independent parallel system's exhaustive mean processing time. The mean processing times for the self-terminating parallel independent case are similarly derived:

$$E_{\text{st}}(t) = \frac{1}{2}E(t_a) + \frac{1}{2}E(t_b),$$

$$E(t_a) = \frac{\nu_a}{\nu_a + \nu_b} \left[\frac{1}{\nu_a + \nu_b} \right] + \frac{\nu_b}{\nu_a + \nu_b} \left[\frac{1}{\nu_a + \nu_b} + \frac{1}{\nu_a} \right] = \frac{1}{\nu_a},$$

$$E(t_b) = \frac{\nu_a}{\nu_a + \nu_b} \left[\frac{1}{\nu_a + \nu_b} + \frac{1}{\nu_b} \right] + \frac{\nu_b}{\nu_a + \nu_b} \left[\frac{1}{\nu_a + \nu_b} \right] = \frac{1}{\nu_b}.$$

This illustrates one of the uses to which one may put the present identifiability results: formulas that are complicated in one type of system can often be derived from simpler formulas in the other type. The values of the serial parameters employed for the overall mean are also those that produce the serial process that mimics the parallel, namely,

$$\lambda_a^{(1)} = \lambda_b^{(1)} = \lambda^{(1)} = \nu_a + \nu_b, \quad \lambda_a^{(2)} = \nu_a, \quad \lambda_b^{(2)} = \nu_b \quad \text{and} \quad p = \frac{\nu_a}{\nu_a + \nu_b}.$$

As a simple numerical example, assume that a is processed three times as fast as b , i.e. $\nu_a = 3\nu_b = 3\nu$ and that $\nu = 1$. Hence the mean completion times for a , b are $\frac{1}{3}$ and 1, respectively, and the total processing energy is $\nu_a + \nu_b = 4$.

The mimicking serial system is generated by making $p = \frac{3}{4}$ (i.e. a is processed first with probability 0.75),

$$\lambda_a^{(1)} = \lambda_b^{(1)} = \lambda^{(1)} = 4, \quad \lambda_a^{(2)} = \nu_a = 3 \quad \text{and} \quad \lambda_b^{(2)} = \nu_b = 1.$$

This example illustrates two properties characterizing serial systems that mimic the present independent parallel systems. One is that p reflects the relative magnitudes of ν_a , ν_b and the other is that there is a general diminution in $\lambda^{(i)}$

as i increases. Since the sum of the ν 's at each stage gives the total processing energy available until the next element is completed, and in the present system the individual ν 's are constant across stages, the result is a lengthening of inter-completion times as processing takes place. The decreasing λ 's in the mimicking serial model reflect this slowing down. As is expected, this serial model predicts independent completion times; for example, when processing time T is fixed and exactly two elements are available for processing, $P_T(a|b) = P_T(a)$.

6. A FUNCTIONAL EQUATION TO TEST PARALLEL-SERIAL IDENTIFIABILITY

The probability distribution function for a two-element system leads to a functional equation that may be valuable in testing for parallel-serial mimicking and may help define the classes of distributions that produce non-identifiability. The serial systems are mixed but the parallel systems are unmixed.

Let $t_{a_i}(t_{b_i})$ be the time taken to process a (b) if it is processed in stage i ; $\Gamma_{a_i}(\Gamma_{b_i})$ be the set of parameters for the serial distribution function during the i th processing stage for elements a (b); $\Lambda_{a_i}(\Lambda_{b_i})$ be the set of parameters for the parallel distribution function during stage i for elements a (b); $F(\alpha, \beta)$ is the general distribution function of both systems where α gives the time variable and β the set of parameters; $f(\alpha, \beta)$ is the general density function for the parallel and serial systems (if it exists); and p is the (serial) probability that a is processed in the first stage (i.e. first). Since we will probably be first interested in testing identifiability within particular classes of distributions, we have made the parallel and serial distribution functions the same (except for parameter specification).

It is assumed that the intercompletion times for different elements are independent. That is, after each completion, processing of the remaining elements starts from zero with distribution parameters independent of previous completion times but not necessarily of the elements completed or their order. Also, we will consider only the processing order $\langle a, b \rangle$. The developments for $\langle b, a \rangle$ are similar in all respects and the parameter mappings that yield equivalence for $\langle a, b \rangle$ also work for $\langle b, a \rangle$. The joint distribution function on the waiting times for parallel systems where (not 'given') a is processed first is then given by

$$P^{(p)}(t_{a_1} \leq t_a; t_{b_2} \leq t_b) = \int_{t_{a_1}=0}^{t_a} [1 - F(t_{a_1}; \Lambda_{b_1})] F(t_b - t_{a_1}; \Lambda_{b_2}) dF(t_{a_1}; \Lambda_{a_1}).$$

Similarly, the analogous joint distribution function for the serial process can be written

$$P^{(s)}(t_{a_1} \leq t_a; t_{b_2} \leq t_b) = p \int_{t_{a_1}=0}^{t_a} F(t_b - t_{a_1}; \Gamma_{b_2}) dF(t_{a_1}; \Gamma_{a_1}).$$

We now apply the principle that at the last stage, given a particular specified order of elements processed, the distribution functions on the last element can be made identical for the parallel and serial models. This principle was seen earlier, for instance, in the setting $\nu_{a_2} = \lambda_{a_2}$ and $\nu_{b_2} = \lambda_{b_2}$ in two-element systems

with exponential densities. Hence $\Gamma_{b_2} \equiv \Lambda_{b_2}$, and the last terms under the integrals in the parallel and serial expression become equal. Next, by this result and the Radon-Nikodym theorem (Halmos, 1950),

$$P^{(s)}(t_{a_1} \leq t_a; t_{b_2} \leq t_b) \equiv P^{(p)}(t_{a_1} \leq t_a; t_{b_2} \leq t_b)$$

if and only if

$$p dF(t_{a_1}; \Gamma_{a_1}) = [1 - F(t_{a_1}; \Lambda_{b_1})] dF(t_{a_1}; \Lambda_{a_1})$$

for all (Lebesgue) measurable sets on the real line, except those of measure zero. We will investigate the stronger condition that the above quality holds for all t_{a_1} . We now consider several special cases in terms of this condition. Since the following cases are Riemann integrable, we employ the density functions $dF(t_{a_1}; \Gamma_{a_1}) = f(t_{a_1}; \Gamma_{a_1}) dt_{a_1}$, and $dF(t_{a_1}; \Lambda_{a_1}) = f(t_{a_1}; \Lambda_{a_1}) dt_{a_1}$.

First, the exponential processes are used to illustrate a positive case. We ask if there exist parallel-to-serial and serial-to-parallel mappings for all $t_{a_1} \geq 0$ that give $p\lambda_{a_1} \exp(-\lambda_{a_1} t_{a_1}) = \nu_{a_1} \exp(-\nu_{a_1} t_{a_1}) \exp(-\nu_{b_1} t_{a_1})$.

Clearly, our former mapping,

$$p = \frac{\nu_{a_1}}{\nu_{a_1} + \nu_{b_1}}, \quad \lambda_{a_1} = \nu_{a_1} + \nu_{b_1},$$

yields

$$\frac{\nu_{a_1}}{\nu_{a_1} + \nu_{b_1}} (\nu_{a_1} + \nu_{b_1}) \exp[-(\nu_{a_1} + \nu_{b_1}) t_{a_1}] = \nu_{a_1} \exp(-\nu_{a_1} t_{a_1}) \exp(-\nu_{b_1} t_{a_1})$$

and the equivalence is established.

Secondly, consider the system with intercompletion times that are uniformly distributed subject to the previously imposed constraints. Specifically, let

$$f(t_{a_1}; \Gamma_{a_1}) = \begin{cases} \frac{1}{h} & 0 \leq t_{a_1} \leq h, \Gamma_{a_1} = h, \\ 0 & \text{otherwise,} \end{cases}$$

$$f(t_{a_1}; \Lambda_{a_1}) = \begin{cases} \frac{1}{k} & 0 \leq t_{a_1} \leq k, \Lambda_{a_1} = k, \\ 0 & \text{otherwise} \end{cases}$$

and

$$1 - F(t_{a_1}; \Lambda_{b_1}) = \frac{l - t_{a_1}}{l}.$$

Then the functional equation requires that

$$\frac{p}{h} = \frac{1}{k} \frac{l - t_{a_1}}{l}, \quad \Lambda_{b_1} = l.$$

Thus it is immediately evident that

$$p = \frac{h}{k} \frac{l - t_{a_1}}{l}$$

for all $t_{a_1} \geq 0$, and since h, k, l are constant, p must be a function of t_{a_1} . This violates the rule of parameter specification for the serial model; also the uniform distribution cannot produce non-identifiability under the present assumptions.

The Weibull distribution (Cox, 1962) is interesting to investigate since it can be obtained from the exponential distribution by a power transformation of the time scale. The Weibull density function is given by

$$f(x) = \alpha \rho (\rho x)^{\alpha-1} \exp[-(\rho x)^\alpha],$$

and therefore forms a two-parameter distribution. Let

$$\Gamma_{a_1} = (\alpha, \rho), \quad \Lambda_{a_1} = (\beta, \tau), \quad \Lambda_{b_1} = (\gamma, \omega) \quad \text{and} \quad t_{a_1} = t.$$

Then the functional equation assumes that

$$p \frac{\alpha \rho (\rho t)^{\alpha-1} \exp[-(\rho t)^\alpha]}{\beta \tau (\tau t)^{\beta-1} \exp[-(\tau t)^\beta]} = \int_{t'=t}^{+\infty} \gamma \omega (\omega t')^{\gamma-1} \exp[-(\omega t')^\gamma] dt'.$$

A change of variable in the right-hand expression, $(t')^\gamma = r$, leads to the term

$$\int_{r'}^{+\infty} \omega^\gamma \exp(-\omega^\gamma r) dr = \exp[-(\omega t)^\gamma].$$

To produce the same expression for the left-hand side, let $\gamma = \alpha = \beta$ and

$$p = \frac{\tau^\alpha}{\tau^\alpha + \omega^\alpha}, \quad \rho^\alpha = \tau^\alpha + \omega^\alpha.$$

This again gives $\exp[-(\omega t)^\gamma]$. Hence it is possible to find parallel and serial processes having the Weibull distribution that mimic one another. However, this is at the expense of using the same exponent in the power transformation on t . An open problem is determination of the total class of distributions that are capable of satisfying the above functional equations.

7. A CASE OF HYBRID MIMICKING

As a generalization of parallel processes with exponential distributions consider a two-element system based on a two-step gamma distribution. That is, each element must complete two steps in processing, and the completion time of these two steps is governed by a gamma distribution for each element. It is assumed that the rate parameter can depend on stage but not on the steps within a stage. Thus ν_{a_1}, ν_{b_1} are the parameters within stage one but when, say, both steps of a are completed the parameter for b can become ν_{b_2} . The analysis of this distribution is conveniently performed in terms of the density functions on inter-completion times. The variables $t_1^{(a)}, t_1^{(b)}$ are the times to completion of a, b where a or b is completed first and $t_2^{(a)}, t_2^{(b)}$ are the times from completion of b or a , respectively, to the completion of a or b .

The joint density function on these is given by

$$f^{(p)}(t_1^{(a)}, t_2^{(b)}; \langle a, b \rangle) = \nu_{a_1}^2 t_1^{(a)} \exp [-(\nu_{a_1} + \nu_{b_1}) t_1^{(a)}] \nu_{b_2}^2 t_2^{(b)} \exp [-\nu_{b_2} t_2^{(b)}] \\ + \nu_{b_1} \nu_{a_1}^2 (t_1^{(a)})^2 \exp [-(\nu_{a_1} + \nu_{b_1}) t_1^{(a)}] \nu_{b_2} \exp [-\nu_{b_2} t_2^{(b)}],$$

and

$$f^{(p)}(t_1^{(b)}, t_2^{(a)}; \langle b, a \rangle) = \nu_{b_1}^2 t_1^{(b)} \exp [-(\nu_{a_1} + \nu_{b_1}) t_1^{(b)}] \nu_{a_2}^2 t_2^{(a)} \exp [-\nu_{a_2} t_2^{(a)}] \\ + \nu_{a_1} \nu_{b_1}^2 (t_1^{(b)})^2 \exp [-(\nu_{a_1} + \nu_{b_1}) t_1^{(b)}] \nu_{a_2} \exp [-\nu_{a_2} t_2^{(a)}],$$

where (p) stands for 'parallel'.

The expression $f(t_1^{(a)}, t_2^{(b)}; \langle a, b \rangle)$ is made up of the sum of two terms; the first gives the probability density when both stages of a finish before even one stage of b is completed. Thus

$$\nu_{a_1}^2 t_1^{(a)} \exp [-\nu_{a_1} t_1^{(a)}]$$

gives a two-step density function for element a , $\exp [-\nu_{b_1} t_1^{(a)}]$ is the probability that step one of the element b 's processing was not completed during time $t_1^{(a)}$, and

$$\nu_{b_2}^2 t_2^{(b)} \exp [-\nu_{b_2} t_2^{(b)}]$$

is a two-step gamma density for b during stage 2. The second term in the sum is composed of

$$[\nu_{a_1}^{(2)} t_1^{(a)} \exp (-\nu_{a_1} t_1^{(a)})] [\nu_{b_1} t_1^{(a)} \exp (-\nu_{b_1} t_1^{(a)})] [\nu_{b_2} \exp (-\nu_{b_2} t_2^{(b)})],$$

where the first term in the product is a two-step gamma as before, the second gives the Poisson probability that exactly one step was completed on b during stage one, and the third is a one-step gamma, or exponential, for the second step of processing b . The joint density for the other order, $\langle b, a \rangle$, can be derived in a similar manner.

A natural serial distribution to consider as possibly equivalent to this parallel distribution is a mixed path two-step gamma distribution given by

$$f^{(s)}(t_1^{(a)}, t_2^{(b)}; \langle a, b \rangle) = p[\lambda_{a_1}^2 t_1^{(a)} \exp (-\lambda_{a_1} t_1^{(a)}) \lambda_{b_2}^2 t_2^{(b)} \exp (-\lambda_{b_2} t_2^{(b)})],$$

and

$$f^{(s)}(t_1^{(b)}, t_2^{(a)}; \langle b, a \rangle) = (1-p) [\lambda_{b_1}^2 t_1^{(b)} \exp (-\lambda_{b_1} t_1^{(b)}) \lambda_{a_2}^2 t_2^{(a)} \exp (-\lambda_{a_2} t_2^{(a)})],$$

where (s) stands for 'serial'. However, it is evident from consideration of these expressions that they are unable to produce the complexity found in the present parallel system. Basically, a product of two, two-step gamma densities cannot in general be converted to a sum of products of one- and two-stage gamma densities. (The reader may, as an exercise, verify this with the earlier functional equation.)

Nevertheless, it is possible to mimic the two-step, two-element parallel system by a hybrid time-sharing system that is engaged in processing only one element at a time but does not necessarily complete one before going on to the other. It is necessary for this analysis to break down processing into steps completed so

cases where it is reasonable to assume reallocation of energy previously devoted to elements now completed, the hybrid model may be just as intuitive since it merely presumes that the processor skips or ignores completed elements.

8. INCOMPLETE MIMICKING: IDENTIFIABILITY AT DIFFERENT LEVELS

We turn our attention to cases that involve only partial non-identifiability, as when mimicking occurs at the level of the means, for example, but not for all aspects (e.g. all moments) of the distributions. A special case was noted earlier in the exponential systems.

The investigation of means only, of course, is equivalent to studying only deterministic systems. But if mimicking is not complete, relations may be discovered between parameters and statistics of one process and the parameters and/or statistics of the other.

As a simple example, consider a mixed path serial process with the same processing speed for different elements and an independent parallel system with possibly different speeds for different elements. Assume also that both processes are based on one-parameter distributions. In the case when two elements must be processed, the means (expectations) of the processing times for the two elements are $E(t_a) = pg(r) + 2(1-p)g(r)$ and $E(t_b) = (1-p)g(r) + 2pg(r)$, where we assume that $g(r)$, the expected time to process a single element, is a positive, monotonically decreasing (continuous for $r > 0$ but unbounded as $r \rightarrow 0$) function of the rate parameter, $r > 0$, and p is as before. Solving these two equations for p and $g(r)$ gives

$$p = \frac{2E(t_b) - E(t_a)}{E(t_a) + E(t_b)} \quad \text{and} \quad g(r) = \frac{E(t_a) + E(t_b)}{3}.$$

Next, the expectations of the parallel system are just $E(t_a) = g(s_a)$ and $E(t_b) = g(s_b)$, where s_a and s_b are the two rate parameters for a and b and the use of g implies that we are staying within the same family of probability distributions for the parallel and serial systems. As long as our means are positive, this parallel system puts no constraint on g or, for that matter, s_a and s_b . The present serial system puts no constraint on g but does on p : if $E(t_a) > 2E(t_b)$ or $E(t_b) > 2E(t_a)$, then p attains inadmissible values and hence this serial system is falsifiable even at the level of the means, whereas the parallel system is not. On the other hand, if the serial system is verified (at the level of means), so is the parallel system. Another way of viewing this result, which also extends to other uses, is to solve for the serial parameters, assuming mimicking of the means, in terms of the parallel parameters and vice versa:

$$r = g^{-1}\left(\frac{g(s_a) + g(s_b)}{3}\right), \quad p = \frac{2g(s_b) - g(s_a)}{g(s_a) - g(s_b)},$$

$$s_a = g^{-1}[g(r)(2-p)], \quad s_b = g^{-1}[g(r)(1+p)].$$

Since g is a monotonic function of its argument, its inverse, g^{-1} , is well defined. Clearly, within the acceptable range of parameter values for the serial system, the rate parameters for the parallel system are not allowed to attain their full

range of values. One consequence is that the difference between s_a and s_b , $|s_a - s_b|$, is limited.

These relations can be employed to investigate differences in predictions at higher levels. For instance, under the parameter mappings given above, the variances for the parallel and serial systems can be written as functions of the parameters of the other systems and therefore possible constraints on predictions derived. Thus if $h(r)$ is the second raw moment of processing time for a single element, then the serial element variances are

$$\begin{aligned}\text{var}^{(s)}(t_a) &= (2-p)h(r) + [2(1-p) - (2-p)^2]g^2(r), \\ \text{var}^{(s)}(t_b) &= (1+p)h(r) + [2p - (1+p)^2]g^2(r).\end{aligned}$$

These equations can now be expressed as functions of s_a and s_b and the parallel variances for the two elements,

$$\text{var}^{(p)}(t_a) = h(s_a) - g^2(s_a), \quad \text{var}^{(p)}(t_b) = h(s_b) - g^2(s_b),$$

can be expressed in terms of p, r . Hence we find by substitution that

$$\begin{aligned}\text{var}^{(p)}(t_a) &= h\{g^{-1}[g(r)(2-p)]\} - g^2\{g^{-1}[g(r)(2-p)]\} \\ &= h\{g^{-1}[g(r)(2-p)]\} - g^2(r)(2-p)^2\end{aligned}$$

and similarly for $\text{var}^{(p)}(t_b)$. Note that for $p = 1$, $\text{var}^{(p)}(t_a) = h(r) - g^2(r)$, as would be expected. Since all parameters are functions of the element means, the variances could also be written as functions of the expectations and expectation versus variance comparisons obtained.

We can apply these developments to our exponential systems. Under the present constraints, $\nu_{a_1} = \nu_{a_2} = \nu_a$, $\nu_{b_1} = \nu_{b_2} = \nu_b$, $\lambda_{a_1} = \lambda_{a_2} = \lambda_{b_1} = \lambda_{b_2} = \lambda$, and p is the remaining parameter.

Let $r = \lambda$, $s_a = \nu_a$, $s_b = \nu_b$, $g(\lambda) = 1/\lambda$, $h(\lambda) = 2/\lambda^2$, then by this last result above,

$$\begin{aligned}\text{var}^{(p)}(t_a) &= h\left[g^{-1}\left(\frac{2-p}{\lambda}\right)\right] - \frac{(2-p)^2}{\lambda^2} \\ &= \frac{2(2-p)^2}{\lambda^2} - \frac{(2-p)^2}{\lambda^2} = \left(\frac{2-p}{\lambda}\right)^2.\end{aligned}$$

We may compare the above expression with that for $\text{var}^{(s)}(t_a)$:

$$\text{var}^{(s)}(t_a) = \frac{2(2-p)}{\lambda^2} + \frac{[2(1-p) - (2-p)^2]}{\lambda^2} = \frac{2-p^2}{\lambda^2}.$$

We next see that

$$\text{var}^{(p)}(t_a) - \text{var}^{(s)}(t_a) = \frac{2(1-p)^2}{\lambda^2} = \begin{cases} 0 & p = 1, \\ \frac{1}{2\lambda^2} & p = \frac{1}{2}, \\ \frac{2}{\lambda^2} & p = 0. \end{cases}$$

A similar and symmetric result is obtained with $\text{var}(t_b)$:

$$\text{var}^{(p)}(t_b) - \text{var}^{(s)}(t_b) = \frac{2p^2}{\lambda^2} = \begin{cases} 0 & p = 0, \\ \frac{1}{2\lambda^2} & p = \frac{1}{2}, \\ \frac{2}{\lambda^2} & p = 1. \end{cases}$$

Hence if both models are able to fit the means (the fit is guaranteed for the parallel model but not for the serial model), then the individual element's variances are predicted, in general, to be larger by the parallel model than by the serial model. It appears that this result will generalize to cases of $n > 2$.

A somewhat different sense of partial mimicking refers to approximation of parallel (serial) systems by serial (parallel) systems when the processing time distributions of single elements are selected from different families or classes of distributions for the two types of systems.

Consider the case where the only observable statistic from a parallel system is the time required to finish processing all the elements. If the number of elements is not known, then regardless of the parallel probability distribution on this statistic, there is a probability mixture of gamma distributions with an equal number of states in each that is arbitrarily close (in the sense of weak convergence) to the parallel distribution. The resultant distribution can be interpreted as a mixed serial system that mimics the parallel system.

A complete exposition of this result would require many pages and would repeat several standard developments in probability theory and analysis and would hence go beyond the purpose of the present paper. Nevertheless, because of the theorem's potential importance in stochastic approximation as well as in parallel-serial mimicking, and since it does not appear to be available in a form accessible to psychologists, we proceed to offer an informal proof that includes the main steps. A more recondite statement on approximation with mixed gamma distributions may be found in Kingman (1966). For an up-to-date account of weak convergence, the reader is referred to Billingsley (1968).

First, we note that weak convergence of probability distribution functions on the real line, $F_n \xrightarrow{w} F$, holds if and only if the sequence of distribution functions F_n converges to F on the set of continuity points and (therefore) on a set of points that is dense in R_1 (the real line). Next, any distribution function can be decomposed into, and written as a convex combination of, a discrete and a continuous distribution function. It is then possible to approximate any probability distribution, in the sense of weak convergence, by a sequence of discrete probability distribution functions.

Consider the open intervals $(a^{(i)}, b^{(i)})$, $i = 1, 2, 3, \dots$, where $a^{(i)} < b^{(i)}$ are contained in the continuity set of F , $C(F)$, for all i and $a^{(i+1)} < a^{(i)}$, $b^{(i)} < b^{(i+1)}$ for all i . This gives a sequence of expanding intervals whose end points are in the domain of the continuous part of F .

Divide up $(a^{(i)}, b^{(i)})$ into $k-1$ disjoint sections, formed from

$$a^{(i)} = a_1^{(i)} < a_2^{(i)} < a_3^{(i)} < \dots < a_r^{(i)} = b^{(i)} \text{ with } a_j^{(i)} \in C(F),$$

$j = 1, 2, \dots, r, r = 1, 2, \dots$, and construct a step function, F_{ir} , with values constant within any one section. Since within any bounded interval, the number of discontinuities is finite for any distribution function, we can construct our step function to take advantage of the discontinuities by letting $F_{ir} = F$ on those step intervals. Note that the assumption that $a_j^{(i)} \in C(F)$ is needed to ensure that the discontinuities are not end-points of our open intervals.

Next, make k sufficiently large that

$$\sup_{t \in C(a^{(i)}, b^{(i)})} |F_{ir}(t) - F(t)| < \varepsilon_r = \varepsilon/r,$$

that is, so that the maximum difference on points in $C(F)$ between F_{ir} and F is less than ε/r . Since F is bounded and has at most discontinuities of the jump variety, this construction is guaranteed.

Noting that $\varepsilon_r \rightarrow 0$ as $r \rightarrow +\infty$ and $(a^{(i)}, b^{(i)})$ grows without limit as i becomes arbitrarily large, it follows that $F_{ir} \xrightarrow{w} F$ as $i, r \rightarrow \infty$, and, in fact, that for any fixed $t \in C(F)$, when $i = i^*$ is large enough to include t in $(a^{(i^*)}, b^{(i^*)})$, $\lim_{r \rightarrow \infty} F_{i^*r}(t) = F(t)$.

This completes the demonstration that any distribution function can be arbitrarily closely approximated in the sense of weak convergence by a step-function distribution function. This result is necessary for the next step in our proof.

We may regard any discrete distribution function as a (generalized) integral of a probability mixture of Dirac delta functions:

$$F(t_m) = P(t \leq t_m) = \int_{-\infty}^{t_m^*} \sum_{i=-\infty}^m p_i \delta(t - t_i) dt,$$

where

$$\int_{-\infty}^{+\infty} \delta(t - t_0) dt = \begin{cases} 0, & t < t_0, \\ 1, & t \geq t_0, \end{cases} \quad \text{and} \quad \sum_{i=-\infty}^{+\infty} p_i = 1$$

(t_m^* implies that the point t_m , in the domain of F , is included in the integration). If we now show that we can form a sequence of gamma distributions that will approach any given Dirac delta function arbitrarily close (again in the sense of weak convergence), then we can form probability mixtures of as many of these as we need, and this mixture is our serial system.

It is again helpful to use the characteristic functions of gamma distributions for this

$$E[\exp(its)] = \left(\frac{\lambda}{\lambda - is} \right)^k = \phi(s)$$

for one with k stages. Suppose we wish a sequence that converges to a Dirac delta function with mass concentrated at t_0 . Such a sequence is found by letting $\lambda = k/t_0$, which assures us that the mean is always t_0 and the variance goes to 0

as k goes to infinity. In fact, we may write

$$\phi(s) = \left(\frac{k/t_0}{(k/t_0) - is} \right)^k = \left(\frac{1}{1 - (ist_0/k)} \right)^k \quad \text{and} \quad \lim_{k \rightarrow \infty} \left(\frac{1}{1 - (ist_0/k)} \right)^k = \exp(ist_0).$$

The letter i in the characteristic functions here is $\sqrt{-1}$. But $\exp(ist_0)$ is the characteristic function of the Dirac delta function with unit mass concentrated at

$$t = t_0: \quad \delta(t - t_0) = \int_{-\infty}^{\infty} \exp(ist_0) \exp(-ist) dt.$$

As noted earlier, equality of characteristic functions implies equivalence of distribution functions. Hence some probability mixture of gamma distributions converges to any mixture of Dirac delta functions and the latter converges weakly to any probability distribution whatsoever, and therefore there will exist a probability mixture of gamma distributions that will approach (weakly) any distribution function whatsoever.

If it is desired that each of the gamma distributions in a given approximating mixture have the same number of stages, say K , this may be accomplished by making K sufficiently large that the distribution function made up of the gamma processes is as close as is required to the approximating step function distribution, which in turn is as close as desired to the 'true' parallel distribution function. This can be carried out using standard δ , ε approximation arguments. The value of K might be interpretable as the (unknown) number of elements.

It is possible to extend this development to the n -dimensional joint distribution function of the parallel system's n elements, where n may now be known. The rationale is the same, but now the characteristic functions of the multidimensional probability mass functions are approximated by a weighted series of products of gamma characteristic functions, each characteristic function drawn from one of the n dimensions. For example, a probability mass placed at the point $(t_{01}, t_{02}, \dots, t_{0n-1}, t_{0n})$ is associated with the Dirac delta function product

$$\prod_{j=1}^n \delta(t_j - t_{0j}),$$

and the characteristic function of this, in turn, is approximated by

$$\prod_{j=1}^n \left(\frac{1}{1 - (is_j t_{0j}/k_j)} \right)^{k_j},$$

where we note that s_j denotes the transform variable for dimension j (i.e. element j). In this manner, a multidimensional distribution function can be approximated by a probability mixture of general gamma distributions, each component gamma characteristic function (i.e. for each s_j) in one of the products representing one of the elements. In this representation, the number of stages in the gamma distributions will not be the number of elements, n . The development is somewhat more unwieldy but follows the line illustrated by the one-dimensional case (e.g. for the time to completion of the last element, given above). We have thus shown that mimicking by approximation can occur in the context of parallel-serial identifiability.

There are several factors which, if present, could ameliorate this particular identifiability problem: (1) if a parallel distribution were discovered that was more natural or intuitive than the mixed-serial process, (2) if the particular means for the gamma process have no intuitive justification and the parameters of a parallel process do, (3) if the total number of processes in the mixture seems unusually large (as it is likely to, for a good approximation). On the other hand, the results can be viewed in a positive manner in the sense that we can always construct a serial system with manageable gamma distributions to approximate even the most complicated parallel distribution function.

In closing this section, we make passing reference to related simpler, but possibly very useful, facts concerning less global approximation abilities of gamma distributions. For example, it can be shown that a probability mixture of two exponential distributions with density function

$$f(t) = p\lambda_1 \exp(-\lambda_1 t) + (1-p)\lambda_2 \exp(-\lambda_2 t)$$

can produce distributions having any desired mean μ and fractional coefficient of variation σ/μ (σ = standard deviation) between 1 and ∞ . For smaller coefficients of variation, one may use a single k -stage process with characteristic function

$$\phi(s) = \prod_{j=1}^k \frac{\lambda_j}{\lambda_j - is},$$

which yields any mean and any coefficient of variation between 1 and $1/\sqrt{k}$. For a discussion of these and related facts, the reader is referred to Cox (1962).

9. QUALITATIVE CONSIDERATIONS

In addition to the possibilities for testing parallel versus serial systems that flow directly from the preceding mathematical treatment, there are some qualitative considerations that also relate to the mathematics and structure of parallel and serial systems.

Neisser *et al.* (1963) reported scanning for ten targets as rapidly as for one in a vigilance search task. Brand (1971), using a similar task, found strong category effects such that looking for one of a category of symbols placed in 'noise' symbols selected from a different category was as fast for practised subjects as looking for a single symbol in this type of array. In one of their visual span of apprehension tasks, Egeth *et al.* (unpublished manuscript, Expt. 1) recently described a mean reaction time function for detection of the presence of any number in an array of letters that was invariant across the number of letters in the array (2-6).

Assuming error rates and other possible artifacts have been properly controlled, these results argue for a parallel interpretation of the underlying mechanisms, at least for the specific conditions under which they were found (not all of the conditions in Egeth *et al.* yielded flat functions). Each of the studies reports reaction times that are essentially flat as functions of the number of elements that must be involved in processing. The serial processes that can mimic this

result are of two types. The first, not mentioned earlier, assumes that the serial processor operates so fast that the small increments in processing time due to an increased number of elements are unnoticeable. In most cognitive tasks, where even one-element processing reveals reaction times significantly larger than simple reaction times, this type may be summarily excluded. The other type, which actually mimics the parallel system mathematically, must assume that the processing rates (e.g. the λ 's) increase linearly as a function of the number of elements to be processed. Although this is mathematically unassailable, we do not expect, on the basis of general knowledge of biological and psychological processes, to often discover mechanisms that function in such a manner. If a system is capable of multiplying its speed with larger numbers of elements, why does it not go faster with one element? It should also be mentioned here that flat processing-time functions indicate that the pertinent parallel systems (and mimicking serial systems) must not only be of unlimited capacity, but also possess zero variance, and hence either be deterministic or based on fixed output times that are always longer than the processing times.

An asymmetry thus can exist in testing serial versus parallel processes: in many situations, an increasing straight-line reaction time function can be as intuitively identified with a limited capacity parallel system as with a serial system, but a flat reaction time function appears to offer fairly strong support for an unlimited capacity parallel system as opposed to a serial system.

This kind of asymmetry, to the detriment of serial processes, can also be found in the Estes & Taylor (1964) detection paradigm (described earlier). Not only does their later finding of independence seem more intuitively connected to parallel systems but invariance of probability correct with mutual distance of redundant test letters (Estes & Taylor, 1966) implies that a serial processor must be able to jump to any place in the display's 'image' with no loss in accuracy. With a fast fading icon, this seems unlikely. Nevertheless, these asymmetries in intuition do not seem quite as strong as those for the flat reaction time functions.

Finally, possibilities for testing serial versus parallel mechanisms by manipulation of element presentation should be mentioned. For example, consider the Estes & Taylor (1964) paradigm with, say, five letters in a visual array and assume the facility of presenting the letters one at a time in sequence for duration Δt each (this is easily arranged with a contemporary fast-phosphor oscilloscope in conjunction with a real-time computer system) in one condition. In the second condition, present the letters simultaneously for $5\Delta t$. If the amount of time available to process each letter can affect accuracy significantly, equality of performance would provide rather strong support for a serial system. Either a limited or unlimited capacity parallel system should predict superiority on the simultaneous condition, excluding other possible artifacts. In fact, the most parsimonious system would be deterministic and serial; e.g. a processor that divides the total available time into equal segments and samples a fixed amount of information from one letter during each segment. In general, a stochastic serial system would predict superiority on the simultaneous condition, since if the processor finished before its time segment was over, it could go on to the

next letter in the simultaneous but not the sequential condition. Also, it might be that a process, normally serial, could not time-lock itself into an externally imposed segmented presentation. For these reasons, superiority of the simultaneous condition would be a more ambiguous result. Superiority of the sequential condition would be difficult to explain with either type of model without the aid of auxiliary notions such as lateral inhibition.

An alternative experiment where the simultaneous condition employed a presentation duration of Δt and the sequential condition presented each letter for Δt apiece would provide a test in the opposite direction. If the simultaneous condition were equal to the sequential, the support would accrue to the parallel system, but only if the serial processor were not being adversely affected by the external manipulation of the visual field. Superiority of the sequential condition could be predicted by a serial system or by a parallel system with the facility to reallocate attentional energy during the sequence of letters, thus becoming effectively serial. Inferiority of the sequential condition would seem to suggest some interference produced by the method of presentation. Eriksen & Spencer (1969) have performed experiments that approximate these latter conditions, obtaining the equality result and therefore apparent support for parallel processing under their conditions. In fact, their results, along with those discussed above, provide mounting credibility of parallel processing ability in 'detection-type' visual span of apprehension experiments.

Neither of the discussions in the preceding two paragraphs negates the earlier mathematical results. For example, a parallel model could predict equality in the $5\Delta t$ simultaneous and Δt sequential conditions by assuming decreased processing rates in the $5\Delta t$ condition, perhaps due to a small upper bound on the amount of information that can be processed by a single 'channel'. Similarly, in the Δt simultaneous and Δt sequential conditions a serial model could predict equality by assuming that its rate is five times faster in the simultaneous condition. Other assumptions or artifacts are also possible. Nevertheless, despite the mathematical non-identifiability, these conditions appear to be the strongest suggested to date, in terms of intuitive strength, for investigations using only means as evidence. It appears that for strong conclusions to be drawn using means, several converging operations may be necessary.

10. CONCLUDING REMARKS

With respect to processing a single set of elements, there are several logically independent aspects that have particularly engaged the attention of experimentalists. These are whether processing is exhaustive or self-terminating (e.g. on trials where only a subset of the elements carry all the information), limited or unlimited capacity (usually with respect to variation in number of elements), independent or dependent on the various elements, and whether it is parallel or serial or hybrid. It appears that the first three are relatively tractable compared to the parallel-serial issue. For example, it is possible to generate serial position effects in the Sternberg (1966) paradigm and in the Atkinson *et al.* (1969)

paradigm (Townsend & Roos, in preparation). If these effects are due to events in the processor under study, they imply self-termination. When taken in conjunction with the parallel reaction time lines of positive and negative trials (i.e. of trials with test letter present and test letter absent, respectively), they suggest the existence of a dual processing mechanism (see Bamber, 1969).

Capacity and independence hypotheses are even more readily testable. Although in black box modelling it is probably always possible to insert extra mechanisms to obscure the functioning of the unit under study, these would often seem contrived in the context of these latter two problems. Such would seem to be the case for the Estes and Taylor independence finding (Estes & Taylor, 1966). And thus the Neisser *et al.* (1963) type or flat search time function, assuming error rate is held constant, is a strong argument for (relatively) unlimited capacity. However, the above results suggest that despite its theoretical importance (see Introduction, Haber, 1969) the parallel-serial question should be ruled as unresolved in most of the contexts in which it has been studied.

Viewed in a broader context, the parallel versus serial problem may be addressed to general questions of allocation of processing energy in time and space, not only within a single processing unit as to how it processes input elements, but also as to distribution of energy to any of a number of subsets of processing units, e.g. multimodality perception and distribution of processing energy to encoding as opposed to short-term memory rehearsal processes. From this viewpoint, further mathematical work on parallel and serial types of questions in psychological systems seems of paramount importance.

REFERENCES

- ATKINSON, R. C., HOLMGREN, J. E. & JUOLA, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Percept. Psychophys.* **6**, 321-326.
- AVERBACH, E. & SPERLING, C. (1961). Short-term storage of information in vision. In C. Cherry (ed.), *Information Theory*. London: Butterworth.
- BAMBER, H. K. (1969). Reaction times and error rates for 'same'-'different' judgments of multidimensional stimuli. *Percept. Psychophys.* **6**, 169-174.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. New York: Wiley.
- BRAND, J. (1971). Classification without identification in visual research. *Q. J. exp. Psychol.* **23**, 178-186.
- COX, D. R. (1962). *Renewal Theory*. New York: Wiley.
- EGETH, H. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Percept. Psychophys.* **1**, 245-252.
- EGETH, H., JONIDES, J. & WALL, S. Unlimited-capacity parallel processing of multi-element displays. (Unpublished manuscript.)
- ERIKSEN, C. W. & SPENCER, T. (1969). Rate of information processing in visual perception: some results and methodological considerations. *J. exp. Psychol. Monogr.* **79**, no. 2, part 2, 1-16.
- ESTES, W. & TAYLOR, A. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Proc. Nat. Acad. Sci.* **52**, 446-454.
- ESTES, W. K. & TAYLOR, H. A. (1966). Visual detection in relation to display size and redundancy of critical elements. *Percept. Psychophys.* **1**, 9-16.
- GARDNER, G. T. (1970). Spatial processing characteristics in the perception of brief visual arrays. (Doctoral dissertation, University of Michigan. Appears as Technical Report no. 23, University of Michigan, Human Performance Center.)

- GREENO, J. G. & STEINER, T. E. (1964). Markovian processes with identifiable states: general considerations and applications to all-or-none learning. *Psychometrika* **29**, 309-333.
- HABER, R. N. (1969). *Information Processing Approaches to Visual Perception*. New York: Holt, Rinehart & Winston.
- HALMOS, P. R. (1950). *Measure Theory*. Princeton: Van Nostrand.
- KINGMAN, J. F. C. (1966). *On the Algebra of Queues*. London: Methuen.
- KOOPMANS, T. C. (ed.) (1950). *Statistical Inference in Dynamic Economic Models*. New York: Wiley.
- MCGILL, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R. R. Bush & E. Galanter (eds.), *Handbook of Mathematical Psychology*, vol. 1. New York: Wiley.
- MCGILL, W. J. & GIBBON, J. (1965). The general-gamma distribution and reaction times. *J. math. Psychol.* **2**, 1-18.
- NEISSER, U., NOVICK, R. & LAZAR, R. (1963). Searching for ten targets simultaneously. *Percept. mot. Skills* **17**, 955-961.
- RUMELHART, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *J. math. Psychol.* **7**, 191-218.
- SAATY, T. L. (1961). *Elements of Queueing Theory*. New York: McGraw-Hill.
- STERNBERG, S. (1966). High-speed scanning in human memory. *Science, N.Y.* **153**, 652-654.
- THOMAS, E. A. C. (1969). Alternative models for information processing: constructing non-parametric tests. *Br. J. math. statist. Psychol.* **22**, 105-113.
- TOWNSEND, J. T. (1969). Stochastic representations of parallel and serial processes. (Paper presented at Mathematical Psychologists Conference, Indiana University.)
- TOWNSEND, J. T. (1971a). Theoretical analysis of an alphabetic confusion matrix. *Percept. Psychophys.* **9**, 40-50.
- TOWNSEND, J. T. (1971b). Alphabetic confusion: a test of models for individuals. *Percept. Psychophys.* **9**, 449-454.
- TOWNSEND, J. T. (1971c). A note on the identifiability of parallel and serial processes. *Percept. Psychophys.* **10**, 161-163.
- TOWNSEND, J. T. & FIAL, R. (1968). Spatio-temporal characteristics of multi-symbol perception. (Paper presented at Mathematical Psychology Meetings, Stanford University.)
- TOWNSEND, J. T. & ROOS, R. N. Search reaction time for single targets in multiletter stimuli with brief visual displays. (In preparation.)