James T. Townsend

Reprinted from JOURNAL OF MATHEMATICAL PSYCHOLOGY
All Rights Reserved by Academic Press, New York and London

Vol. 14, No. 1, August 1976
Printed in Belgium

A Stochastic Theory of Matching Processes[1]

JAMES T. TOWNSEND

Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907

A theory of matching processes is developed within which serial models and parallel models based on within-stage independent intercompletion times are defined. These models are then specialized to a class of processes possessing exponential intercompletion time densities and equivalence properties of parallel and serial models within this class are investigated. Nonequivalence theorems are proved that designate possible differences in "same" (+) and different (−) matching rates as important in testing parallel and serial models. An experimental paradigm is then derived with the property that if + rates are different from − rates, and if processing is self-terminating, then the parallel and serial models are distinguishable at the level of mean reaction times. The serial class of models that is tested by this paradigm includes a large number of stochastic distributions whose central assumption is additivity of element processing times. The corresponding parallel class of models is currently limited to those assuming exponential intercompletion times. A numerical example and an example of nonparametric relations predicted by the serial or parallel models are given. Some advantages and limitations of the present treatment are discussed.

1. INTRODUCTION

Within the realm of human cognition, there are many instances where information from one source or subprocess must make contact with and be compared with information from another source or subprocess. This paper is concerned with the stochastic properties of such comparison processes as can be represented by one set of elements matched against another set. Each element in one set either matches (is identical to) or mismatches any given element in the other set. The applications of the theorems discussed in this paper are primarily directed to reaction time paradigms with low error rates. However, the mathematical theory provides the distribution of uncompleted elements at any point in time and this may be paired with deadline and response bias assumptions to derive predictions of speed–accuracy tradeoff.

[1] I am grateful to W. K. Estes, Rockefeller University and the National Institute for General Medical Sciences for support of a visiting Associate Professorship under Grant GM16735, during which time (1972–73) about 40% of the present work was completed. I am appreciative of the comments on an early draft of this manuscript by W. K. Estes, J. C. Falmagne, S. Link, F. Norman, and D. Vorberg, and for helpful remarks on exposition by W. Batchelder and two referees.

Some of the kinds of comparison processing to which this work may be relevant are: (1) *Pattern recognition*: A single pattern (e.g., a letter) is presented on each trial and the subject's task is to give a response identifying the pattern (e.g., Townsend, 1971a). The match is between the stimulus pattern and memory information, corresponding to the total set of possible patterns, stored in some fashion in the subject's memory. The term "element," used below in theorizing, might refer to a feature of a pattern and matching would refer to comparing the stimulus pattern features with features stored in memory. (2) *Multidimensional pattern discrimination*: On each trial, two patterns that may differ on one or more dimensions (size, color, etc.) are compared and the subject responds "same" if the patterns are identical on each dimension and "different" otherwise (e.g., Egeth, 1966). The elements that undergo comparison or match processing are the dimension values present on a given trial. (3) *Display and memory search*: One or more symbols or patterns are given to the subject on each trial followed by a second display containing one or more symbols to be compared with the memory symbols. Often, visual span of apprehension is studied by putting only one target symbol in the subject's memory and then giving him a multisymbol display presented for a short duration (Atkinson, Holmgren, & Juola, 1969). Memory scanning often involves the opposite, namely presenting several symbols for memory storage followed by presentation of a single target or critical symbol (e.g., Sternberg, 1966). In both of the latter cases, the subject responds "yes" indicating the presence of the target in the multisymbol set or "no" indicating its absence. Both of these designs are closely related to a two-alternative forced choice situation invented by Estes and Taylor (1964). The usual breakdown in these experiments would have the symbols being the elements, but the symbol elements may themselves be considered as made up of smaller features. In the latter instance, processing questions may be found on both levels.

This paper is divided into seven sections. In the remainder of the Introduction, previous germane work is briefly reviewed and the approach and goals of the present study are discussed. The reader interested in moving immediately to the formal development may wish to bypass the remainder of this section. The second section builds a theoretical framework for explication of the structure of comparison processing and its modeling. Section 3 introduces the models with which we are specifically concerned here and gives them a mathematical characterization. Next, Section 4 provides general equivalence and nonequivalence results for the present serial and parallel models. Following this, a special case (Section 5) closely related to a contemporary type of matching experiment is developed in detail and is used to illustrate model structure and equivalence results. The sixth section gives a parallel–serial testing paradigm (PST), derived from considerations of the earlier theorems, that tests the present serial and parallel models against one another. Some concluding remarks on PST appear in the seventh and final section. Reading the first two sections (excluding the Introduction) is necessary for understanding the rest of the

paper. However, the proofs of the theorems of Section 4 may be skipped and their content garnered from their exemplification in Section 5. It may also be helpful to read Section 5 in conjunction with Section 3. The last two sections, which develop and discuss PST, require familiarity with earlier notation and rest upon the theorems of Section 4, but may be read independently of the latter.

Accompanying the significant interest in theories of how information is processed, stored and used (not to be confused with information theory, per se are experiments and models addressed to questions of how processing resources are deployed to complete some finite number of comparisons, as well as how much in the way of resources are available. However, in at least certain major areas of research not only has there been a general lack of theoretical framework, but rather little in the way of mathematical treatment seems to have been accorded either models or data analysis.

Attempts have recently been initiated to set up a qualitative framework for some important aspects of processing (Gardner, 1973; Townsend, 1971b) and to provide a mathematical framework and properties of identifiability for these and similar aspects (Townsend, 1969, 1972, 1974; Townsend & Roos, 1973). Among such aspects of processing mechanisms are the dichotomies of self-terminating vs exhaustive processing, dependence vs independence of processing, limited vs unlimited capacity and serial vs parallel processing. (For detailed discussions of these issues the reader is referred to Gardner, 1973; Townsend, 1974.) Although all of these can be facilely discussed in the mathematical context developed in the above papers, the parallel vs serial issue (roughly simultaneous vs one-at-a-time processing of several elements) has been emphasized because of its great interest to investigators (see, e.g., Haber, 1969, Introduction) and the fact that problems of parallel–serial testability were shown to be much more refractory than was previously recognized. Further, the other issues differ in their interpretation according as processing is parallel or serial. A brief, more specific review of recent related work follows.

Historically, Christie and Luce (1956) appear to have been the first to engage in a mathematical investigation of parallel and serial processes within a psychological context. Their work was primarily concerned with processing of all the elements with constant and equal rates on distinct elements. In their case, an "element" was a component of a decision process, and they considered the general case of unknown numbers of objects. Typically, experimenters have unfortunately not employed these results in other than a qualitative manner. Further, it later became apparent that there existed severe problems of parallel–serial equivalence even when the number of objects to be processed was assumed known (Townsend, 1969, 1971b). Although the most general results to date on the latter problem have not yet been embedded in the case where the number of objects is not known, we can probably expect the difficulties in testability to be substantially augmented.

McGill and Gibbon (1965) provided an exposition on the general gamma distribution, in particular with regard to its use in modeling RT. Reaction time was viewed as

possibly consisting of a sequence of stages, the random duration of each stage being governed by an exponential density. This hypothesis corresponds to Townsend's (1971b, 1972) assumption of exponential intercompletion times within stages of processing a finite number of inputs. That assumption was employed in establishing the primary equivalence results early in the latter paper (specifically, Townsend, 1972, p. 174–184).

Rumelhart (1970) succeeded in fitting data from a wide array of visual multisymbol processing experiments. His model was parallel across the various symbols but features were assumed to be extracted one at a time. His intercompletion times for features were distributed exponentially. Like McGill and Gibbon (1965), Rumelhart was not specifically concerned with the parallel–serial equivalence question. However, it follows from recent results that the aspect of his model concerned with processing of features of individual symbols may be equally well expressed as serial or parallel (Townsend, 1971b, 1972). A remark on the overall characteristics of the model will be made below.

In a development directly germane to parallel–serial testability, Thomas (1969) showed that under certain circumstances, nonparametric tests can be constructed to test between independent, self-terminating parallel models and nonindependent self-terminating serial models. The type of independence referred to here is that of total completion times (see Townsend, 1974, p. 152). Some further discussion of Thomas' work will be found in the final section.

A series of studies on parallel and serial processing that the present writer has been involved with over the last few years can be briefly summarized as follows.

The first paper (1969) dealt with the absolute nontestability of serial models that predict, for example, increasing straight lines as n, the number of elements to be processed, is varied, and the mathematically equivalent parallel model, both assuming exponential intercompletion times. The existence of serial models equivalent to certain parallel independent models was also indicated. Atkinson *et al.* (1969) listed a special case of the former results. A qualitative note (1971b) summarized these and some of the results presented in a more mathematical treatment (1972). The latter extended the findings in several directions. (1) Consider the classes of all parallel and serial models based on independent exponential intercompletion times. The parallel models further possess the property of independence of processing among the elements during a single stage (the state of the system between two successive element completions). Mixed serial models assume a set of probabilities, summing to 1, on the different possible processing orders for each trial. Mixed parallel models assume a set of more than one parallel system, one of which is implemented on each trial according to a set of probabilities summing to 1. These classes of mixed serial and mixed parallel models are nonequivalent in that there are models in each class that are not identical to any model in the other class. However, there are substantial degenerate regions where equivalencies exist. Thus, the class of unmixed parallel models is contained in

the class of mixed serial models. The class of unmixed serial models (only one processing order is ever used) may be approximated arbitrarily closely by the class of unmixed parallel models but is never equivalent to a parallel model. The classes of models in which these results are based may be viewed as probability mixtures of general gamma distributions of the type investigated by McGill and Gibbon (1965). (2) There may be difficulty or impossibility of testability at different levels so that even models that are not mathematically equivalent may give identical predictions at a given level, for example, for mean reaction times. (3) Related to (2), it may sometimes be feasible to take two nonequivalent models that give the same predictions at one level, and derive an implied difference in predictions at another level and thus a potential testable consequence. (4) A functional equation that is necessary for mathematical equivalence of parallel and serial models to hold was shown and some special cases that did and did not satisfy this equation were pointed out. (5) It was shown that Rumelhart's (1970) model is a special case of a class of parallel models that is equivalent to a class of hybrid time-sharing models. The nonhomogeneous aspect of the Rumelhart model does not affect the efficacy of the proof and was omitted. Two alternate nonparallel interpretations are either as a serial model with a rather bizarre density function or as a serial model with a variable number of features processed from trial-to-trial rather than the fixed criterial numbers as the parallel model supposes. (6) The problem of mimicking by approximation, rather than by actual equivalence, on one or more statistics was treated. For instance, a parallel independent model is not actually equivalent to typical serial models that predict a straight mean RT function as n increases. Yet such models can predict this function, since it is dependent only on first moments, not on other aspects of the model's probability distribution. Such parallel models apparently had been thought to be falsified by linear RT functions that increase fairly rapidly (Sternberg, 1966), but this is wrong.

Addressing a different type of approximation, a theorem was proved showing that without knowledge of n (e.g., as in the Christie and Luce (1956) case, the elements might be internal to the human processor) or other ameliorating observable quantities, a serial model made up of gamma distributions can always be constructed to arbitrarily closely approximate any parallel processing distribution. (7) Qualitative considerations concerning conclusions based on "naturalness" and "intuitive value" were entertained.

More recent work includes a detailed treatment of the (parallel vs serial) $\times$ (exhaustive vs self-terminating) $\times$ (independent vs dependent) $\times$ (unlimited vs limited capacity) processing dimensions emphasizing their logical and mathematical independence, some of the confusions appearing in recent literature, and the way various combinations of these dimensions may be employed in model building (Townsend, 1974). Also completed are theorems exhibiting fairly broad conditions (i.e., not restricted to models assuming exponential intercompletion times) on models satisfying a functional equation posed earlier (1972) that is necessary for parallel and serial models to be equivalent (Townsend, in press).

One of the first obvious conclusions to be drawn from the above work is that many, if not most, of the experimental studies purporting to test serial vs parallel processing have not in reality done so. Two directions present themselves. One is to develop mathematical theories of information processing that tell where to look for observable differences between parallel and serial processing and indeed, other important aspects of processing as well (e.g., self-termination vs exhaustive processing, etc.). The other is to delve into the physiology of the central nervous system. Although significant progress has been recorded recently in the latter, it appears that a close and certain linkage of the neuroanatomy and physiology with much of the cognitive processing literature is not presently at hand.

The former, mathematical theory building can be roughly divided into two approaches. The first is the utilization of finer grained statistics and distribution information in contemporary and past experiments. To the extent that the underlying processing distributions are not of the type where equivalence obtains between say, parallel and serial models, this type of information may be of value. The second is the construction of experimental paradigms, especially designed from consideration of the pertinent mathematical structures to produce testable empirical consequences.

The second approach is the one followed here. Classes of serial and parallel models are studied in terms of their structure and equivalence properties. Out of this structure comes an experimental paradigm that tests between these classes of serial vs parallel models at the level of observable mean RTs.

More specifically, the present work provides a relatively thorough and rigorous treatment of parallel and serial comparison models based on exponential intercompletion times. Intercompletion time is the time between the successive completion of two elements. Some results presented earlier (e.g., some of the equivalence mappings in Townsend, 1972) in an informal manner are special cases of the theorems given here. The generalizations in the work here take three directions: (1) Pairs of elements are considered to be processed by comparison against each other; previously, only single elements were considered. (2) There are an arbitrary number of elements in each of two sets (n_1 in the first, n_2 in the second) from which one or more pairs are drawn for processing or matching; previous proofs are based on two or three elements. Thus, predictions for special cases of specified numbers may be immediately obtained by plugging the numbers in formulas presented here. (3) The processing rates, whether serial or parallel, are allowed to depend not only upon the individual stimulus locations of the pair of elements and the processing order, but also upon whether an identity (+) match or a mismatch (−) comparison is taking place. Items (1) and (2) allow formulas and theorem results to be applied to any situation involving processing of pairs of elements from two finite sets, with arbitrary numbers of elements in each. Of course, when only a single set of elements is undergoing processing, the present formulation gives model structure and predictions by allowing certain parameters to attain degenerate values.

Generalization (3) is the most important. Within classes of models based on exponential intercompletion times, and with rates dependent only on individual locations and/or processing order, it has been shown that the class of unmixed parallel models is contained within the class of mixed serial models, but not vice versa, but that mixed parallel models were not contained within the class of mixed serial models (Townsend, 1972). However, it is shown in the present study that when processing or comparison rates are different for + comparisons than for − comparisons, the class of unmixed parallel models is neither contained in nor contains the class of mixed serial models.[2] Further, the cases previously found where serial models could not be mimicked by parallel models often precluded experimental exploitation due to the lack of observable quantities in the usual experimental settings (see, e.g., Townsend, 1972; Townsend, 1974). When − rates are distinct from + rates, it is possible to design conditions where fundamental differences are predicted in easily acquired mean RT's. Such an experimental paradigm is exhibited and testability theorems are proved here, along with comments on potential advantages and pitfalls associated with its application.

In previous papers (Townsend, 1971b, 1972), two models, one parallel, the other serial, were said to be nonidentifiable if and only if there existed a mapping between the parallel and serial parameters in such a way that the stochastic processes were mathematically identical; otherwise they were said to be identifiable. This is consistent with earlier uses of "identifiability" (e.g., Hurwicz, 1950) when as "model" we take the union of the parallel and serial models we wish to consider, and when as "data" we take the union of any data defining the parallel and serial processes (e.g., the set of all realizations of the two processes). Although mathematically correct, this approach is not wholly in line with previous terminology in psychology (see, e.g., Greeno & Steiner, 1968). It also potentially obscures questions of identifiability that may arise *within* a parallel or serial model and with respect to either data that is short of defining a process, perhaps a set of expectations or data that is sampled, for example a set of observed means. To attempt a more appropriate terminology the development will follow Greeno and Steiners' usage in referring to "equivalence" between theories (or models, with which we shall be concerned here) rather than "identifiability." "Equivalence" as employed in the present context is made precise below and notions of testability are consequent.

It is pertinent to note here that if it were possible and realistic to assume that processing occurred in discrete time (assuming the concurrent observability problem had been solved) then it might be a simple matter to discriminate parallel from serial processing, since serial processing excludes the possibility that more than one element can be completed in any single interval, but parallel processing does not. Even the analog to exponential processing of elements, geometric processing, would in this case

[2] Hereafter, the discussion will be centered on *mixed* serial models (more than one processing order may occur) and *unmixed* parallel models (the same set of parallel rates apply on every trial) and the mixing designation dropped for brevity.

present no difficulty. Unfortunately, attempts to represent human information processing as involving immutable time-quanta have not been notably successful.

In the present paper, the mathematics for particular serial and parallel models are developed that may be applicable when an experiment requires matching or comparison (matching is a primitive with the usual interpretation of detecting whether or not two elements differ) and show under what conditions they are nonequivalent. Further, as noted above, it is shown that experiments can be designed where parallel and serial differences in predictions are obtained at the level of overall mean reaction times.

2. Matching: Systems, Models, and Parallel–Serial Equivalence

We shall follow the tack of using informal definitions to aid in the somewhat formidable task of bookkeeping and exposition. It should not be supposed, however, that the present formulation attempts to be axiomatic in a strict sense.

Definition 1. An experimental trial is a matching trial if and only if a subject matches the elements from one (nonempty) finite stimulus set, referred to as S_1, with the elements from another (nonempty) finite stimulus set, S_2.

It is assumed that the elements of both S_1 and S_2 may be ordered along at least one dimension, called serial position. It is also assumed that S_1 and S_2 have been presented to the subject and that they are themselves subsets of other "alphabets" $\mathscr{S}_1$, $\mathscr{S}_2$. Often, $\mathscr{S}_1 \equiv \mathscr{S}_2$ (e.g., Atkinson *et al.*, 1969) but occasionally one or more elements in one, say S_2 may be drawn from a set with members not contained in $\mathscr{S}_1$ although all elements in $\mathscr{S}_1$ are in $\mathscr{S}_2$. Hence, in the latter case $\mathscr{S}_1 \subseteq \mathscr{S}_2$ (e.g., Egeth, Jonides, & Wall, 1972). As illustration, let $\mathscr{S}_1$ be a set of numerals and $\mathscr{S}_2$ be the same set of numerals *union* the set of English uppercase letters. In a typical experimental trial, S_1, a subset of $\mathscr{S}_1$, is presented to the subject. Following this, S_2, a subset of $\mathscr{S}_2$, containing zero or more members of S_1, is presented and the subject's task is to indicate whether or not a member of S_1 was in fact contained in S_2. The use of elements in $\mathscr{S}_2$ that are not in $\mathscr{S}_1$ allows the study of category effects in perception and memory (see, e.g., Brand, 1971; Egeth, Jonides, & Wall, 1972). Hereafter, the terms "element" and "stimulus" will be used interchangeably, although the processing structure developed here is also appropriate to systems matching totally unobservable elements.

It should be noted that in general, the classes of elements from which the sets $\mathscr{S}_1$, $\mathscr{S}_2$ are drawn will be germane in modeling specific psychological situations. However, reference to these classes can be suppressed in most of the development here and to do so will permit simplification of an already complex notation.

Corresponding to the notion of a matching trial is that of a simple matching experiment. The latter is relevant to definitions of parallel and serial models and to establishment of model diversity.

DEFINITION 2. A simple matching experiment is a set of matching trials made up of two types of trials: (a) T_1 trials: None of the elements in S_1 match any of those in S_2 and (b) T_2 trials: One or more of the elements in S_1 and S_2 match each other.

Next, we proceed to give definitions of systems, serial, parallel and hybrid systems, models, total equivalence and x-diversity. Basically, systems are viewed here as composed of mechanisms made out of living or nonliving material and that function in real time and space. Models are abstract descriptions of systems. Although the elements referred to immediately below will subsequently be interpreted as themselves pairs of elements (tuples) from S_1 and S_2, it seems preferable to employ "element" in a universal sense here to lend generality to certain of the following definitions. A postulate that applies throughout the paper is that rates of processing, in systems or models, are finite.

DEFINITION 3. A serial system is a *material* (i.e., real) entity that (*a*) processes members of a set of elements one at a time; (*b*) each element is completed before another is begun; (*c*) processing on each successive element may be instituted without delay when the preceding one is completed; (*d*) the order of processing may, in principle, be established before processing actually begins according to the serial positions of the elements; (*e*) the processing time of an individual element may depend on its nature (e.g., identity); (*f*) but neither the order of processing nor the processing time of an element can depend on the nature (e.g., identity) of unfinished elements.

Items (*d*), (*e*), and (*f*) of Definition 3 will be particularly important in establishing differences between parallel and serial models.

DEFINITION 4. A parallel system is a material entity that (*a*) begins processing on all members of a set of elements simultaneously; (*b*) processing of each element proceeds until it is completed; (*c*) the order of completion of the elements is determined entirely by their processing times which in turn may depend on (*d*) the nature of the elements (e.g., identity) and their serial positions; (*e*) but the processing time of an element on a given trial cannot depend on the order (relative to serial position) of completion of elements completed after it.

As in the serial case, items (*d*), (*e*) will be critical in deriving distinctions between serial and parallel models. Definition 5 establishes a category of systems which by fiat exhausts the remaining types of processing systems.

DEFINITION 5. A hybrid system is one that does not satisfy Definitions 3 or 4.

Figure 1a shows the activity of a serial and a parallel system, respectively, when two elements $o_1 = (a)$, $o_2 = (b)$, are being processed. The "z" variables are the processing times on the individual elements. That is, z_{a_1} is the amount of time actually spent by the system on (a) when it happens to be completed first and z_{b_2} is the corresponding time spent on (b) on the same type of trial ((a) completed first) and so on. In Fig. 1b, the same activity is viewed in terms of the durations between completion of the elements, hereafter called *intercompletion times*. So, t_{a_1} is the time from $t = 0$ up until element (a) is completed, when it is finished first; t_{b_1} is the corresponding time when (b) is completed first (not shown in Fig. 1). Similarly, t_{b_2} is the duration from

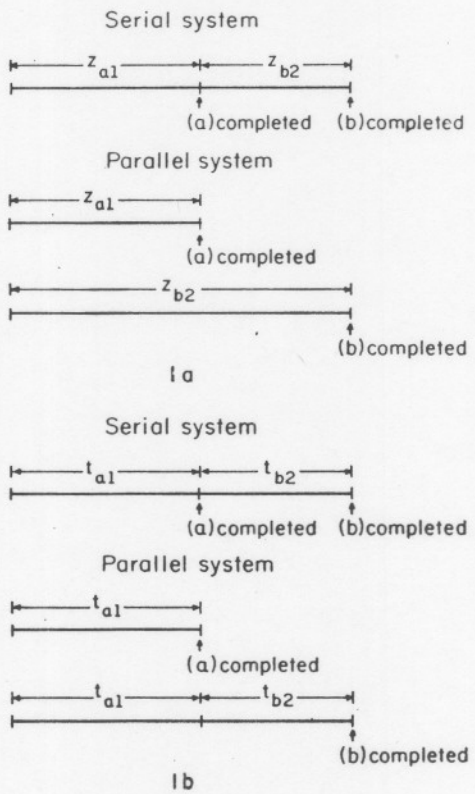

FIG. 1. (a) Schema showing total time system is engaged in processing (a) and (b) in a serial or parallel system when (a) is completed first. (b) Schema showing intercompletion times in processing (a) and (b) in a serial or parallel system when (a) is completed first. Figures 1a and 1b reprinted with permission from J. T. Townsend, "Issues and Models Concerning the Processing of a Finite Number of Inputs," in B. H. Kantowitz (Ed.), *Human Information Processing: Tutotials in Performance and Cognition* (Potomac, Md.: Erlbaum, 1974).

the instant of completion of (a) up until (b) is also finished. Finally, t_{a_2} is the corresponding time between the instant of completion of (b) and the completion of (a) where (b) is completed first (not shown in Fig. 1). The intercompletion times t_{a_1}, t_{b_1} can also be said to be the durations of *stage* 1 (no elements completed) and t_{a_2}, t_{b_2} the durations of *stage* 2, the time between the first and the second completions.

It is very important to observe that in a serial system, the element processing times *are* the intercompletion times. Thus, in Fig. 1, $z_{a_1} = t_{a_1}$ and $z_{b_2} = t_{b_2}$ in the case of serial processing. In parallel systems, on the other hand, only the first element to be completed will possess a processing time equal to an intercompletion time (e.g., $t_{a_1} = z_{a_1}$). All other parallel element processing times will be sums of intercompletion times. For instance, in Fig. 1, $z_{b_2} = t_{a_1} + t_{b_2}$ in parallel processing. A primary source of difficulty in experimentally discriminating parallel and serial systems lies in the nonobservability of the element times themselves, the z's. All we can see in outputs (at best) are the intercompletion times. As may be inferred from Fig. 1 and this fact, the intercompletion times are an important class of theoretical variables. When intercompletion times are independent, as they will be in the models considered here, the resulting convenient convolution properties of characteristic or moment-generating functions are very useful. We shall depend on them a great deal.

The systems of Definitions 3 to 5 are mutually exclusive and together they exhaust the class of processing systems. Examples of processing systems might be subsystems of a brain, a specific queuing system, a digital computer or a pattern recognition device. The more subsystems that make up a system and the larger the class of elements to be processed by a system, the less likely it can be defined as serial or parallel and the more likely it is hybrid. Again, it is critical to note that "system" is employed here to refer to actual biological or mechanical entities. In many situations, "system" may be (and is) employed as an abstraction much like we use "model" here. It is probably preferable to use a common word and establish a convention for its meaning rather than invent neologisms.

Corresponding to systems are their numerical or quantitative descriptions. It is typical of many biological and psychological systems that such descriptions are stochastic, so that even the most refined and stable data lead to a probability distribution on observable events and on the events internal and external to the system that together yield the observed distribution. Thus, investigation of the internal mechanisms and their interaction must be complemented by a theory of estimation leading to "best" estimates of the parameters of a probability distribution and by a theory of contribution of external factors to the observed distribution. We will here be concerned with the mathematical description of the internal mechanisms of systems (of matching processes) with little reference to the latter two problems. For example, the problem of estimation per se will not be touched on, but consequences of the existence of certain observable statistics (and therefore a priori assumed susceptibility to estimation) will be considered.

Our quantitative description will be termed a "model" and "model" will be taken to be a stochastic process. Deterministic systems (fixed x-variables) may be considered as degenerate stochastic processes.

DEFINITION 6. A model of a processing system is a subclass of the class of stochastic processes whose domain and sigma field correspond to the processing event possibilities in the system (see, e.g., Loeve, 1955).

Definitions 7 to 9 attempt to capture the basic characteristics of parallel, serial, and hybrid models. In order to conserve the intuitive aspects of the structure of such processes and because the processes examined in detail below can be expressed in such a way, Definitions 7 to 9 will be given in terms of probability densities.

It would, in a sense, be most natural to first define serial and parallel models in terms of the x-variables—the actual processing times. However, all consequent theorems on equivalence are greatly simplified by development in terms of the (potentially observable) intercompletion times (the t's). Furthermore, the definitions may be written in terms of the intercompletion times in such a way that no loss of generality is incurred, and this is the course taken here.

The earlier definitions of systems explicitly allow for dependence of an element's or (element-pair's) processing time upon the specific individual nature of that element. However, the major results in this paper are a joint result of parallel–serial differences in dependence on serial position, processing order and match (+) vs mismatch (−) times. The identity of the elements will only be important here in defining a match vs a mismatch. Therefore, the following definitions and derivations will not take into account processing time or possible order differences that depend explicitly on element identity per se.

In Townsend (1972), only single elements were considered to be processed. This corresponds, of course, to letting S_1 or S_2 contain one element in a matching experiment so that the "processing" of the multielement set is the matching of them against the single element in the other set. Since the processing times, and in a serial system the order of processing, may depend on the serial positions of both elements in a pair, we adopt a lexicographic ordering $\lesssim$ on the element pairs $\langle o_{1i}, o_{2j}\rangle$, $o_{1i} \in S_1$, $o_{2j} \in S_2$ so that $\langle o_{1i}, o_{2j}\rangle \lesssim \langle o_{1k}, o_{2m}\rangle$ if and only if either $i < k$ or $i = k$ and $j \leqslant m$. This amounts to taking the serial position of the S_1 element as predominant and secondarily ordering according to the S_2 element. For example, $\langle o_{13}, o_{25}\rangle \lesssim \langle o_{14}, o_{24}\rangle$ but $\langle o_{13}, o_{24}\rangle \lesssim \langle o_{13}, o_{25}\rangle$. We may then assign the numbers 1 to n_1n_2 to the consequent ordered list of pairs and call each number the "serial position index" or simply "serial position" of any particular pair. Therefore the term "serial position" (of a pair) will be used with this meaning henceforth and $a_k = \langle o_{1i}, o_{2j}\rangle$ will be an element pair with serial position k.

Consider a set of $N = n_1n_2$ element pairs $\{a_i\}$ (indexed by serial position) to be processed, $i = 1, 2, \ldots, N$ and suppose that stage 1 occupies the time interval from the

start of processing up until the first pair completes processing, stage 2 occupies the time from the completion of the first pair to the completion of the second and so on. Let $A(k_r, r, O, \delta, \bar{t}_r)$ be the parameter set of a specified probability distribution within a serial model where k_r designates the serial position of the pair processed during stage r, $O = \langle a_{k_1}, a_{k_2}, \ldots, a_{k_N} \rangle$ is a vector denoting a particular order of processing the pairs, according to serial position (k_1 indexes the serial position of the element completed first), $\delta = +1$ if the element pair is a match and -1 if it is a mismatch, and $\bar{t}_r = \langle t_1, t_2, \ldots, t_{r-1} \rangle$, is a vector representing the successive intercompletion times and their order, up to stage r.

We then have the following definition of a serial model.

DEFINITION 7. A serial model for matching is a stochastic process specified by the following type of joint probability density on intercompletion times between the successive completions of pairs of elements.

$$f^{(s)}(t_1, t_2, \ldots, t_N, O) = P(O) \prod_{r=1}^{N} f(t_r \mid A(k_r, r, O, \delta, \bar{t}_r)),$$

where (s) refers to "serial," $A(k_r, r, O, \delta, \bar{t}_r)$ is as described above, $P(O)$ is the probability that the order of processing the pairs is that given in the argument, and t_r is the intercompletion time between pair $a_{k_{r-1}}$ and a_{k_r}. The term $f(t_r \mid A(k_r, r, O, \delta, \bar{t}_r))$ is a probability density conditioned via its parameter set on the order of the various serial positions, whether or not the pair in position k_r is composed of matching or mismatching stimuli, and the previous $r-1$ intercompletion times. By convention, $\bar{t}_1 = \langle 0 \rangle$; that is, the zeroth intercompletion time is zero. Notably, any term in the product

$$P(O) = P(a_{k_1})\, P(a_{k_2} \mid a_{k_1}) \cdots P(a_{k_N} \mid a_{k_1}, a_{k_2}, \ldots, a_{k_{N-1}})$$

say $P(a_{k_r} \mid a_{k_j}, \ldots, a_{k_{r-1}})$ is independent of whether $a_{k_j}, j > r$ is a matching $(+)$ or mismatching $(-)$ pair. Similarly, "A" at stage r is independent of whether any element completed later is a $+$ or $-$ comparison (cf. Definition 3).

In cases of particular serial models where the order is selected as processing occurs, rather than before processing begins, it simply happens that at a stage r, if say, pair a_i is processed, that its rate will be independent of what happens next. This in no way invalidates Definition 7. What it does mean is that the parameter set A at stage r will, in that case, be completely determined by $\bar{t}_r$ and by the first r-1 positions of O. It would also be natural, but no more correct in this instance where the order is determined as processing proceeds, to write

$$P(O) = P(a_{k_1})\, P(a_{k_2} \mid a_{k_1}) \cdots P(a_{k_N} \mid a_{k_1}, a_{k_2}, \ldots, a_{k_{N-1}}).$$

It can be seen from the constraints on the parameters (P, A) that a serial model is not just a general chain-rule stochastic definition of a matching trial. It can further

be seen that certain parallel or hybrid models would, in the case where element processing is not assumed to be an all-or-none event, specify a very different event space (or sigma field) for a trial. For instance, a time-sharing model with possible partial processing would not be associated with the same event space as the set of serial models.

The stochastic processes defined by Definition 7 are quite general. If two further conditions hold, processes satisfying Definition 7 form a two-dimensional Markov process with the state of the system on one dimension and t_r on the other. The first condition is that $A(k_r, r, O, \delta, \bar{t}_r) = A(k_r, r, O, \delta)$, that is, that succeeding inter-completion times are independent of the preceding ones. The second condition is that $f(t_r \mid A(k_r, r, O, \delta))$ be dependent only on the state of the system at time $t = \sum_{k=1}^{r-1} t_k$, in particular the order up to that point, and not the order of later completed elements. Further, it follows in this special case that the process formed by recording the state of the system across time, is a semi-Markov or Markov renewal process (see, e.g., Pyke, 1961a,b). The present paper is concerned with different matters but it may be interesting in the future to attempt application of semi-Markov machinery to serial and parallel systems.

Assume that Fig. 1 depicts a case where S_1 consists of two elements, (a) and (b), and S_2 consists of one element, say (c), and that (a) and (b) are processed by being matched against the S_2 element. Let (a) have serial position 1 and (b) serial position 2. Since S_2 has but one member, it is possible in this case to simply write the order in terms of the elements of S_1 and the serial density of Fig. 1b is

$$f^{(s)}(t_1, t_2 \mid \langle a, b \rangle) = P(\langle a, b \rangle) f(t_1 \mid A(1, 1, \langle a, b \rangle, \delta, \bar{t}_1)) f(t_2 \mid A(2, 2, \langle a, b \rangle, \delta, \bar{t}_2)).$$

Now, let $B(h_r, r, O(r), \delta, \bar{t}_r)$ play the parameter role for parallel models analogous to that of A in serial models. Therefore, $i = h_r, i_r, j_r$ will represent an indexing of element pairs that remain unfinished until at least the end of stage $r + 1$, where there are N-r such indices at stage r. The only difference in A and B is that B uses $O(r) = \langle a_{k_1}, a_{k_2}, \ldots, a_{k_{r-1}} \rangle$, that is, the order of processing through stage $r - 1$ whereas A is potentially dependent on $O = O(N)$. The term $\varnothing$ will be used to demarcate portions of the order vector not yet filled up. The index k_r will, as above, designate the element actually completed at stage r. The supercript (p) below will refer to parallel processing. The function F is just one minus the cumulative distribution function. So, for example, $F(t_2 \mid B(h_2, 2, O(2), -1, \bar{t}_2))$ is the probability that the pair in serial position h_2 has not yet been completed during stage 2 when t_2 time units of that stage have elapsed, given that $\bar{t}_2 = \langle t_1 \rangle$, that is that the time necessary for stage 1 was t_1 and h_2 refers to a mismatching pair. We could also equivalently write, for arbitrary stage r:

$$F(t_r \mid B(h_r, r, O(r), \delta, \bar{t}_r)) = \int_{t_r' = t_r}^{\infty} f(t_r' \mid B(h_r, r, O(r), \delta, \bar{t}_r))\, dt_r',$$

where f is again a density function.

DEFINITION 8. A parallel model for matching is a stochastic process specified by the following joint probability density on intercompletion times.

$$
\begin{aligned}
f^{(p)}(t_1, t_2, \ldots, t_N ; O) = {} & F(t_1 \mid B(h_1, 1, O(1), \delta, \bar{t}_1)) F(t_1 \mid B(i_1, 1, O(1), \delta, \bar{t}_1)) \\
& \cdots F(t_1 \mid B(j_1, 1, O(1), \delta, \bar{t}_1)) f(t_1 \mid B(k_1, 1, O(1), \delta, \bar{t}_1)) \\
& \cdots F(t_2 \mid B(h_2, 2, O(2), \delta, \bar{t}_2)) F(t_2 \mid B(i_2, 2, O(2), \delta, \bar{t}_2)) \\
& \cdots F(t_2 \mid B(j_2, 2, O(2), \delta, \bar{t}_2)) f(t_2 \mid B(k_2, 2, O(2), \delta, \bar{t}_2)) \\
& \cdots F(t_{N-1} \mid B(h_{N-1}, N-1, O(N-1), \delta, \bar{t}_{N-1})) \\
& \cdot f(t_{N-1} \mid B(k_{N-1}, N-1, O(N-1), \delta, \bar{t}_{N-1})) \\
& \cdot f(t_N \mid B(k_N, N, O(N), \delta, \bar{t}_N)),
\end{aligned}
$$

where the remaining notation is as in Definition 7.

The most striking difference between Definitions 7 and 8 is that in parallel processing, when say, pair k_1 finishes first at time t_1, the probability that each of the other pairs has not been completed during that time must be included. Consider as a special instance the parallel processing illustrated in Fig. 1. The parallel density function for this example, again assuming two elements (a), (b) in S_1 and only one in S_2, is

$$
\begin{aligned}
& f^{(p)}(t_1, t_2 ; \langle a, b \rangle) \\
& \quad = [F(t_1 \mid B(2, 1, O(1), \delta, \bar{t}_1)) f(t_1 \mid B(1, 1, O(1), \delta, \bar{t}_1))] f(t_2 \mid B(2, 2, O(2), \delta, \bar{t}_2)).
\end{aligned}
$$

The quantity in brackets is the probability expression for stage 1 and the $f(t_2 \mid \cdot)$ expression is the probability density function for element (b) during stage 2, after (a) has completed comparison against the element in S_2.

Observe too that the parallel models are capable of less a priori dependence on order of processing according to serial position, than the serial models, since the order "evolves" in a parallel system. But, in a general serial system that order may be viewed as selected beforehand and thus allow say, the parameter for the first pair processed to depend on which serial position finishes last. A parallel system (and hence a parallel model) cannot realistically do this, since that would make the model futuristic, in knowing what chance event will later occur (cf. Definition 4).

To illustrate this principle, an example requiring at least three element pairs is required. Let $S_1 = \{a_1, b_2, c_3\}$ and $S_2 = \{d\}$ where the subscripts in S_1 denote serial positions. Now consider the parallel probability density for the order of completion $\langle a_1, b_2, c_3 \rangle$ vs that for the order $\langle a_1, c_3, b_2 \rangle$:

$$
\begin{aligned}
& f^{(p)}(t_1, t_2, t_3 ; \langle a, b, c \rangle) \\
& \quad = [F(t_1 \mid B(2, 1, \langle \varnothing, \varnothing, \varnothing \rangle, \delta, \bar{t}_1)) \\
& \qquad \cdot F(t_1 \mid B(3, 1, \langle \varnothing, \varnothing, \varnothing \rangle, \delta, \bar{t}_1)) f(t_1 \mid B(1, 1, \langle \varnothing, \varnothing, \varnothing \rangle, \delta, \bar{t}_1))] \\
& \qquad \cdot [F(t_2 \mid B(3, 2, \langle a, \varnothing, \varnothing \rangle, \delta, \bar{t}_2)) f(t_2 \mid B(2, 2, \langle a, \varnothing, \varnothing \rangle, \delta, \bar{t}_2))] \\
& \qquad \cdot f(t_3 \mid B(3, 3, \langle a, b, \varnothing \rangle, \delta, \bar{t}_3)),
\end{aligned}
$$

and

$$f^{(p)}(t_1, t_2, t_3; \langle a, c, b\rangle)$$
$$= [F(t_1 \mid B(2, 1, \langle \varnothing, \varnothing, \varnothing\rangle, \delta, \bar{t}_1))\, F(t_1 \mid B(3, 1, \langle \varnothing, \varnothing, \varnothing\rangle, \delta, \bar{t}_1))$$
$$\cdot f(t_1 \mid B(1, 1, \langle \varnothing, \varnothing, \varnothing\rangle, \delta, \bar{t}_1))]\, [F(t_2 \mid B(2, 2, \langle a, \varnothing, \varnothing\rangle, \delta, \bar{t}_2))$$
$$\cdot f(t_2 \mid B(3, 2, \langle a, \varnothing, \varnothing\rangle, \delta, \bar{t}_2))]\, f(t_3 \mid B(2, 3, \langle a, c, \varnothing\rangle, \delta, \bar{t}_3)),$$

respectively.

The serial terms for the two orders are

$$f^{(s)}(t_1, t_2, t_3; \langle a, b, c\rangle)$$
$$= P(\langle a, b, c\rangle)\, f(t_1 \mid A(1, 1, \langle a, b, c\rangle, \delta, \bar{t}_1))$$
$$\cdot f(t_2 \mid A(2, 2, \langle a, b, c\rangle, \delta, \bar{t}_2))\, f(t_3 \mid A(3, 3, \langle a, b, c\rangle, \delta, \bar{t}_3))$$

and

$$f^{(s)}(t_1, t_2, t_3; \langle a, c, b\rangle)$$
$$= P(\langle a, c, b\rangle)\, f(t_1 \mid A(1, 1, \langle a, c, b\rangle, \delta, \bar{t}_1))$$
$$\cdot f(t_2 \mid A(3, 2, \langle a, c, b\rangle, \delta, \bar{t}_2))\, f(t_3 \mid A(2, 3, \langle a, c, b\rangle, \delta, \bar{t}_3)),$$

respectively.

The main difference in the serial vs the parallel terms is that in the serial terms, there is explicit dependence on the order in the density at the very first stage and its alternate with the other order. In contrast, the terms of the first stage of the parallel model do not express such dependence, that is, the joint probability term for stage one is identical for the two orders.

It is important to note that this does not mean that the parallel density of the first stage, *when conditioned on order*, cannot depend on that order. But the dependence on order has to come from the dependence on the successive intercompletion times. Thus, since the time to complete (a) may influence whether (b) or (c) is more likely to finish next, the likelihood that (a) is completed at time t_1 may depend in a Bayesian fashion on whether (b) or (c) finished second. In models that assume independent successive intercompletion times, the order-conditional densities are independent of order. With one exception, all the models in the present paper are based on independent successive intercompletion times. Thus, even the conditional densities of the present paper are order-independent. This is not an immutable property of models developed from exponential intercompletion times, as is well known in the study of continuous Markov processes. The exception mentioned above is that the serial models covered in Theorem 5 are essentially distribution free.

An additional characteristic of the parallel models as defined here is the independence of the processing on the various elements during a single stage. For example $P((a)$ *is* 1st at time $t_1) = P((a)$ takes time t_1 and (b) takes longer$) = F(t_1 \mid B(2, 1, O(1), \delta, \bar{t}_1))$ $f(t_1 \mid B(1, 1, O(1), \delta, \bar{t}_1))$ reflects this within-stage independence. This is not a necessary property of parallel models by any means, and could be altered with an increment

in complexity of notation. Processes that involve exponential intercompletion times (*not* necessarily completion times) possess this property of within stage independence, and are the ones on which we concentrate in this paper.

In summary, then, the basic property of a within-stage independent parallel model is that the joint probability density on intercompletion times is given by a product-series of products, each component product being a measure on the minimum waiting time of a group of separate processes. This is shown in Definition 8 by the appearance of parameters of all uncompleted pairs at each of stages 1, 2,..., N.

Since parallelity and seriality are quite obviously very special types of processing, it is most convenient to define as "hybrid" all other models, or systems. For consistency, Definition 9 is written at the level of stochastic processes that possess densities but extension to more general processes is possible.

DEFINITION 9. A hybrid model for matching is a stochastic process that possesses a probability density on the joint intercompletion times of a set of pairs but that does not satisfy Definitions 6 or 7.

Since models have been equated, for our purposes, with stochastic processes, it is natural to define model equivalence in terms of identicality of the stochastic processes.[3] However, it is obvious that two models may be equivalent for one type of trial or experiment (i.e., context) but not so for another, so the definitions below are stated with regard to context.

DEFINITION 10. Two models, M_1, M_2, are totally equivalent, $M_1 \equiv M_2$ with respect to context C, if and only if they are defined by equivalent stochastic processes within context C.

DEFINITION 11. Two models M_1, M_2 are x-diverse with respect to context C, if and only if there exists some aspect x associated with their respective stochastic processes that is different for M_1 and M_2.

The aspect x is purposely left unspecified in Definition 10; some examples of x are (1) a measurable function on the underlying event space, (2) one or more moments, (3) a set of possible realizations of the processes, (4) the distribution itself. Number 4 will be especially useful and will be termed "distribution diverse." Number 4 is the weakest type of diversity and simply means the two stochastic processes are not totally equivalent. To be distribution diverse does not, of course, imply that there is no stronger diversity.

DEFINITION 12. Two models M_1, M_2 are x-testable with respect to context C, if and only if they are x-diverse and x can be observed or can be estimated within context C.

[3] For a reasonably general statement on equivalence of stochastic processes, the reader is referred to Loève (1955, p. 499).

In some contexts it may be useful to distinguish between cases where x can be actually observed (for example, a sequence of responses in a learning experiment) and those where x is a statistic with a sampling distribution (for example, mean reaction time). This distinction will not be critical for discussion in the present study.

The definitions of parallel and serial *model* equivalence, diversity and testability are directly implied by Definitions 10 to 12 and are formally omitted. The definitions of parallel and serial *systems* are centered on simultaneous processing of all elements versus one-at-a-time processing. It is important to keep the parallel–serial question separate from other related questions such as limited vs unlimited capacity systems, and independent vs dependent processing systems (see especially, Gardner, 1973; Townsend, 1974). These aspects of processing as well as others (e.g., self-termination vs exhaustiveness of processing) are logically independent of the parallel–serial issue, although certain combinations may seem more plausible in certain experimental contexts.

3. Models Based on Exponential Intercompletion Times

Among the probability distributions so far examined, those involving exponential intercompletion times (times between two successive element processing completions) appear to offer the most severe problems of equivalence (Townsend, 1971b, 1972). Certain models defined within rather large classes of exponential-intercompletion-time processes are totally equivalent. On the other hand it has been shown that there are other models in these classes that are not totally equivalent. Investigation of exponential processing appears to provide a conservative view of the possibilities for testing serial against parallel processing. This fact may be of importance in psychological experimentation, since useful information pertaining to specific characteristics of distributions (e.g., higher moments) is often hard to come by. If parallel processing is distinguishable in principle from serial processing when exponential processing is assumed, the two will probably be distinguished for other, more complex distributions. We proceed to develop a theory of serial and parallel processing with exponential intercompletion times for matching paradigms.

Definition 13. A serial-exponential model (SEM) for matching experiments is a processing model which is characterized by (1) a probability distribution over the processing order of pairs of elements to be compared on a given trial, and (2) a set of exponential processing rates for the different pairs which may, in turn, depend on whether a pair matches (+) or mismatches (−), and on the entire processing order specified in terms of the serial position index of each pair in $S_1 \times S_2$.

As noted earlier, a "serial position index" may be assigned to each of the pairs in $S_1 \times S_2$ that preserves the information of the serial positions of the individual elements

and hence can be of potential use to a serial processor (or ourselves as classifiers) in assigning processing orders on each trial. It will be convenient in the following to assign an index to each of the $N!$ processing orders, namely the numerals 1, 2,..., $N!$ and to employ this index in parameter descriptions instead of the order itself (O).

The parameter specification is then implied for the serial model when n_1 elements in S_1 are compared with n_2 elements in S_2 ; there being $n_1 \cdot n_2 = N$ potential comparisons available for processing.

DEFINITION 14. The serial sets of parameters for matching experiments are given by

(1) $\{P(K)\}$, where $K = 1, N!, \sum_{K=1}^{N!} P(K) = 1$, and $\{P(K)\}$ is the set of probabilities on the different processing orders, and K is the index defined above;

(2) $\{\lambda(k_r, r, K, \delta)\}$, where $\{\lambda(\cdot)\}$ is the set of exponential processing rates ($\lambda \geqslant 0$), which depend on (i) $\delta = +1$ or -1, i.e., whether the pair consists of matching or mismatching elements and, (ii) K, i.e., the processing order given in terms of the serial positions with k_r denoting the serial position of the element processed at stage r.

As an example of serial parameters, $P(3)$ would be the probability that the third out of the $N!$ possible processing orders is taken. The parameter $\lambda(6, 5, 20, -1)$ specifies a processing rate for the stimulus pair with serial position index 6, processed 5th in the 20th order out of the possible $N!$ orders and they mismatch. There are $N! - 1$ unconstrained processing order probabilities and $N(N)!$ processing rate parameters for each of the two match possibilities, $\delta = \pm 1$. This formulation therefore results in a total of $N! - 1 + 2N(N!)$ parameters under nondegenerate conditions for the serial model. Although K is obviously sufficient to specify r, the inclusion of r in $\lambda(\cdot)$ will facilitate later theoretical remarks.

In general parallel processing models, we allow for the possibility that as various stimulus objects are completed, the rates of the remaining stimuli may change. Recall that the term "stage" refers to how many comparisons have been completed; thus, stage 1 means no comparisons have been finished, stage 2 that one comparison has been completed and so on.

DEFINITION 15. A parallel-exponential model (PEM) for matching experiments is a parallel processing model characterized by a set of exponential rates that can depend on specific comparisons (+ or − match) and on the past order of completions up to any specified stage on a given trial.

Just as in the serial model, the rates of the parallel model can depend on processing order. However, unlike the serial model, the parallel parameters cannot depend on that part of the order that comes at later stages. As noted above, this follows from a parallel model being basically evolutionary with the probability of a particular order depending on the relative magnitudes of the processing rates at the successive stages

while a serial model's processing rates on a particular pair can be selected with regard to what happens after this pair is processed as well. This historical constraint on parallel processing is illustrated in the next definition.

DEFINITION 16. The parallel set of parameters is given by $\{\nu(h_r, r, m(r), \delta)\}$, where h_r denotes the serial position of a pair in $S_1 \times S_2$ that has not completed processing by the beginning of stage r; r denotes the processing stage, $1 \leqslant r \leqslant N$; $m(r)$ is an ordered list of the serial positions of the pairs completed in stages 1 through stage $r - 1$, and δ is as before.

As an example $\nu(7, 2, \langle 5\rangle, -1)$ is the processing rate for the pair with serial position 7 at stage 2 when the pair with serial position 5 was completed in stage 1 and the pair 7 mismatches. That is $m(2) = \langle 5\rangle$ since the complete processing history prior to stage 2 consists of the pair processed in stage 1 which is indexed by serial position 5 for our example.

By convention we let $m(1) = \{\varnothing\}$ to denote that prior to stage 1 there is only one (null) processing history. Then $m(1)$ has N possible members, $m(2)$ has $N(N-1)$ possible members, and $m(r)$ has $N(N-1)\cdots(N-r+1)$ possible members. At stage r there are $N - r + 1$ unprocessed elements and these elements can take any of $N(N-1)\cdots(N-r+1)$ processing rates depending on the serial positions of the pairs already processed, their order, as well as whether the unfinished pairs consist of matching or mismatching elements.

Note that although an enumeration of the total set of possible processing orders K was employed in the serial parameter definition, an ordering of the element pairs by serial position, $m(r)$, is again used in the parallel definition. This is because it is an aid in some of the proofs below to retain the order itself in the parallel case. In fact, it would be technically correct to write $m(r) = O(r)$ where O is as before, but a specialization of notation is necessary later that is best kept separate from the earlier more general case.

It may be seen that the total possible number of parameters for the parallel model is $2N!\sum_{j=0}^{N-1}(1/j!)$. This is a compact way of writing $2[(N) + (N)(N-1) + \cdots + (N)(N-1)\cdots 2 + (N)(N-1)\cdots 2\cdot 1]$. The "2" simply comes from the doubling of the number of rate parameters due to the $+$, $-$ match distinction. Inside the brackets, at stage 1 there are N parameters for each of the N pairs. At stage 2, for each of the possible N pairs there are $N - 1$ different possible pairs that could have been completed at stage 1 and hence there is a set of parameters at stage 2 numbering $N(N-1)$. This continues: The set of possible histories increases as the stage number gets larger up until at the last stage $(N-1)!$ histories are possible for each of the N potential last elements giving $N!$ parameters at the last stage. Note that order of the completions in the histories is retained as a possible source of rate differences. Note again that the N serial rate parameters can depend on the entire set of $N!$ orders as contrasted with the parallel model's historical constraints.

We now are in a position to derive a unique description of the stochastic processes that result from Definitions 13–16. The device we shall employ exhibits less complexity of notation than would specification of the density functions. The method is that of joint characteristic functions.[4] The "true" characteristic function for the present models would involve summing up our joint characteristic functions. The set of joint characteristic functions is actually a more precise description of the stochastic process than is the sum of the members of this set, since each joint characteristic function specifies a certain order of processing, and together the whole set uniquely determines the stochastic process. Thus, our joint characteristic functions involve a Fourier transform of the time aspect of the process but also include the transformed probability measure on the particular processing order.

It is necessary at this point to develop a slightly more detailed notation for the vector of pairs associated with K in order to indicate individual pairs and their completion position. We let $k_r(K)$ be the serial position of the pair that completes processing *rth* when the total order is given by $K = \langle k_1(K), k_2(K),..., k_N(K)\rangle$. Thus, when useful we may, for example, write $P(K) = P(k_1(K), k_2(K),..., k_{N-1}(K), k_N(K))$.

The terms $T(k_r, r, K, \delta)$, $T(k_r, r, m(r), \delta)$ refer to the serial and parallel intercompletion time random variables from the $(r-1)$th to the rth pair, respectively. It is to be understood that (k, r, K) here corresponds in a 1–1 fashion to serial position k_r in the order K. The terms $s(k_r, r, K, \delta)$ and $s(k_r, r, m(r), \delta)$ are the respective Fourier transform variables of the previous two terms. Since the first two $(T(k_r, r, K, \delta), T(k_r, r, m(r), \delta))$ refer to the same variable in the time domain and the latter two $(s(k_r, r, K, \delta), s(k_r, r, m(r), \delta))$ refer to the same variable in the transform domain we see that it is possible to set

$$T(k_r, r, K, \delta) = T(k_r, r, m(r), \delta),$$
$$s(k_r, r, K, \delta) = s(k_r, r, m(r), \delta),$$

when equivalence mappings are established between serial and parallel models. It will facilitate exposition to use the same "time" (t or T) and dummy variable (s) terms for the parallel and serial expressions.

In order to further simplify the still complex notation, we will employ the convention of writing $t(k_r, r, K, \delta) = t_r$, $s(k_r, r, K, \delta) = s_r$, under the integral sign, but the full notation, (k_r, r, K), is always implied, and can be immediately retrieved by way of the other terms in the characteristic function.

With the foregoing considerations in mind, we can write the appropriate joint characteristic function for the serial model; this is the joint expectation of the terms

[4] Where the moment generating function is defined, as is the case here with exponential intercompletion times, it serves as well as the characteristic function. In fact, by dropping the "$i = (-1)^{1/2}$" in the characteristic function, the moment generating function is obtained in the following characteristic function expressions.

$\exp(is_rT_r)$, where $i = (-1)^{1/2}$, $r = 1, 2, \ldots, N$ and the probability function includes the probability for the order of processing. It will shorten the exponential terms if we use $e(x)$ rather than the usual $\exp(x)$ to stand for e^x.

LEMMA 1. *If a model of a processing system in a matching experiment is serial-exponential, then the joint characteristic function* $\phi^{(s)}$ (s = *serial*) *for a specified order* K, *of the intercompletion times, is given by*

$$\phi^{(s)}(T(k_1, 1, K, \delta), T(k_2, 2, K, \delta), \ldots, T(k_N, N, K, \delta))$$
$$= P(K) \cdot \frac{\lambda(k_1, 1, K, \delta)}{\lambda(k_1, 1, K, \delta) - is_1} \cdot \frac{\lambda(k_2, 2, K, \delta)}{\lambda(k_2, 2, K, \delta) - is_2} \cdots \frac{\lambda(k_N, N, K, \delta)}{\lambda(k_N, N, K, \delta) - is_N}.$$

Proof. Following directly from the above definitions and that of joint characteristic functions we have

$$\phi^{(s)} = E^{(s)}[e(is(k_1, 1, K, \delta)\, T(k_1, 1, K, \delta))\, e(is(k_2, 2, K, \delta)\, T(k_2, 2, K, \delta)) \cdot \cdots$$
$$\cdot e(is(k_N, N, K, \delta)\, T(k_N, N, K, \delta))]$$
$$= P(K) \int_0^\infty \int_0^\infty \cdots \int_0^\infty \{[\lambda(k_1, 1, K, \delta)\, e(-\lambda(k_1, 1, K, \delta)\, t_1)\, e(is_1t_1)]\}$$
$$\cdot [\lambda(k_2, 2, K, \delta)\, e(-\lambda(k_2, 2, K, \delta)\, t_2)\, e(is_2t_2)] \cdot \cdots$$
$$\cdot [\lambda(k_N, N, K, \delta)\, e(-\lambda(k_N, N, K, \delta)\, e(is_Nt_N)]\, dt_1\, dt_2 \cdots dt_N.$$

Clearly, the separate terms involving the t_i may be segregated into the product of the integrals,

$$\phi^{(s)} = P(K) \int_0^\infty \lambda(k_1, 1, K, \delta)\, e(-\lambda(k_1, 1, K, \delta)\, t_1)\, e(is_1t_1)\, dt_1$$
$$\cdot \int_0^\infty \lambda(k_2, 2, K, \delta)\, e(-\lambda(k_2, 2, K, \delta)\, t_2)\, e(is_2t_2)\, dt_2 \cdot \cdots$$
$$\cdot \int_0^\infty \lambda(k_N, N, K, \delta)\, e(-\lambda(k_N, N, K, \delta)\, t_N)\, e(is_Nt_N)\, dt_N,$$

where $\lambda(k_r, r, K, \delta)\, e(-\lambda(k_r, r, K, \delta)\, t_r)\, dt_r$ ($r = 1, \ldots, N$) is the density function for the rth intercompletion time, and these may be integrated to yield the proposed result. Q.E.D.

The next lemma gives an analogous result for the parallel models.

LEMMA 2. *If a model of an element processing system in a matching experiment is parallel-exponential then the joint characteristic function for a specified history $m(N+1)$ of the intercompletion times, $\phi^{(p)}$ (p = parallel) is given by*

$$\phi^{(p)}(T(k_1, 1, m(1), \delta), T(k_2, 2, m(2), \delta), \ldots, T(k_N, N, m(N), \delta))$$

$$= \frac{\nu(k_1, 1, m(1), \delta)}{\sum_{j=1}^{N} \nu(h_{1_j}, 1, m(1), \delta) - is_1}$$

$$\cdot \frac{\nu(k_2, 2, m(2), \delta)}{\sum_{j=2}^{N} \nu(h_{2_j}, 2, m(2), \delta) - is_2} \cdots \frac{\nu(h_{N.N}, N, m(N), \delta)}{\nu(h_{N.N}, N, m(N), \delta) - is_N},$$

where h_{rj} denotes the serial position of one of the $N - r + 1$ elements as yet uncompleted at the beginning of stage r, and of course, $h_{N.N} = k_N$.

Proof. The density for each intercompletion time is slightly more complicated here. This quantity is, for the rth to be finished,

$$\nu(k_r, r, m(r), \delta)\, e(-\nu(k_r, r, m(r), \delta)\, t_r) \cdot e\left(-t_r \sum_{j=r+1}^{N} \nu(h_{rj}, r, m(r), \delta)\right) dt_r;$$

the first two terms are the typical exponential probability density elements and the third is the probability that none of the other (remaining) stimulus pairs completes processing during the interval t_r, which begins at time $t = t_1 + t_2 \cdots + t_{r-1}$.

The expectation of the product of all such appropriate terms is

$$\phi^{(p)} = E^{(p)}[e(is(k_1, 1, m(1), \delta)\, T(k_1, 1, m(1), \delta))\, e(is(k_2, 2, m(2), \delta)\, T(k_2, 2, m(2), \delta)) \cdot \cdots$$

$$\cdot e(is(k_N, N, m(N), \delta)\, T(k_N, N, m(N), \delta))]$$

$$= \int_0^\infty \int_0^\infty \cdots \int_0^\infty \left\{\left[\nu(k_1, 1, m(1), \delta)\, e(-\nu(k_1, 1, m(1), \delta)\, t_1)\right.\right.$$

$$\cdot e\left(-t_1 \sum_{j=2}^{N} \nu(h_{1_j}, 1, m(1), \delta)\right) e(is_1 t_1)\Big]$$

$$\cdot \left[\nu(k_2, 2, m(2), \delta)\, e(-\nu(k_2, 2, m(2), \delta)\, t_2)\, e\left(-t_2 \sum_{j=3}^{N} \nu(h_{2_j}, 2, m(2), \delta)\right)\right.$$

$$\left.\left.\cdot e(is_2 t_2)\right] \cdots [\nu(k_N, N, m(N), \delta)\, e(-\nu(k_N, N, m(N), \delta)\, t_N)\, e(is_N t_N))]\right\} dt_1\, dt_2 \cdots dt_N.$$

Again, we may segregate the terms for the t_j's producing a product of integrals

$$\phi^{(p)} = \int_0^\infty \nu(k_1, 1, m(1), \delta)\, e(-\nu(k_1, 1, m(1), \delta)\, t_1)\, e\left(-t_1 \sum_{j=2}^{N} \nu(h_{1_j}, 1, m(1), \delta)\right)$$
$$\cdot\, e(is_1 t_1)\, dt_1 \int_0^\infty \nu(k_2, 2, m(2), \delta)\, e(-\nu(k_2, 2, m(2), \delta)\, t_2)$$
$$\cdot\, e\left(-t_2 \sum_{j=3}^{N} \nu(h_{2_j}, 2, m(2), \delta)\right) e(is_2 t_2)\, dt_2 \cdot \cdots$$
$$\cdot \int_0^\infty \nu(k_N, N, m(N), \delta)\, e(-\nu(k_N, N, m(N), \delta)\, t_N)\, e(is_N t_N)\, dt_N .$$

Integration of these obtain the joint characteristic function indicated in the statement of the lemma. Q.E.D.

The two preceding lemmas exhibit the basic canonical forms of the pertinent characteristic functions. They can also be proven employing more explicitly the fact that the succeeding intercompletion times are independent in the present models. That fact means that the joint characteristic function can be written as the product of the joint characteristic functions of the separate intercompletion times and this leads to the above results. Lemmas 1, 2 were proven more directly in order to expose the structure of the probability functions and their relationships. We should observe that for any fixed experimental arrangement of duplicated stimuli in S_1 and S_2 there will be $N!$ different such joint characteristic functions corresponding to the various $N!$ completion orders. Alternation of the stimulus serial position locations of members of identical pairs in S_1 and/or S_2 defines a new set of $N!$ characteristic functions. This new set of characteristic functions may differ from the old since matching and mismatching pairs, upon which the rates depend, may be located in new serial positions.

In the example of Fig. 1, the serial joint characteristic function is thus found to be

$$\phi^{(s)}(T(k_1, 1, \langle a, b\rangle, \delta),\ T(k_2, 2, \langle a, b\rangle, \delta))$$
$$= P(\langle a, b\rangle)\, \frac{\lambda(k_1, 1, \langle a, b\rangle, \delta)}{\lambda(k_1, 1, \langle a, b\rangle, \delta) - is_1} \cdot \frac{\lambda(k_2, 2, \langle a, b\rangle, \delta)}{\lambda(k_2, 2, \langle a, b\rangle, \delta) - is_2}$$

and that for the corresponding parallel case is

$$\phi^{(p)}(T(k_1, 1, \{\varnothing\}, \delta),\ T(k_2, 2, a, \delta))$$
$$= \frac{\nu(k_1, 1, \{\varnothing\}, \delta)}{\nu(k_1, 1, \{\varnothing\}, \delta) + \nu(h_{12}, 1, \{\varnothing\}, \delta) - is_1} \cdot \frac{\nu(k_2, 2, a, \delta)}{\nu(k_2, 2, a, \delta) - is_2},$$

where, of course, $h_{12} = k_2$ specifying the same pair of elements. Note that $m(1) = \{\varnothing\}$ since nothing has been processed at stage 1 and hence no history is yet available, but

at stage 2 in this example, $m(2) = k_1$ since the pair specified by serial position k_1 was completed at the end of stage 1. Thus, at stage 2 the rate for the element pair $a_{k_2} = a_{h_{12}}$ may be changed from $\nu(k_2, 1, \{\varnothing\}, \delta)$ to $\nu(k_2, 2, a, \delta)$ and they are not necessarily equal. The parallel and serial joint characteristic functions for the other orders can be found in a similar manner.

4. Theorems on Equivalence and Diversity

Equivalence will be investigated for a simple matching type of experiment (Definition 2), in particular, one that includes two types of trials, the first being with no S_2 stimulus elements identical to any S_1 stimulus elements (trial type T_1) and the second being with exactly one of the S_2 stimulus elements identical to one of the S_1 stimulus objects (trial type T_2).

Theorem 1 considers the context of T_1 trials, the results are equivalent to those found from considering (only) trials where all the stimulus pairs match. Theorem 1A states sufficient conditions to ensure that *total equivalence* does *not* hold, that is, that at least *distribution diversity does* hold. Basically, Theorem 1A says that the serial model's rate parameters can differ in a way depending on order that the parallel model's rate parameters cannot, in order for the parallel and serial models to be different. Theorem 1A works as long as either $n_1 > 1$, $n_2 > 1$ or both.

The second part of Theorem 1 (1B) states that the negation of the condition in 1A implies total equivalence. Thus, the condition is necessary and sufficient for total equivalence not to hold.

Theorem 1A. *On T_1 trials if $P(K) \neq 0$ for all K in* SEM *then the serial exponential model* (SEM) *and the parallel exponential model* (PEM) *are distribution diverse and therefore not totally equivalent if* (1) *there exist at least two orders K_1, K_2 (or $m_1(N+1)$, $m_2(N+1)$) in which* (2) *there is at least one stage r such that the orderings (or history) up until r are the same in K_1, K_2 and such that $\lambda(k_r^{(1)}, r, K_1, -1) \neq \lambda(k_r^{(2)}, r, K_2, -1)$, where $k_r^{(1)}$ and $k_r^{(2)}$ may be the same but are not necessarily, and are not processed before r in K_1, K_2, but each is assumed to be completed at the end of stage r.*

1B. *If for all K_i, K_j; for all r, and for all $k_r^{(i)}$, $k_r^{(j)}$, $\lambda(k_r^{(i)}, r, K_i, -1) = \lambda(k_r^{(j)}, r, K_j, -1)$, where K_i, K_j are the same up until stage r, then* SEM *and* PEM *are totally equivalent.*

Proof. A. Assume there exist K_1, K_2, and r such that $\lambda(k_r^{(1)}, r, K_1, -1) \neq \lambda(k_r^{(2)}, r, K_2, -1)$; then note that at stage r the PEM survivor function (1 minus the cumulative distribution function) is given for K_1, K_2, for the intercompletion time from $r-1$ to r, conditionalized on order, by $F_{\text{PEM}}(t_r, K_i) = e(-t_r \sum_{j=r+1}^{N} \nu(h_{rj}, j, m_i(r), -1))$, $i = 1, 2$, and hence $F_{\text{PEM}}(t_r, K_1) = F_{\text{PEM}}(t_r, K_2)$ by virtue of $m_1(r) = m_2(r)$ and the historical constraint on parallel processing, irrespective of the

specific pair indexed by serial positions $k_r^{(1)}, k_r^{(2)}$. Therefore it holds for the pairs indexed by $k_r^{(1)}$ and $k_r^{(2)}$. On the other hand, the corresponding quantities for SEM are given by $F_{\text{SEM}}(t_r, K_1) = e(-\lambda(k_r^{(1)}, r, K_1, -1)\, t_r) \neq F_{\text{SEM}}(t_r, K_2) = e(-\lambda(k_r^{(2)}, r, K_2, -1)\, t_r)$ by hypothesis that the λ's are unequal. Thus, even if a mapping is otherwise possible, the unequal λ hypothesis precludes total equivalence and yields distribution diversity.

Proof. B. Part B is much more complex since a mapping must be given between the parameter spaces of PEM and SEM that yields total equivalence. Assume that hypothesis B is true, that is, for all K_i, K_j, and for all r, such that $m_i(r) \equiv m_j(r)$ up until stage r, $\lambda(k_r^{(i)}, r, K_i, -1) = \lambda(k_r^{(j)}, r, K_j, -1)$. Then there exists a homeomorphic mapping G, carrying the space of PEM parameters onto the space of SEM parameters and vice versa. It is well known that this will guarantee equivalent probability measures for the related processes.

First consider the two pertinent parameter spaces. Each space includes a number of orthogonal dimensions equal to the number of independent parameters. In the case of PEM, there are thus $N! \sum_{j=0}^{N-1} (1/j!)$ dimensions. SEM now, (under our hypothetical constraint) also includes the same number: $N! - 1$ probabilities, and one λ for stage 1, N for stage 2, $N(N-1)$ for stage 3, and so on down to $N(N-1)\cdots 2$ λ's at stage $m(N)$. This results in a total of $N! - 1 + [\sum_{j=0}^{N-2} \prod_{i=0}^{j} (N-i) + 1] = \sum_{j=0}^{N-1} \prod_{i=0}^{j} (N-i)$ which can be written as $N! \sum_{j=0}^{N-1} (1/j!)$ which is the same number of parameters as that for PEM. The serial probabilities $\{\mathbf{P}\}$ are, of course, constrained to the open interval (0, 1) by hypothesis.

Now let G be expressed as the proposed PEM $\rightarrow$ SEM mapping: For any K, set

$$\begin{aligned} P(K) &= P[k_1(K), k_2(K), \ldots, k_{N-1}(K), k_N(K)] \\ &= P[k_1(K)] \cdot P[k_2(K) \mid k_1(K)] \cdots P[k_N(K) \mid k_1(K), k_2(K), \ldots, k_{N-1}(K)] \end{aligned}$$

and let

$$P[k_1(K)] = \nu(k_1, 1, m(1), -1) \Big/ \sum_{j=1}^{N} \nu(h_{1j}, 1, m(1), -1),$$

$$P[k_2(K) \mid k_1(K)] = \nu(k_2, 2, m(2), -1) \Big/ \sum_{j=2}^{N} \nu(h_{2j}, 2, m(2), -1),$$

$$\vdots$$

$$P[k_{N-1}(K) \mid k_1(K), \ldots, k_{N-2}(K)] = \frac{\nu(k_{N-1}, N-1, m(N-1), -1)}{\nu(h_{N-1.N-1}, N-1, m(N-1), -1) + \nu(h_{N-1.N}, N-1, m(N-1), -1)},$$

$$P[k_N(K) \mid k_1(K), \ldots, k_{N-1}(K)] = \frac{\nu(k_N, N, m(N), -1)}{\nu(k_N, N, m(N), -1)} = 1;$$

where, of course, $h_{ii} = k_i$. Further, set

$$\lambda(k_1, 1, K, -1)$$
$$= \sum_{j=1}^{N} \nu(h_{1j}, 1, m(1), -1),$$

$$\lambda(k_2, 2, K, -1)$$
$$= \sum_{j=2}^{N} \nu(h_{2j}, 2, m(2), -1),$$
$$\vdots$$
$$\lambda(k_{N-1}, N-1, K, -1)$$
$$= \nu(h_{N-1,N-1}, N-1, m(N-1), -1) + \nu(h_{N-1,N}, N-1, m(N-1), -1),$$
$$\lambda(k_N, N, K, -1)$$
$$= \nu(h_{N,N}, N, m(N), -1),$$

for all k_r processed at stages $r = 1, N$.

Performing this for every K yields the PEM $\rightarrow$ SEM mapping. It may be ascertained that this mapping produces "P's" that sum to 1 by noting that all the denominators for each total joint probability giving the K's are equal; factoring these out and adding the numerators over the K's yields the product of sums of ν's found in the denominator.

At each stage r, there are $N(N-1)\cdots(N-r+2)$ histories $m(r)$, which we shall designate $m(r,1), m(r,2),\ldots, m(r, \prod_{i=0}^{r-2}(N-i))$. For each one of these histories, there is a set of $N-r+1$ pairs yet to be processed which we call $h(r,1), h(r,2),\ldots, h(r, N-r+1)$. Corresponding to these are the probabilities $P[h(r,1) \mid m(r,j)]$, $P[h(r,2) \mid m(r,j)],\ldots, P[h(r, N-r+1), \mid m(r,j)]$ for each $j = 1, 2,\ldots, \prod_{i=0}^{r-2}(N-i)$, for $r > 1$; $j = 1$ for $r = 1$. These probabilities correspond, of course, to a particular sequence of pairs so that the product

$$P[h(1,i_1) \mid m(1,j_1)]\, P[h(2,i_2) \mid m(2,j_2)] \cdots P[h(N,i_N) \mid m(N,j_N)] = P(K).$$

What we have just done is to show how the (serial) probability of a given order K, can be decomposed into a product of terms written via the historical sequence $m(1,j_1)$, $m(2,j_2),\ldots, m(N,j_N)$. The term $P(h(r,i_r) \mid m(r,j))$ gives the probability that the pair with serial position $h(r,i_r)$ is processed at stage r, conditioned on the specific previous history $m(r,j)$. This writing of serial probabilities in terms of some parallel structured terms, the $m(r,j)$, is convenient since the specification of parallel parameters depended

on the developing history from the start and secondly, it allows a natural language of serial and parallel expressions in G^{-1}, the prototypical SEM $\rightarrow$ PEM *mapping*:

$$\nu(h(1, 1), m(1, 1), -1)$$
$$= P(h(1, 1) \mid m(1, 1))\, \lambda(h(1, 1), m(1, 1), -1),$$
$$\nu(h(1, 2), m(1, 1), -1)$$
$$= P(h(1, 2) \mid m(1, 1))\, \lambda(h(1, 1), m(1, 1), -1),$$
$$\vdots$$
$$\nu(h(1, N), m(1, 1), -1)$$
$$= P(h(1, N) \mid m(1, 1))\, \lambda(h(1, 1), m(1, 1), -1),$$
$$\vdots$$
$$\nu(h(r, 1), m(r, 1), -1)$$
$$= P(h(r, 1)\, m(r, 1))\, \lambda(h(r, 1), m(r, 1), -1),$$
$$\vdots$$
$$\nu(h(r, N - r + 1), m(r, 1), -1)$$
$$= P(h(r, N - r + 1) \mid m(r, 1))\, \lambda(h(r, 1), m(r, 1), -1),$$
$$\vdots$$
$$\nu\left(h(r, N - r + 1), m\left(r, \prod_{i=0}^{r-2} (N - i)\right), -1\right)$$
$$= P\left(h(r, N - r + 1) \,\middle|\, m\left(r, \prod_{0}^{r-2} (N - i)\right)\right)$$
$$\cdot \lambda\left(h(r, 1), m\left(r, \prod_{0}^{r-2} (N - i)\right), -1\right),$$
$$\vdots$$
$$\nu\left(h(N, 1), m\left(N, \prod_{i=0}^{N-2} (N - i)\right), -1\right)$$
$$= P\left(h(N, 1) \,\middle|\, m\left(N, \prod_{0}^{N-2} (N - i)\right), -1\right) \lambda\left(h(N, 1) \,\middle|\, m\left(N, \prod_{0}^{N-2} (N - i)\right), -1\right).$$

Note that we have supposed, say,

$$\lambda\left(h(j, N - j + 1), m\left(j, \prod_{0}^{j-2} (n - i)\right), -1\right) = \lambda\left(h(j, 1), m\left(j, \prod_{0}^{j-2} (n - i)\right), -1\right)$$

as is required by hypothesis. The stage was explicitly given by the argument of h, m, etc., so was otherwise omitted from its usual position in ν, λ.

Next let $V_\sigma \in \Sigma$, $V_\pi \in \Sigma$ be the vectors with $N! \sum_{j=0}^{N-1} (1/j!)$ places which are contained in the serial (Σ) or parallel (Π) parameter spaces, respectively. The assumption that the ν's are in $(0, +\infty)$, the p's in $(0, 1)$, λ's in $(0, +\infty)$, establishes the nature of the dimensions. The parallel space given by $\Pi = \times\nu$ and the serial by $\Sigma = (\times p) \times (\times\lambda)$ are of the same dimension. Thus, a homeomorphism is at least conceivable, it not being possible to provide homeomorphisms between arbitrary, or even Euclidian, spaces of arbitrary dimensions (Hurewicz & Wallman, 1948). We now prove that G is a homeomorphism.

(a) G is $1 - 1$. Observation of G and G^{-1} will indicate that each parameter vector is a single valued vector function of the other for both G and G^{-1} and therefore G is $1 - 1$.

(*b*) Next it is obvious that for every point in the spaces Π, Σ there is a corresponding point in the other space via G; therefore G is onto. This leaves the proposition that G is bicontinuous to be proved.

(c) This proposition will be proved if it can be shown that each coordinate in Π and Σ is a continuous function of the coordinates in the other space, since such a relationship implies overall continuity. But this follows immediately from the form of these functions, i.e., these functions are of the form

$$\nu = p\lambda,$$
$$p = \nu/\Sigma\nu_i,$$
$$\lambda = \Sigma\nu_i,$$

and are undoubtedly continuous, with respect to the Euclidean metric, for example, in the (open) domains of the respective arguments. Q.E.D.

We will withhold discussion of most implications of this theorem, until the other type of trial is considered and the question of equivalence within the context of the whole simple matching paradigm can be treated. However, it may be observed that Theorem 1A implies that λ may differ for different pairs at stage r and produce diversity with PEM or, more significantly, may differ for the same pair with the same ordering prior to r and a different ordering thereafter and still produce diversity.

On trials when exactly one of the S_1 stimulus elements (say o^*) is the same as one of the S_2 stimulus elements ($o_1^* = o_2^* = o^*$, a T_2 type of trial), consider as prototype the order when o_1^* is in S_1 position u, and in S_2 is in position v_1. Assume further that this pair (o_1^*, o_2^*) with serial position h^{*+} is processed r_1th and designate $r_2 > r_1$ as the processing position of a specific pair that includes o_1^* as its S_1 member and another different stimulus from S_2, the serial position of the latter pair to be called h^{*-}. We assume for later reference (Theorem 3) that the second, different, stimulus is in position v_2 in S_2. It will be helpful to write the joint characteristic functions for this type of trial; these formulas follow directly from Lemmas 1 and 2.

For the serial model the joint characteristic function is

$$\phi^{(s)}(T(k_1, 1, K, -1),\ldots, T(h^{*-}, r_1, K, +1),\ldots, T(h^{*-}, r_2, K, -1),\ldots, T(k_N, N, K, -1))$$
$$= P(K)\,\frac{\lambda(k_1, 1, K, -1)}{\lambda(k_1, 1, K, -1) - is_1}\cdots\frac{\lambda(h^{*+}, r_1, K, +1)}{\lambda(h^{*+}, r_1, K, +1) - is_{r_1}}$$
$$\cdots\frac{\lambda(h^{*-}, r_2, K, -1)}{\lambda(h^{*-}, r_2, K, -1) - is_{r_2}}\cdots\frac{\lambda(k_N, N, K, -1)}{\lambda(k_N, N, K, -1) - is_N}.$$

It is useful for the parallel joint characteristic function to separate at each stage the $+$ part from the others. Then we may express

$$\phi^{(p)}(T(k_1, 1, K, -1),\ldots, T(h^{*+}, r_1, K, +1),\ldots, T(h^{*-}, r_2, K, -1),\ldots, T(k_N, N, K, -1))$$
$$= \frac{\nu(k_1, 1, m(1), -1)}{\sum_{j=1, h_{ij}\neq h^{*+}}^{N}\nu(h_{ij}, 1, m(1), -1) + \nu(h^{*+}, m(1), +1) - is_1}$$
$$\cdots\frac{\nu(h^{*+}, r_1, m(r_1), +1)}{\sum_{j=r_1+1, h_{r_1 j}\neq h^{*+}}^{N}\nu(h_{r_1 j}, r_1, m(r_1), -1) + \nu(h^{*+}, r_1, m(r_1), +1) - is_{r_1}}$$
$$\cdots\frac{\nu(h^{*-}, r_2, m(r_2), -1)}{\sum_{j=r_2, h_{r_2 j}\neq h^{*-}}^{N}\nu(h_{r_2 j}, r_2, m(r_2), -1) + \nu(h^{*-}, r_2, m(r_2), -1) - is_{r_2}}$$
$$\cdots\frac{\nu(k_N, N, m(N), -1)}{\nu(k_N, N, m(N), -1) - is_N}.$$

Theorem 2 states that differential processing speeds for $+$ as opposed to $-$ matches makes distribution diversity a certainty; now restricting our attention to the *context* of T_2 trials. As before, it is postulated that no order K exists such that $P(K) = 0$.

THEOREM 2. *If $+$ comparisons are processed at different rates than are $-$ comparisons then* SEM *and* PEM *are distribution diverse with respect to T_2 trials.*

Proof. Consider without loss of generality the alternative order K' which permutes the *processing* positions of h^{*+} and h^{*-} but which leaves the history up until r_1, $(m(r_1))$, unchanged. Then if total equivalence were to hold, it must be the case that

$$\lambda(h^{*+}, r_1, K, +1) = \sum_{j=r_1+1}^{N}\nu(h_{r_1 j}, r_1, m(r_1), -1) + \nu(h^{*+}, r_1, m(r_1), +1)$$
$$= \lambda(h^{*-}, r_1, K', -1).$$

But, $\lambda(h^{*+}, r_1, K, +1) \neq \lambda(h^{*-}, r_1, K', -1)$ by hypothesis; therefore SEM and PEM are distribution diverse. Q.E.D.

Note that the sum of ν's in the proof is the same for K and K' since the ordering up until r_1, given by $m(r_1)$ is the same for K, K'. Note also, that Theorem 2 depends on a change in order after stage r as did Theorem 1A, but now even if the λ's do not differ

according to serial position or the order after stage r they may because of $+$, $-$ processing differences and thereby produce diversity.

An obvious corollary follows.

COROLLARY. *If the rates do not differ for* $+$, $-$ *comparisons* and *the* λ's *otherwise satisfy the constraint of Theorem* 1B, SEM *and* PEM *are totally equivalent on* T_2 *trials.*

Theorem 3 also obtains a result of distribution diversity, again following from different processing rates on $+$ and $-$ comparisons and again involving the context of T_2 trials. However, in this case it is PEM that is the more general. Basically, Theorem 3 uses the fact that the serial (SEM) probability of selecting a given element (stimulus pair) with specified positions in S_1 and S_2, for a certain *processing* position, cannot depend on whether the stimuli in that pair match or do not match. On the other hand, the $+$ and $-$ rates differ in the parallel model (as in the serial), and processing order in parallel processing is entirely dependent on rates of processing. It therefore follows that the parallel probability of a pair with specified S_1 and S_2 positions being completed at a given stage *will* depend on whether the pair matches or does not.

THEOREM 3. *If* $+$ *comparisons are processed at different rates than* $-$ *comparisons, then when* T_2 *trials with different placement of* o^* *in* S_1 *and* S_2 *are considered,* SEM *and* PEM *are distribution diverse, if* $P[m(N+1)]$ *is different for the distinct placements.*

Proof. Hold K constant and permute o^* in S_1 and S_2 such that ${o_1}^*$ is still in position u, but the second, different stimulus is now in position v_1, and in position v_2 we place ${o_2}^*$ $(= {o_1}^*)$.

Since the order K on serial positions is unchanged, according to SEM, $P(K)$ should be invariant with respect to what is in the stimulus positions of S_1 and S_2. In particular, $P(h^{*+}_{r_1} \mid k_1, k_2, \ldots, k_{r_{1-1}}) = P(h^{*-}_{r_1} \mid k_1, k_2, \ldots, k_{r_{1-1}})$, where $h^{*+}_{r_1}$ is the serial position of the matching pair (${o_1}^*$ at u and ${o_2}^*$ at v_1) and $h^{*-}_{r_1}$ is that of a mismatching pair (${o_1}^*$ at u and a different mismatching stimulus at v_1). Thus, $h^{*+}_{r_1}$ and $h^{*-}_{r_1}$ represent the same *stimulus positions* in S_1 and S_2 (and therefore the same pairwise serial position, $h^{*+}_{r_1} = h^{*-}_{r_1}$), but the first pair matches and the second mismatches.

These facts together with Theorem 1 imply that for total equivalence to hold we must simultaneously satisfy

$$P(h^{*+}_{r_1} \mid k_1, k_2, \ldots, k_{r_{1-1}})$$

$$= \frac{\nu(h^{*+}, r_1, m(r_1), +1)}{\sum_{j=r_1+1}^{N} \nu(h_{r_1 j}, r_1, m(r_1), -1) + \nu(h^{*+}, r_1, m(r_1), +1)}$$

$$= P(h^{*-}_{r_1} \mid k_1, k_2, \ldots, k_{r_{1-1}})$$

$$= \frac{\nu(h^{*-}, r_1, m(r_1), -1)}{\sum_{\substack{j=r_1+1 \\ j \neq r_2}}^{N} \nu(h_{r_1 j}, r_1, m(r_1), -1) + \nu(h^{*-}, r_1, m(r_1), -1) + \nu(h^{*+}, r_2, m(r_2), +1}.$$

on the other hand, if there are unequal, PEM and SEM are distribution diverse. Q.E.D.

COROLLARY. *If the processing rates do not differ for + and − comparisons and the hypothesis of Theorem* 1B *holds, then* PEM *and* SEM *are totally equivalent even across trials with different (experimental) placement of the "same" stimulus.*

Proof. The proof follows immediately from the homeomorphism employed in Theorem 1B. Q.E.D.

Yet another facet of diversity is discovered when T_1 trials are compared with T_2 trials; that is, in the context of simple matching experiments. More specifically, if a certain mismatching pair is processed at some stage then its rate may be presumed to be the same in a serial system whether the other stimulus pairs match or not. But in a parallel system, the intercompletion rate of that stage quite obviously depends on the composition of the other uncompleted pairs. This type of diversity is captured in Theorem 4. Here, as in Theorem 3, it is the parallel model that is the more general.

It is assumed that the S_1 stimuli are identical in the T_1 and T_2 stimulus sets to be considered, but that exactly one S_2 stimulus is removed from the T_1 set and replaced with a stimulus that matches exactly one of the S_1 stimuli. Arbitrarily select a − pair that is found in both the T_1 and T_2 sets with serial position h^{*-}.

THEOREM 4. *If + comparisons are processed at different rates than − comparisons, then* SEM *and* PEM *are distribution diverse with respect to simple matching experiments.*

Proof. Fix the order for both the T_1 and T_2 trials. Refer to the pairs on the T_2 trial containing o_2^* $(=o_1^*)$ as their S_2 member, as "b" pairs and those not containing o_2^* as "a" pairs. The T_1 trial contains only "a" elements. Now, designate the serial position of "a" tuples as h_a and those of "b" tuples as h_b and suppose an "a" pair (say a^{*-}) completed at stage r has serial position h^{*-}.

The SEM exponential intercompletion time survivor function for element pair a^{*-} at stage r when conditionalized on order is $e[-\lambda(h^{*-}, r, K, -1)\, t_r]$ and this is identical on the T_1 and T_2 trials. Note that this expression is independent of the composition of the other element pairs, in particular, the uncompleted pairs. On the other hand, the intercompletion time survivor function of PEM on a^{*-} at stage r, again conditionalized on order, is

$$e\left\{-\left[\nu(h^{*-}, r, m(r), -1) + \sum_{j=r+1}^{N} \nu(h_{rj}, r, m(r), -1)\right] t_r\right\} \qquad \text{for } T_1 \text{ trials,}$$

and

$$e\left\{-\left[\nu(h^{*-}, r, m(r), -1) + \sum_{\{h_a \mid h_a \neq h^{*-}\}} \nu(h_r(a), r, m(r), -1) + \sum_{\{h_b\}} \nu(h_r(b), r, m(r), +1)\right] t_r\right\}$$

for T_2 trials. As indicated, the T_1 and T_2 expressions need not be equal; therefore, λ cannot be made identical to both simultaneously and PEM and SEM are distribution diverse in this circumstance. Q.E.D.

COROLLARY. *If the processing rates do not differ for + and − comparisons and the hypothesis of Theorem* 1B *holds, then* PEM *and* SEM *are totally equivalent with respect to simple matching experiments.*

Potential testability (which requires "diversity" but also observability) between SEM and PEM based on the above theorems then necessitates selecting materials and/or aspects of the processing mechanisms that promote:

(a) Different rates on different serial positions and different possible "pathways" through the (T_1) pairs (Theorem 1).

(b) Different rates on +, − comparisons and different possible "pathways" through the (T_2) pairs (Theorem 2).

(c) Same as (b) but also ensure that sets of experimental stimuli for different trials include pairs where the duplicated stimulus appears in different stimulus positions (Theorem 3).

(d) Different rates on +, − comparisons and use of simple matching experiments (Theorem 4).

In concluding discussion of the above theorems, we note that they were stated in terms of parallel or serial processing of pairs of stimuli. The theorems are directly applicable to any set of elements that can be characterized with logical structure similar to that for pairs of stimuli selected from two finite sets.

Another possibility of producing distribution diversity, apart from the above theorems lies in letting some processing orders K possess zero probability of occurrence; in the extreme, letting one path occur with probability 1. Recall that $P(K) \neq 0$ for all K was assumed as a postulate for the above developments. The idea here would be that when, for example, $P(K) = 1$ for serial K in SEM (and clearly when based on other distributions than the exponential as well), no PEM can be totally equivalent to it, since to do so violates the part of PEM definition requiring nonzero rates at all times for uncompleted pairs. As can be seen (with minor extrapolation) from the mapping in Theorem 1B, PEM can approximate SEM at these cases to an arbitrarily close degree by letting all but one ν be arbitrarily close to zero at any stage, assuming the other hypotheses for total equivalence are satisfied. In a negative sense, this means that a parallel model can act very much like a single-processing-order serial model. In a positive sense, it means that if a parallel model is the correct model, it is nevertheless evolving into or acting like a serial model, since when all the ν's but one are zero at any given stage, it *is* a serial model.

Before proceeding it should be pointed out that it is possible to generalize these results to situations where differences between particular stimuli and therefore stimulus comparison rates, may be present (e.g., it may take longer to compare I and J than I and O), and be of potential aid in parallel–serial testability. A related line of generalization is to situations where some of the stimuli are of one category (say numbers) and the others are of a different category (e.g., letters) and the two categories are processed at different rates. It appears that anything that enhances different processing rates for different pairs will tend to increase the chances of distribution diversity of the sort shown above for different stimulus positions, and + and − comparisons.

5. A Simple Matching Design to Illustrate Models and Theorems

Although the above distribution diversity developments suggest some indication of where testability of PEM and SEM may lie, they do not prove that such testability exists. The problem that arises in attempted utilization of the distribution diversity results, consists in the subtlety of the manifestations in data of PEM and SEM structural differences. Not only is there the usual possibility of a fairly realistic model possessing more parameters than an experiment possesses degrees of freedom, but the statistics typically observable may not reveal distinctive differences between parallel and serial processing.

It is pertinent to consider a simple special case of SEM and PEM in conjunction with a typical type of experiment in order to

(1) realize certain intuitions and clarify PEM–SEM structure that may have been partially obscured by the generality of the above treatment,

(2) indicate how PEM, SEM differences may become submerged or "averaged out" in traditional statistics, and

(3) suggest statistics that do test SEM–PEM models, when and if they can be obtained.

Following that, a more complex type of experimental design that yields SEM–PEM testability will be examined.

Consider a simple matching experiment where the basic types of trials are (1) T_1 : Stimulus o_1 absent from S_2 , (2) T_2 : Stimulus o_1 present in position 1 in S_2 , (3) T_2 : Stimulus o_1 present in position 2 in S_2 ; where, for example, o_1 , o_2 , o_3 might be the randomly selected letters c, v, 1. The potential stimulus configurations are determined by the serial positions of the S_2 elements, since it is assumed the serial position of o_1 in S_1 is fixed. The S_2 configurations, after ordering by serial position, are of the form $\langle o_1 , o_2 \rangle$, $\langle o_2 , o_1 \rangle$, and $\langle o_2 , o_3 \rangle$. Hence, the serial position index of a pair is $h_r = 1$ if the S_2 element of a pair not yet finished at stage r is from the first serial position

and $h_r = 2$ if it is from the second. Of course, we may still use $h_r = k_r$ to denote the element actually completed at the end of stage r. Note that $S_1 = o_1$.

Recall that the serial rates are of the form λ(pair-serial position, stage, order, type of comparison). Thus, the 2 in $\lambda(2, \cdot, \cdot, \cdot)$ gives the processing rate of the pair made up of o_1 and the second (right-hand) member of S_2 which might be o_1, o_2, o_3. However, the 2 (in $\lambda(\cdot, 2, \cdot, \cdot)$) says that the stage is number 2 and the 1 (in $\lambda(\cdot, \cdot, 1, \cdot)$) means that the order of processing is number 1 (i.e., compare o_1 with the leftmost member of S_2 then do the rightmost). This, plus $\lambda(\cdot, \cdot, \cdot, +1)$ tell us that this particular λ applies to an ordering of the S_2 elements, $S_2 : \langle o_2, o_1 \rangle$ or $\langle o_3, o_1 \rangle$ when o_1 and o_1 are being compared in stage 2 after o_1 and o_2 (or o_3) have already been compared in stage 1. In conjunction with our previous definitions, this results in the parameter space determined by the following set of parameters for SEM with $n_1 = 1$, $n_2 = 2$.

$$\langle p, \lambda(1, 1, 1, +1), \lambda(1, 1, 1, -1), \lambda(2, 2, 1, +1), \lambda(2, 2, 1, -1),$$
$$\lambda(1, 2, 2, +1), \lambda(1, 2, 2, -1), \lambda(2, 1, 2, +1), \lambda(2, 1, 2, -1)\rangle,$$

where p is the probability of comparing the *first* (S_2) position with the early stimulus first, and the λ's are, of course, the rates for the two pairs at the various stages and duplicative possibilities.

This formulation gives a total of nine parameters which can be verified with the earlier formula

$$(n_1 \cdot n_2)! - 1 + 2(n_1 \cdot n_2)\,((n_1 \cdot n_2)!) = (1 \cdot 2)! - 1 + 2(1 \cdot 2)\,((1 \cdot 2))! = 9.$$

In an analogous manner, the parameter space for PEM is generated by

$$\langle \nu(1, 1, \langle \varnothing, \varnothing \rangle, +1), \nu(1, 1, \langle \varnothing, \varnothing \rangle, -1), \nu(2, 2, \langle 1, \varnothing \rangle, +1), \nu(2, 2, \langle 1, \varnothing \rangle, -1),$$
$$\nu(1, 2, \langle 2, \varnothing \rangle, +1), \nu(1, 2, \langle 2, \varnothing \rangle, -1), \nu(2, 1, \langle \varnothing, \varnothing \rangle, +1), \nu(2, 1, \langle \varnothing, \varnothing \rangle, -1)\rangle,$$

the rate parameters which govern the behavior of PEM. Now, $\langle \varnothing, \varnothing \rangle = m(1)$, the null history before the first stage is over and $\langle h_1, \varnothing \rangle = m(2)$, where $h_1 = 1, 2$ is the serial position of the pair (and thus the serial position of the S_2 element) that is completed at the end of stage 1. Of course, the parallel–serial correspondence of order of processing is $\langle 1, 2 \rangle = 1$ and $\langle 2, 1 \rangle = 2$, although to be sure, the parallel order is determined in this case when the first pair is completed. The other entries are as in the SEM case. The formula $2(n_1 \cdot n_2)! \sum_{j=0}^{n_1 n_2 - 1} (1/j!) = 2 \cdot (1 \cdot 2)!\,(1 + 1) = 8$ yields the correct number of parameters.

In this specific situation, the parameter notation can be simplified. In fact, we can write $\lambda_{ij}(\cdot)$ to be the SEM rate for $h_i = 1, 2$ at stage j ($j = 1, 2$) for a $(\cdot)$ comparison ($\cdot = +, -$). Note that K would be redundant here. The parameter $\nu_{ij}(\cdot)$ is defined in exactly the same way.

There are six joint characteristic functions for the intercompletion times; two each for the three trial types above. Each characteristic function will be written as conditionalized on trial type and placement of the duplicated stimulus in S_2. Thus, $E(e^{is_{11}^{+}T_{11}^{+}} \cdot e^{is_{22}^{-}T_{22}^{-}} \mid \langle +, - \rangle)$ is the defining function of the joint characteristic function of intercompletion times when the duplicated (same) stimulus is in S_2 stimulus position 1 and the model represents the element processing system as comparing the early stimulus with that late stimulus in (S_2) position 1 first, and so on. Similarly, the order is facilely given in terms of the $+$, $-$ comparisons $\langle +, - \rangle$.

To enhance comparison of the serial and parallel expressions, the two corresponding formulas for the joint characteristic function will be given together, serial first (S), parallel second (P).

$$E[e^{is_{11}^{-}T_{11}^{-}}, e^{is_{22}^{-}T_{22}^{-}} \mid \langle -, - \rangle] = p \frac{\lambda_{11}^{-}}{\lambda_{11}^{-} - is_{11}^{-}} \cdot \frac{\lambda_{22}^{-}}{\lambda_{22}^{-} - is_{22}^{-}} \quad \text{(S)}, \quad (1)$$

$$= \frac{\nu_{11}^{-}}{\nu_{11}^{-} + \nu_{21}^{-} - is_{11}^{-}} \cdot \frac{\nu_{22}^{-}}{\nu_{22}^{-} - is_{22}^{-}} \quad \text{(P)}; \quad (2)$$

$$E[e^{is_{21}^{-}T_{21}^{-}}, e^{is_{12}^{-}T_{12}^{-}} \mid \langle -, - \rangle] = (1 - p) \frac{\lambda_{21}^{-}}{\lambda_{21}^{-} - is_{21}^{-}} \cdot \frac{\lambda_{12}^{-}}{\lambda_{22}^{-} - is_{12}^{-}} \quad \text{(S)}, \quad (3)$$

$$= \frac{\nu_{21}^{-}}{\nu_{11}^{-} + \nu_{21}^{-} - is_{21}^{-}} \cdot \frac{\nu_{12}^{-}}{\nu_{12}^{-} - is_{12}^{-}} \quad \text{(P)}; \quad (4)$$

$$E[e^{is_{11}^{+}T_{11}^{+}}, e^{is_{22}^{-}T_{22}^{-}} \mid \langle +, - \rangle] = p \frac{\lambda_{11}^{+}}{\lambda_{11}^{+} - is_{11}^{+}} \cdot \frac{\lambda_{22}^{-}}{\lambda_{22}^{-} - is_{22}^{-}} \quad \text{(S)}, \quad (5)$$

$$= \frac{\nu_{11}^{+}}{\nu_{11}^{+} + \nu_{21}^{-} - is_{11}^{+}} \cdot \frac{\nu_{22}^{-}}{\nu_{22}^{-} - is_{22}^{-}} \quad \text{(P)}, \quad (6)$$

$$E[e^{is_{21}^{-}T_{21}^{-}}, e^{is_{12}^{+}T_{12}^{+}} \mid \langle +, - \rangle] = (1 - p) \frac{\lambda_{21}^{-}}{\lambda_{21}^{-} - is_{21}^{-}} \cdot \frac{\lambda_{12}^{+}}{\lambda_{12}^{+} + is_{12}^{+}} \quad \text{(S)}, \quad (7)$$

$$= \frac{\nu_{21}^{-}}{\nu_{11}^{+} + \nu_{21}^{-} - is_{21}^{-}} \cdot \frac{\nu_{12}^{+}}{\nu_{12}^{+} - is_{12}^{+}} \quad \text{(P)}; \quad (8)$$

$$E[e^{is_{11}^{-}T_{11}^{-}}, e^{is_{22}^{+}T_{22}^{+}} \mid \langle -, + \rangle] = p \frac{\lambda_{11}^{-}}{\lambda_{11}^{-} - is_{11}^{-}} \cdot \frac{\lambda_{22}^{+}}{\lambda_{22}^{+} - is_{22}^{+}} \quad \text{(S)}, \quad (9)$$

$$= \frac{\nu_{11}^{-}}{\nu_{11}^{-} + \nu_{21}^{+} - is_{11}^{-}} \cdot \frac{\nu_{22}^{+}}{\nu_{22}^{+} - is_{22}^{+}} \quad \text{(P)}; \quad (10)$$

$$E[e^{is_{21}^{+}T_{21}^{+}}, e^{is_{12}^{-}T_{12}^{-}} \mid \langle -, + \rangle] = (1-p)\frac{\lambda_{21}^{+}}{\lambda_{21}^{+} - is_{21}^{+}} \cdot \frac{\lambda_{12}^{-}}{\lambda_{12}^{-} - is_{12}^{-}} \qquad (S),\ (11)$$

$$= \frac{\nu_{21}^{+}}{\nu_{11}^{-} + \nu_{21}^{+} - is_{21}^{+}} \cdot \frac{\nu_{12}^{-}}{\nu_{12}^{-} - is_{12}^{-}} \qquad (P)\ (12)$$

First observe those characteristic functions which as a set are intrinsically related to Theorem 1. These correspond to the trials with no + comparisons (Eqs. (1)–(4)). We note as in the general case, a type of greater generality of the serial expressions. This can be seen directly by setting the first and second parts of (1), (2) and (3), (4) equal to one another and solving for the serial variables (using the method of undetermined coefficients of the variable "is").

By way of short cut, however, use of the previously developed mappings shows, and inspection of (1)–(4) makes it intuitively reasonable that

$$p = \nu_{11}^{-}/(\nu_{11}^{-} + \nu_{21}^{-}), \qquad \lambda_{11}^{-} = \nu_{11}^{-} + \nu_{21}^{-}, \qquad \lambda_{21}^{-} = \nu_{11}^{-} + \nu_{21}^{-}$$

if equivalence is to hold, even for (1)–(4) alone. But, this constrains $\lambda_{11}^{-} = \lambda_{21}^{-} = \lambda_{1}^{-}$ which, in general, does not have to hold. The reverse mappings, SEM → PEM, are of course,

$$\nu_{11}^{-} = p\lambda_{1}^{-}, \qquad \nu_{21}^{-} = (1-p)\lambda_{1}^{-},$$

and in both SEM → PEM and PEM → SEM, $\nu_{22}^{-} = \lambda_{22}^{-}$, $\lambda_{12}^{-} = \nu_{12}^{-}$.

Next, examine the joint characteristic functions for trials when the duplicated stimulus is in S_2 position 1 (Eqs. (5)–(8)). Equations (5)–(8) as a set correspond to Theorem 2, since + rates may differ from − rates, a + match occurs, but different orders are pictured. Employing the same type of mapping as earlier, we find that if total equivalence is to hold, that

$$p = \nu_{11}^{+}/(\nu_{11}^{+} + \nu_{21}^{-}), \qquad \lambda_{11}^{+} = \nu_{11}^{+} + \nu_{21}^{-}, \qquad \lambda_{21}^{-} = \nu_{11}^{+} + \nu_{21}^{-},$$

or equivalently,

$$\nu_{11}^{+} = p\lambda_{11}^{+} = p\lambda_{21}^{-}, \qquad \nu_{21}^{-} = (1-p)\lambda_{11}^{+} = (1-p)\lambda_{21}^{-},$$

and the second stage parameter mappings, as usual, are able to equal one another with complete freedom. Again, the constraints necessary for total equivalence are placed on the λ's, i.e., $\lambda_{21}^{-} = \lambda_{11}^{+}$. Notice that even if λ did not otherwise differ according to order, if $\lambda^{+} \neq \lambda^{-}$, the parallel and serial processes could not be totally equivalent.

Equations (9)–(12) and the two characteristic functions associated with them are

also pertinent to Theorem 2. The only difference is that here, it will be necessary that $\lambda_{21}^{+} = \lambda_{11}^{-}$ in order that total equivalence occur. In particular,

$$p = \nu_{21}^{+}/(\nu_{11}^{-} + \nu_{21}^{+}), \qquad \lambda_{11}^{-} = \nu_{11}^{-} + \nu_{21}^{+} = \lambda_{21}^{+} = \lambda_1 ,$$

or, in reverse,

$$\nu_{11}^{-} = p\lambda_{11}^{-} = p\lambda_{21}^{+} , \qquad \nu_{21}^{+} = (1 - p)\,\lambda_{11}^{-} = (1 - p)\,\lambda_{21}^{+}$$

and the second stage parameters are equal.

As noted, together Eqs. (5)–(12) depict behavior relating to the function with a matching stimulus in S_1 and S_2. We have shown immediately above that total equivalence necessitates $\lambda_{11}^{+} = \lambda_{21}^{-}$ and $\lambda_{11}^{-} = \lambda_{21}^{+}$. However, $\nu_{11}^{+}, \nu_{21}^{-}, \nu_{11}^{-}, \nu_{21}^{+}$ can all be different, although these ratios, $\nu_{11}^{+}/(\nu_{11}^{+} + \nu_{21}^{-})$, $\nu_{11}^{-}/(\nu_{11}^{-} + \nu_{21}^{+})$ must be equal if total equivalence is to hold. These two ratios must be equal to one another because the PEM $\rightarrow$ SEM mapping shows that both of these quantities must equal p. This illustrates Theorem 3. It was shown there that placement of the matching stimulus in a different place and calculation of the probability that the same *position* (in S_1, S_2) pair is completed during a given stage, resulted in the same serial probability for the two placements but two different parallel probabilities. Solving for p in (5), (6) and then in (9), (10) proves this supposition in the present situation. Again, it should be noted that this is an aspect of greater generality of parallel processing.

The constancy of the p parameter is not the only place where PEM is more general than the serial class. The rates are also involved via Theorem 4. Note for instance, that λ_{21}^{-} and λ_{11}^{-} both have to equal two different sums of ν's for total equivalence to hold (observe expressions (3), (4), (7), (8), and (1), (2), (9), (10), respectively), $\nu_{11}^{-} + \nu_{21}^{-}$ and $\nu_{11}^{+} + \nu_{21}^{-}$ both equal λ_{21}^{-}, and $\nu_{11}^{-} + \nu_{21}^{-}$ and $\nu_{11}^{-} + \nu_{21}^{+}$ both equal λ_{11}^{-}. Hence, if we look separately at trials with and without a duplicated stimulus in S_2 position 2, then conditioned on order of processing, say position 1 first, then the distribution of completion times for the first comparison will be the same (whether or not a duplicated stimulus is present) for the serial models; the latter given by $\lambda_{11}^{-}/(\lambda_{11}^{-} - is)$. The parameters $\lambda_{11}^{+}, \lambda_{21}^{+}$ do not appear. The analogous conditioned distributions for the parallel models, on the other hand, will be given by $(\nu_{11}^{-} + \nu_{21}^{-})/(\nu_{11}^{-} + \nu_{21}^{-} - is)$, for trials with no duplicated stimulus and by $(\nu_{11}^{-} + \nu_{21}^{+})/(\nu_{11}^{-} + \nu_{21}^{+} - is)$, with a duplicated stimulus object, which are, in general, not equal.

When *all* the expressions for this particular paradigm are taken together ((1)–(12)) we find that $\lambda_{11}^{+} = \lambda_{21}^{-}, \lambda_{11}^{-} = \lambda_{21}^{+}, \lambda_{11}^{-} = \lambda_{21}^{-}$ and so, if total equivalence is to follow it must be that $\lambda_{11}^{+} = \lambda_{21}^{-} = \lambda_{21}^{+} = \lambda_{11}^{-} = \lambda_1$. That is, there can be no + and − match-rate differences or order-dependent rate differences during stage 1. Similarly, in PEM we learn from the fact that $\lambda_{11}^{+} = \nu_{11}^{+} + \nu_{21}^{-}, \lambda_{11}^{-} = \nu_{11}^{-} + \nu_{21}^{+}$, etc., and the equality of the λ's that $\nu_{11}^{-} = \nu_{11}^{+} = \nu_{11}$ and $\nu_{21}^{-} = \nu_{21}^{+} = \nu_{21}$ to ensure total equivalence. Thus, at

stage 1, the serial rates may not differ according to position or +, − distinctions and the parallel rates may differ according to position but not the +, − match difference. On the other hand, SEM has p to represent position effects. At stage 2, $\nu_{22}^{-} = \lambda_{22}^{-}$, $\nu_{12}^{-} = \lambda_{12}^{-}$, $\nu_{22}^{+} = \lambda_{22}^{+}$, $\nu_{12}^{+} = \lambda_{12}^{+}$, hence at this stage a full latitude of dependence on order (on history) and + or − matching differences can occur.

The price, of course, of conforming to the equivalence mappings (and thus yielding total equivalence) is that models must give up the ability to predict different processing rates for + and − comparisons during the first stage. Further, the serial models even give up the latitude for distinct rates on different positions, being forced to discern position effects by way of p and the stage 2 rates.

As noted above, in the class of potential statistics that in principle might test PEM vs SEM are mean intercompletion times conditioned on serial position and order events. For instance, in discussing trials containing a + match, it may be inferred from Theorem 2 that the distribution to first completion time, conditioned on which S_2 stimulus position contains the duplicated stimulus and on which comparison takes place first, should exhibit a particular type of greater generality for SEM. When the duplicated stimulus is in (S_2) position 1, the characteristic function for PEM is given by $(\nu_{11}^{+} + \nu_{21}^{-})/(\nu_{11}^{+} + \nu_{21}^{-} - is)$, independent of which comparison actually occurs first. In the same fashion, the characteristic function when the second (S_2) position is occupied by the duplicated stimulus is $(\nu_{11}^{-} + \nu_{21}^{+})/(\nu_{11}^{-} + \nu_{21}^{+} - is)$, independent of whether S_2 position 1 or 2 is completed first. But, the analogous distributions for the serial models need not be equal:

$$\frac{\lambda_{11}^{+}}{\lambda_{11}^{+} - is} \neq \frac{\lambda_{21}^{-}}{\lambda_{21}^{-} - is}, \qquad \frac{\lambda_{11}^{-}}{\lambda_{11}^{-} - is} \neq \frac{\lambda_{21}^{+}}{\lambda_{21}^{+} - is}.$$

Hence, a statistic that is capable of testing PEM and SEM on this basis is the associated mean (minimum) completion time (first intercompletion time) conditionalized as above, with resulting values

$$\frac{1}{\nu_{11}^{+} + \nu_{21}^{-}}, \frac{1}{\nu_{11}^{-} + \nu_{21}^{+}} \quad \text{for PEM and} \quad \frac{1}{\lambda_{11}^{+}}, \frac{1}{\lambda_{21}^{-}}; \frac{1}{\lambda_{11}^{-}}, \frac{1}{\lambda_{21}^{+}} \quad \text{for SEM}$$

so that if $\lambda_{11}^{+} \neq \lambda_{21}^{-}$, $\lambda_{11}^{-} \neq \lambda_{21}^{+}$, PEM predicts a single mean value when the duplicated stimulus is in position 1 (or 2) but SEM can predict 2 if a + comparison takes place at a different rate than a − comparison, during stage 1 on position 1 or 2. Similar remarks can be made for the SEM–PEM differences connected with the other theorems.

The difficulty in applying the above results and the related statistics is connected with the difficulty of discerning (observing) which stimulus comparisons actually occur first on any given trial in a typical experiment. Apart from such information,

which would allow us to use means, stage 2 processing will be blended in and the overall mean statistics will not suffice to distinguish the models. In the above experiment, for example, the information typically obtainable in a reaction time experiment, neglecting additional residue latency involvements, are serial position means on + comparison trials and the overall mean on − comparison trials. These are, for SEM and PEM:

$$E(T^-) = p\left(\frac{1}{\lambda_{11}^-} + \frac{1}{\lambda_{22}^-}\right) + (1-p)\left(\frac{1}{\lambda_{21}^-} + \frac{1}{\lambda_{12}^-}\right) \quad \text{(S)},$$

$$= \frac{1}{\nu_{11}^- + \nu_{21}^-} + \frac{\nu_{11}^-}{\nu_{11}^- + \nu_{21}^-} \cdot \frac{1}{\nu_{22}^-} + \frac{\nu_{21}^-}{\nu_{11}^- + \nu_{21}^-} \cdot \frac{1}{\nu_{12}^-}, \quad \text{(P)};$$

$$E(T_1^+) = p\,\frac{1}{\lambda_{11}^+} + (1-p)\left(\frac{1}{\lambda_{21}^-} + \frac{1}{\lambda_{12}^+}\right) \quad \text{(S)},$$

$$= \frac{1}{\nu_{11}^+ + \nu_{21}^-} + \frac{\nu_{21}^-}{\nu_{11}^+ + \nu_{21}^-} \cdot \frac{1}{\nu_{12}^+}, \quad \text{(P)};$$

$$E(T_2^+) = p\left(\frac{1}{\lambda_{11}^-} + \frac{1}{\lambda_{22}^+}\right) + (1-p)\,\frac{1}{\lambda_{21}^+} \quad \text{(S)},$$

$$= \frac{1}{\nu_{11}^- + \nu_{21}^+} + \frac{\nu_{11}^-}{\nu_{11}^- + \nu_{21}^+} \cdot \frac{1}{\nu_{22}^+}, \quad \text{(P)},$$

where $E(T^-)$ = comparison time on (exhaustive) no-stimulus-duplication trials, $E(T_i^+)$ = comparison time on (self-terminating) stimulus duplication trials with $i = 1, 2$ referring to duplication in S_2 1 or 2, respectively.

It is evident that both SEM and PEM have sufficient flexibility relative to these statistics engendered by the convolving of stages and relatively large numbers of parameters, to fit an elephant. As is evident from earlier work (Townsend, 1971b, 1972) this does not mean that special cases may not be testable. For example, a PEM submodel with parameters constant across different histories and values of n_2 is diverse at the level of mean processing times relative to say, a SEM submodel with a constant rate over order and n_2 (i.e., all λ's equal) parameter; assuming self-termination on + matching trials. The rub is that there are other PEM (SEM) submodels (some of them are quite reasonable in certain contexts) that are totally equivalent relative to either of these two special cases, and hence, the experiment is not really testing seriality vs parallelity.

In the particular experiment described above, manipulating the number of elements in the early and late sets does not seem to be very helpful either, since the capacity

question then enters; that is, there is no a priori reason to expect the rates to be unrelated to the number of elements that need be compared. Also, the processing path distribution (SEM) or *distribution* of processing capacity via the rates (PEM or SEM) may change under these conditions. What is needed is an experimental design that can test a broad subset of PEM (SEM) models against some subset of the opposing class, preferably at the level of means of some easily obtainable dependent variable.

6. A Parallel–Serial Testing Paradigm (PST)

Optimally, no submodel of the serial model (SEM) should be able to make predictions, at the level of observable means, that any submodel of the parallel model (PEM) can make, and vice versa. Further, the ideal experiment would serve, given sufficient trials, etc., to specify the specific member of the proper model that applies.

This completely optimal aim probably can never be achieved in the realm of "black box" modeling. Nevertheless, the development below gives an experimental paradigm that can test the total PEM, against the total SEM. In fact, PEM is tested against a rather large class of additive serial models (i.e., a class including SEM) with the following provision. The rates for matches must differ from the rates for mismatches. If mean reaction time behavior is of the specified serial nature, then no PEM can be totally equivalent to this class of serial models and vice versa. Moreover, the models are diverse at the level of *observable means* and thus are mean-testable.

The paradigm (PST) may be thought of as an elaboration of simple matching experiments where three converging conditions contain T_1, T_2 trial types and a third type of trial with all S_2 elements matching the S_1 element. These conditions together with PEM and SEM meet the assumptions of Theorems 1–4 in such a manner that observable mean diversity results. The pertinent theorems for PST are 3 and 4 and it is therefore likely that other paradigms resting on Theorems 1, 2 or various combinations of 1–4 may also elicit observable forms of diversity.

The conditions are shown in stylized form in Table 1 with A referring to the (any) S_1 stimulus that is duplicated in S_2 and B refers to (any) S_2 stimulus that is different from A. It is the logical form of the paradigm that is important, not the particular realization. For instance, S_1 could contain two (A, B) rather than one stimulus and S_2 could contain one (A) rather than two. The designations A, B need not refer to separate stimuli. They might refer to categories or stimulus dimensions, for example.

Condition I simply demands of S that he determine the S_2 position of the target. Conditions II and III require that both S_2 stimulus positions contain the target or that at least one contain the target, respectively, in order for a "yes" response to occur. We will investigate predictions of processing time rather than, say, probability correct. Thus, the present predictions correspond to a high-accuracy reaction time experiment,

TABLE 1

Condition		Response
I		
	Early stimulus: A	
	Late stimulus Set 1, AB	yes
	Set 2, BA	no
II		
	Early stimulus: A	
	Late stimulus Set 1, AA	yes
	Set 2, AB	no
	Set 3, BA	no
	Set 4, BB	no
III		
	Early stimulus: A	
	Late stimulus Set 1, AA	yes
	Set 2, AB	yes
	Set 3, BA	yes
	Set 4, BB	no

although there is no reason predictions cannot be extended to include low accuracy conditions.

It should not be inferred from Table 1 that the same stimuli are presented on each trial. Rather A simply represents the target, whatever it is on any particular trial and B any nontarget. In fact, the BB in Condition II and III of Table 1 need not refer to identical stimuli, as long as neither is identical to the target.

The derivations of the means for the various cases will not be given, since they are completely standard. On the other hand, it is necessary that some rather tedious algebra be dealt with in proving the nonequivalence results. The proofs are useful in making apparent the structure of the models relative to the experimental design and the aspects of the models whence diversity arises.

The present full class of parallel models PEM is considered. The rates can depend on stage, S_2 position and whether a comparison is $+$ or $-$. The parallel parameters are, therefore (using our simplified notation),

$$\langle \nu_{11}^{+}, \nu_{11}^{-}, \nu_{22}^{+}, \nu_{22}^{-}, \nu_{12}^{+}, \nu_{12}^{-}, \nu_{21}^{+}, \nu_{21}^{-} \rangle,$$

where $\nu_{i1}^{\cdot}$ gives the first stage rate for S_2 position i with a $\cdot$ ($-$ or $+$) comparison and $\nu_{i2}^{\cdot}$ is the same for stage 2.

The SEM assumes that $a +$ comparison can proceed at a rate different from that of $a -$ comparison and that the rates are in general dependent on stages, and S_2 position, thus yielding the parameter vector,

$$\langle p, \lambda_{11}^{+}, \lambda_{11}^{-}, \lambda_{22}^{+}, \lambda_{22}^{-}, \lambda_{12}^{+}, \lambda_{12}^{-}, \lambda_{21}^{+}, \lambda_{21}^{-} \rangle.$$

Table 2 lists the mean processing time predictions of SEM and PEM for PST. It is assumed that processing is self-terminating. Thus, in CI, either S_2 stimulus element gives full information and the matching process (comparator) never makes more than a single comparison. Similarly in CII, CII (1) requires exhaustive processing, CII (2) processing concludes whenever S_2 position 2 is completed, CII (3) whenever S_2 position 1 is finished, and CII (4) processing is curtailed after either of the S_2 positions completes its match against S_1. Self-evident analogous remarks hold for CIII.

It should be remarked here that if B is known to the subject then he might compare B, rather than A with the S_2 stimuli. The theorems below go through just as well in this case. However, if both A and B are compared simultaneously with the S_2 stimuli with self-termination whenever either type of comparison gains sufficient information to determine a response, I do not know if the theorems hold. In a number of interesting cases the B's are selected from a rather large set (e.g., all the letters except a target letter) and it would be very inefficient for the subject to match all members of this set against S_2 rather than the single target against S_2.

An especially interesting aspect, from an experimental point of view, of the succeeding theorem, is that it includes the presence of arbitrary additive residual reaction time components, t_p, t_s, that may be different ($t_p \neq t_s$) for PEM and SEM, respectively. A point in passing is that it is quite feasible to impose distribution assumptions on t_p and t_s and then test the entire set of reaction time distribution predictions associated with Table 1. It certainly follows from the earlier theorems (as well as a fortiori from Theorem 5 below) that the PEM and SEM distributions must be diverse. However, such questions as that of independence of the residuals t_s, t_p with respect to the comparison times rather than simply additivity, immediately arise. So although a distribution approach may prove fruitful, the present development avoids certain pitfalls surrounding a distributional analysis of processing stages. Theorem 5 presents the chief diversity and equivalence results.

THEOREM 5. (A) *If the stage* 1 *processing rates for either* S_2 *positions* 1 *or* 2 *differ for* $+$ *and* $-$ *comparisons* (i.e., $\lambda_{11}^{+} \neq \lambda_{11}^{-} (\nu_{11}^{+} \neq \nu_{11}^{-})$ *or* $\lambda_{21}^{+} \neq \lambda_{21}^{-} (\nu_{21}^{+} \neq \nu_{21}^{-})$) *then* PST *provides for observable mean diversity between* PEM *and* SEM, *even though to each* PEM *and* SEM *expression for mean processing time is added the residual latency component* t_p, t_s, *respectively.*

(B) *If none of the first stage rates of* SEM *or* PEM *differ according to whether comparisons are* $+$ *or* $-$, *then* SEM *and* PEM *are observable mean equivalent relative to* PST *so long as* $p \neq 0, 1$.

TABLE 2

Serial and Parallel Observable Mean RT Predictions for the Parallel–Serial Testing Paradigm (PST)

	Serial models	Parallel models
CI		
(1)	$\frac{p}{\lambda_{11}^{+}} + \frac{(1-p)}{\lambda_{21}^{-}} + t_s$	$\frac{1}{\nu_{11}^{+} \quad \nu_{21}^{-}} + t_p$
(2)	$\frac{p}{\lambda_{11}^{-}} + \frac{(1-p)}{\lambda_{21}^{+}} + t_s$	$\frac{1}{\nu_{11}^{-} + \nu_{21}^{+}} + t_p$
CII		
(1)	$p\left(\frac{1}{\lambda_{11}^{+}} + \frac{1}{\lambda_{22}^{+}}\right) + (1-p)\left(\frac{1}{\lambda_{21}^{+}} + \frac{1}{\lambda_{12}^{+}}\right) + t_s$	$\frac{1}{\nu_{11}^{+} + \nu_{21}^{+}} + \frac{\nu_{11}^{+}}{\nu_{22}^{+}(\nu_{11}^{+} + \nu_{21}^{-})} + \frac{\nu_{21}^{+}}{(\nu_{11}^{+} + \nu_{21}^{+})\nu_{12}^{+}} + t_p$
(2)	$p\left(\frac{1}{\lambda_{11}^{+}} + \frac{1}{\lambda_{22}^{-}}\right) + \frac{(1-p)}{\lambda_{21}^{-}} + t_s$	$\frac{1}{\nu_{11}^{+} + \nu_{21}^{-}} + \frac{\nu_{11}^{+}}{\nu_{22}^{-}(\nu_{11}^{+} + \nu_{21}^{-})} + t_p$

(3)	$\frac{p}{\lambda_{11}^-} + (1-p)\left(\frac{1}{\lambda_{21}^+} + \frac{1}{\lambda_{12}^-}\right) + t_s$	$\frac{1}{\nu_{11}^- + \nu_{21}^+} + \frac{\nu_{21}^+}{\nu_{12}^-(\nu_{11}^- + \nu_{21}^+)} + t_p$
(4)	$\frac{p}{\lambda_{11}^-} + \frac{(1-p)}{\lambda_{21}^-} + t_s$	$\frac{1}{\nu_{11}^- + \nu_{21}^-} + t_p$

CIII

(1)	$\frac{p}{\lambda_{11}^+} + \frac{(1-p)}{\lambda_{21}^+} + t_s$	$\frac{1}{\nu_{11}^+ + \nu_{21}^+} + t_p$
(2)	$\frac{p}{\lambda_{11}^+} + (1-p)\left(\frac{1}{\lambda_{21}^-} + \frac{1}{\lambda_{12}^+}\right) + t_s$	$\frac{1}{\nu_{11}^+ + \nu_{21}^-} + \frac{\nu_{21}^-}{\nu_{12}^+(\nu_{11}^+ + \nu_{21}^-)} + t_p$
(3)	$p\left(\frac{1}{\lambda_{11}^-} + \frac{1}{\lambda_{22}^+}\right) + \frac{(1-p)}{\lambda_{21}^+} + t_s$	$\frac{1}{\nu_{11}^- + \nu_{21}^+} + \frac{\nu_{11}^-}{\nu_{22}^+(\nu_{11}^- + \nu_{21}^+)} + t_p$
(4)	$p\left(\frac{1}{\lambda_{11}^-} + \frac{1}{\lambda_{22}^-}\right) + (1-p)\left(\frac{1}{\lambda_{21}^-} + \frac{1}{\lambda_{12}^-}\right) + t_s$	$\frac{1}{\nu_{11}^- + \nu_{21}^-} + \frac{\nu_{11}^-}{\nu_{22}^-(\nu_{11}^- + \nu_{21}^-)} + \frac{\nu_{21}^-}{\nu_{12}^-(\nu_{11}^- + \nu_{21}^-)} + t_p$

(C) *If the conditions of* (B) *hold with the exception that* $p = 0, 1$, *then* PEM *approximates* SEM *to an arbitrarily close degree in* PST, *but* SEM *is not able to mimic* PEM *for arbitrary* PEM *parameter values.*

Proof A. We proceed by reductio ad absurdum on the assumption that $\lambda_{11}^{+} \neq \lambda_{11}^{-}$. The proofs, if we instead assumed $\lambda_{21}^{+} \neq \lambda_{21}^{-}$, $\nu_{11}^{+} \neq \nu_{11}^{-}$ or $\nu_{21}^{+} \neq \nu_{21}^{-}$, are similar in form.

Assume the premise that PEM is not observable mean diverse under the present characterization. Then from Table 2 we observe that

$$\mathrm{CI}(1) = t_s + \frac{p}{\lambda_{11}^{+}} + \frac{(1-p)}{\lambda_{21}^{-}} = \frac{1}{\nu_{11}^{+} + \nu_{21}^{-}} + t_p,$$

$$\mathrm{CI}(2) = t_s + \frac{p}{\lambda_{11}^{-}} + \frac{(1-p)}{\lambda_{21}^{+}} = \frac{1}{\nu_{11}^{-} + \nu_{21}^{+}} + t_p,$$

and (13)

$$\begin{aligned}\mathrm{CI}(1) + \mathrm{CI}(2) &= p\left(\frac{1}{\lambda_{11}^{+}} + \frac{1}{\lambda_{11}^{-}}\right) + (1-p)\left(\frac{1}{\lambda_{21}^{-}} + \frac{1}{\lambda_{21}^{+}}\right) + 2t_s \\ &= \frac{1}{\nu_{11}^{+} + \nu_{21}^{-}} + \frac{1}{\nu_{11}^{-} + \nu_{21}^{+}} + 2t_p.\end{aligned}$$

But we further note that

$$\begin{aligned}\mathrm{CII}(4) + \mathrm{CIII}(1) &= p\left(\frac{1}{\lambda_{11}^{+}} + \frac{1}{\lambda_{11}^{-}}\right) + (1-p)\left(\frac{1}{\lambda_{21}^{-}} + \frac{1}{\lambda_{21}^{+}}\right) + 2t_s \\ &= \frac{1}{\nu_{11}^{+} + \nu_{21}^{+}} + \frac{1}{\nu_{11}^{-} + \nu_{21}^{-}} + 2t_p.\end{aligned} \tag{14}$$

Now, Eqs. (13) and (14), being the same in SEM, imply a constraint on a function of the ν's, namely,

$$\begin{aligned}&\mathrm{CI}(1) + \mathrm{CI}(2) = \mathrm{CII}(4) + \mathrm{CIII}(1) \\ &\Rightarrow \frac{1}{\nu_{11}^{+} + \nu_{21}^{-}} + \frac{1}{\nu_{11}^{-} + \nu_{21}^{+}} = \frac{1}{\nu_{11}^{+} + \nu_{21}^{+}} + \frac{1}{\nu_{11}^{-} + \nu_{21}^{-}}.\end{aligned} \tag{15}$$

A little algebraic inspection of (15) shows that for PEM $\equiv$ SEM at the level of mean processing times in PST means that PEM $\equiv$ SEM $\Rightarrow [\nu_{21}^{+} = \nu_{21}^{-} = \nu_{21}$ or $\nu_{11}^{+} = \nu_{11}^{-} = \nu_{11}]$, without loss of generality, set $\nu_{21}^{+} = \nu_{21}^{-} = \nu_{21}$. This result comes from the diversity associated with Theorem 4.

It then follows that in PEM, CI(1) = CIII(1) and CI(2) = CII(4), which is only true in SEM if $\lambda_{21}^{+} = \lambda_{21}^{-}$ or $p = 1$. If we had set $\nu_{11}^{+} = \nu_{11}^{-} = \nu_{11}$, then we would have obtained $\lambda_{11}^{+} = \lambda_{11}^{-}$ or $p = 0$. At this point, those PEM representatives supposing

neither $\nu_{21}^{+} = \nu_{21}^{-}$ nor $\nu_{11}^{+} = \nu_{11}^{-}$ are excluded, as are those SEM representatives that assume neither that $\lambda_{21}^{+} = \lambda_{21}^{-}$ nor $\lambda_{11}^{+} = \lambda_{11}^{-}$. But, it must now be verified that the remaining submodels of SEM and PEM also cannot be equivalent.

Assume first that $\lambda_{21}^{+} = \lambda_{21}^{-}$. Then, with little effort we obtain not only

$$\text{CIII}(2) - \text{CIII}(1) = \frac{(1-p)}{\lambda_{12}^{+}} = \frac{\nu_{21}}{\nu_{12}^{+}(\nu_{11}^{+} + \nu_{21})}, \tag{16}$$

$$\text{CII}(3) - \text{CII}(4) = \frac{(1-p)}{\lambda_{12}^{-}} = \frac{\nu_{21}}{\nu_{12}^{-}(\nu_{11}^{-} + \nu_{21})}, \tag{17}$$

$$\text{CII}(1) - \text{CIII}(2) = \frac{p}{\lambda_{22}^{+}} = \frac{\nu_{11}^{+}}{\nu_{22}^{+}(\nu_{11}^{+} + \nu_{21})}, \tag{18}$$

$$\text{CIII}(4) - \text{CII}(3) = \frac{p}{\lambda_{22}^{-}} = \frac{\nu_{11}^{-}}{\nu_{22}^{-}(\nu_{11}^{-} + \nu_{21})}, \tag{19}$$

but also

$$\begin{aligned} \text{CII}(1) - \text{CII}(2) &= p\left(\frac{1}{\lambda_{22}^{+}} - \frac{1}{\lambda_{22}^{-}}\right) + \frac{(1-p)}{\lambda_{12}^{+}} \\ &= \frac{\nu_{11}}{\nu_{11}^{+} + \nu_{21}}\left(\frac{1}{\nu_{22}^{+}} - \frac{1}{\nu_{22}^{-}}\right) + \frac{\nu_{21}}{\nu_{12}^{+}(\nu_{11}^{+} + \nu_{21})}, \end{aligned} \tag{20}$$

or

$$\begin{aligned} \text{CIII}(4) - \text{CIII}(3) &= p\left(\frac{1}{\lambda_{22}^{-}} - \frac{1}{\lambda_{22}^{+}}\right) + \frac{(1-p)}{\lambda_{12}^{-}} \\ &= \frac{\nu_{11}^{-}}{\nu_{11}^{-} + \nu_{21}}\left(\frac{1}{\nu_{22}^{-}} - \frac{1}{\nu_{22}^{+}}\right) + \frac{\nu_{21}}{\nu_{12}^{-}(\nu_{11}^{-} + \nu_{21})}. \end{aligned} \tag{21}$$

It can be observed that the constancy of p in (18), (19) is contradicted by

$$\frac{\nu_{11}^{+}}{\nu_{11}^{+} + \nu_{21}} \neq \frac{\nu_{11}^{-}}{\nu_{11}^{-} + \nu_{21}}$$

in the same expressions, a result associated with Theorem 3.

More formally, plug the equivalences obtained in (16), (18), (19) into (20) to derive the following equation in the ν's.

$$\frac{\nu_{11}^{+}}{\nu_{22}^{+}(\nu_{11}^{+} + \nu_{21})} - \frac{\nu_{11}^{-}}{\nu_{22}^{-}(\nu_{11}^{-} + \nu_{21})} = \frac{\nu_{11}^{+}}{\nu_{11}^{+} + \nu_{21}}\left(\frac{1}{\nu_{22}^{+}} - \frac{1}{\nu_{22}^{-}}\right).$$

And it follows immediately that $\nu_{11}^{+} = \nu_{11}^{-}$; or substituting (17)–(19) in (21) would again imply this result. But now, substituting $\nu_{11}^{+} = \nu_{11}^{-} = \nu_{11}$ back into CI(1), CI(2) yields $\lambda_{11}^{+} = \lambda_{11}^{-}$, which contradicts our original hypothesis.

It remains to show that a contradiction still appears if we assume instead of $\lambda_{21}^{+} = \lambda_{21}^{-}$, that $p = 1$. If $p = 1$ then CII(3) − CII(4) (recall that $\nu_{21}^{+} = \nu_{21}^{-}$) implies that

$$\nu_{21}/\nu_{12}^{-}(\nu_{11}^{-} + \nu_{21}) = 0.$$

Hence either $\nu_{21} = 0$, which is in violation of the tenets of a parallel process, that the rates should be nonzero for all unfinished comparisons at all stages, or $\nu_{12}^{-} = +\infty$ or $\nu_{11}^{-} = +\infty$, which contradicts our postulate concerning finite processing rates.

In any case, $\nu_{12}^{+} = \nu_{12}^{-} = +\infty$ (an attendant manipulation gets $\nu_{12}^{+} = +\infty$) implies that $\nu_{11}^{+} = \nu_{11}^{-}$ (examine CI(1), CII(1), CIII(3) when $p = 1$, $\nu_{12}^{+} = \nu_{12}^{-} = +\infty$) which again implies that $\lambda_{11}^{+} = \lambda_{11}^{-}$ again resulting in a contradiction. If $\nu_{11}^{+} = +\infty$ or $\nu_{11}^{-} = +\infty$ then we cannot even obtain finite predictions for CI.

Proof (B). Note first that the unobservable t_p, t_s were treated as variables that could only obscure diversity, that is, promote PEM–SEM equivalence. Hence it is sufficient to show equivalence results when $t_p = t_s$.

It may then be ascertained through simple algebra that the following mappings constitute equivalent PEM and SEM models.

The SEM → PEM mappings. Set

$$\nu_{11} = \left(\frac{p}{\lambda_{11}} + \frac{1-p}{\lambda_{21}}\right)^{-1} - \nu_{21},$$

where ν_{21} is fixed at some value such that

$$\left(\frac{p}{\lambda_{11}} + \frac{1-p}{\lambda_{21}}\right)^{-1} > \nu_{21}.$$

Next, let

$$\nu_{12}^{(\cdot)} = \lambda_{12}^{(\cdot)}\nu_{21}\left(\frac{p}{\lambda_{11}} + \frac{1-p}{\lambda_{21}}\right)\frac{1}{1-p},$$

and

$$\nu_{22}^{(\cdot)} = \lambda_{22}^{(\cdot)}\left[1 - \nu_{21}\left(\frac{p}{\lambda_{11}} + \frac{1-p}{\lambda_{21}}\right)\right]\frac{1}{p},$$

where $\cdot = +, -$.

The PEM → SEM mappings. Let p be arbitrary but contained in (0, 1). Next, let

$$\lambda_{11} = \left(\frac{1}{\nu_{11} + \nu_{21}} - \frac{1-p}{\lambda_{21}}\right)\frac{1}{p},$$

where λ_{21} is chosen so that the difference in parentheses is greater than zero. Next, let

$$\lambda_{12}^{(\cdot)} = \frac{(1-p)\,\nu_{12}^{(\cdot)}(\nu_{11}+\nu_{21})}{\nu_{21}}, \qquad \lambda_{22}^{(\cdot)} = \frac{p\nu_{22}^{(\cdot)}(\nu_{11}+\nu_{21})}{\nu_{11}},$$

where $\cdot = +, -$, and the proof of B is completed.

Proof C. Without loss of generality, let $p = 1$. Then we obtain from Table 2,

$$\frac{1}{\lambda_{11}} = \frac{1}{\nu_{11}+\nu_{21}}, \frac{1}{\lambda_{22}^{-}} = \frac{\nu_{11}}{\nu_{22}^{-}(\nu_{11}+\nu_{21})},$$

$$0 = \frac{\nu_{21}}{\nu_{12}^{-}(\nu_{11}+\nu_{21})}, \frac{1}{\lambda_{22}^{+}} = \frac{\nu_{11}}{\nu_{22}^{+}(\nu_{11}+\nu_{21})},$$

and

$$0 = \frac{\nu_{11}}{\nu_{12}^{+}(\nu_{11}+\nu_{21})}.$$

It is clear that SEM cannot mimic arbitrary parallel behavior. However, the PEM can clearly approximate SEM by picking appropriate ν's, in particular, letting ν_{12}^{-} and ν_{12}^{+} be so large that

$$\frac{\nu_{21}}{\nu_{12}^{-}(\nu_{11}+\nu_{21})} \quad \text{and} \quad \frac{\nu_{11}}{\nu_{12}^{+}(\nu_{11}+\nu_{21})}$$

are as close to zero as desired. Q.E.D.

It may be noted that letting $\nu_{12}^{+}, \nu_{12}^{-}$ be very large in Theorem 5(C) has different connotations from doing the same thing in 5(A). The reason is that in 5(C) no constraints are imposed on other allowable parameters. But, in 5(A), letting ν_{12}^{+} and ν_{12}^{-} be very large results in constraints on the ability of the models to reflect +, − processing rate differences in data; for example, ν_{11}^{+} must be close to ν_{11}^{-} in value.

As a simple example of 5(A), assume that the serial model is true with parameters, $p = 0.80$, $\lambda^{+} = 0.04$ comparisons/msec, $\lambda^{-} = 0.013$ comparisons/msec. That is, there is a fairly high likelihood of processing the tuple with the first-position member of S_2 first and while there is a +, − processing rate difference, order and identity of the pair do not otherwise matter. The value $\lambda^{+} = 0.04$ and $\lambda^{-} = 0.013$ correspond to average elapsed times of 25 and 75 msec/comparison.

Now, recall that SEM predicts that CI(1) + CI(2) = CII(4) + CIII(1), which is false for PEM unless $\nu_{21}^{-} = \nu_{21}^{+}$ (or $\nu_{11}^{+} = \nu_{11}^{-}$). But, if say, $\nu_{21}^{-} = \nu_{21}^{+}$, then PEM predicts that CI(1) = CIII(1), CI(2) = CII(4). Since the values of the data are CI(1) = 35 msec, CIII(1) = 25 msec, CI(2) = 65 msec, CII(4) = 75 msec, PEM is falsified. Similarly, if all the ν^{-}'s = 0.013 and the ν^{+}'s = 0.04 comparisons/msec, then CI(1) =

CI(2) = 18.9 msec, CII(4) = 38.5 msec, CIII(1) = 12.5 msec, and CI(1) + CI(2) = 37.8 $\neq$ CII(4) + CIII(1) = 51 msec, so in this case SEM would be falsified.

Of course, in the case of SEM being true, more complicated cases could arise, for instance, with $\lambda_{21}^{+} = \lambda_{21}^{-}$ but $\lambda_{11}^{+} \neq \lambda_{11}^{-}$, but similar consequences would nevertheless result. Further, it appears that the structure of the paradigm essentially entails some nonparametric tests; for example, the predictions by additive serial models (inclusive of SEM) that CI(1) + CI(2) = CII(4) + CIII(1). This latter is a strong prediction by a very large class of models.

We neglected t_p, t_s in the example, but with no loss of generality. Methods-of-moments estimates of all the parameters do not appear feasible, at least by hand, but least squares or chi-square estimates should permit numerical fits.

The preceding theorem (Theorem 5) demonstrates that the experimental paradigm given in PST is sufficient assuming that "same" comparisons are processed at a different rate than "different" comparisons to exploit the distribution diversity between SEM and PEM on an observable mean diversity level; even with the inclusion of additive residual latency components that could be different for SEM and PEM.

7. Conclusion

In addition to analytic tests developed so far, certain types of empirical results are sometimes acquired which are more intuitively predicted by, say, parallel than by serial processes, or vice-versa and some of these are mentioned in earlier papers (Townsend, 1971b, 1972).

As noted earlier, Thomas has recently contributed to the development of nonparametric tests of parallel and serial predictions. The tests developed there were limited to mismatching (−1) comparisons and to stage-1 completion times. On these aspects, the present treatment is more general. However, Thomas' (1969) basic parallel assumption was that of within-stage independence and hence was more general than the parallel-exponentiality assumed in PEM.

Thomas' results were based on a two-dimensional experimental paradigm with three sets of conditions. One of the conditions required a positive response if a specified one of two displayed stimulus dimensions was an appropriate value, otherwise a negative response was to be made. The second condition was just a reversal of the first; a positive response was to be made if the other dimension attained the proper value, otherwise a negative response was required. The third condition required a positive response only if both dimensions were at the proper value, otherwise a negative response was to be made. The nonparametric tests were developed relating the way the negative time distribution in the third condition could be serially or parallelly composed of those negative response-time distributions from the other two conditions. Thus, an implicit assumption is that the negative processing distributions remain the

same in all three conditions. This assumption is not too unlikely from an intuitive standpoint if processing is serial but there may well be changes in capacity requirements when the number of stimulus dimensions to be attended to in parallel is altered. One of the prime sources of difficulty in parallel–serial testability has been the intuitively acceptable ability of limited capacity parallel models to mimic standard serial models in just such experiments (see, e.g., Townsend, 1974). We finally remark in these connections that the consequences for Thomas' (1969) results have apparently not been investigated when a residual latency variable (analogous to t_p, t_s) with nonzero variance is convolved with the processing time variables.

As far as I can determine, the present development, cultimating in Theorem 5, represents the first case of an experimental paradigm that can be shown to analytically separate fairly general parallel serial models at the level of observable mean reaction times. To appreciate the generality of the models one merely has to note that there are only two more degrees of freedom in the data than there are parameters in the parallel model, PEM, and only one more than the serial, SEM; or that one can predict positive, negative or no dependence between the processing of individual stimulus elements (that is, between the comparisons of S_1 with the two members of S_2). It is axiomatic that any simpler submodels of PEM and SEM are a fortiori also testable, with the constraint that $\lambda^+ \neq \lambda^-$ or $\nu^+ \neq \nu^-$.

One type of generalization of the above theory would take the direction of letting the rate on one comparison depend on whether another comparison is a match or mismatch. For example, we could write ν^{+-} as the parallel rate for the pair with serial position 1 when it is a + match but the pair in serial position 2 is a − match. The full class of PEM and SEM extensions of this sort are not observable mean diverse with respect to PST, but there are "Gestalt" subcases where the models are still testable (Townsend & Snodgrass, 1974).

Finally, a cautionary note should be sounded concerning the possibility that the converging conditions I, II and III of PST may induce processing of a sort qualitatively different in the three cases (e.g., exhaustive processing in one, self-terminating in the others, etc.). This is a question that must ultimately be referred to empirical test in any given cognitive context.

References

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

Brand, J. Classification without identification in visual research. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 178–186.

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

Egeth, H., Jonides, J., & Wall, S. Parallel processing of multielement displays. *Cognitive Psychology*, 1972, **3**, 764–698.

EGETH, H. Parallel versus serial processes in multidimensional stimulus discrimination. *Perception and Psychophysics*, 1966, **1**, 245–252.

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

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

GREENO, J. G., & STEINER, T. E. Comments on "Markovian processes with identifiable states": General considerations and applications to all-or-none learning. *Psychometrika*, 1968, **33**, 169–173.

HABER, R. N. *Information processing approaches to visual perception*. New York: Holt, Rinehard & Winston, 1969.

HURWICZ, L. Generalization of the concept of identification. In T. C. Koopmans (Ed.), *Statistical inference in dynamic economic models*. New York: Wiley, 1950.

HUREWICZ, W., & WALLMAN, H. *Dimension theory*. Princeton: Princeton University Press, 1948.

LOEVE, M. *Probability theory*. Princeton: Van Nostrand, 1955.

MCGILL, W. J., & GIBBON, J. The general-gamma distribution and reaction times. *Journal of Mathematical Psychology*, 1965, **2**, 1–18.

PYKE, R. Markov renewal processes: Definitions and preliminary properties. *Annals of Mathematical Statistics*, 1961, **32**, 1231–1242. (a)

PYKE, R. Markov renewal processes with finitely many states. *Annuals of Mathematical Statistics*, 1961, **32**, 1243–1259. (b)

RUMELHART, D. E. A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, **7**, 191–218.

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

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

TOWNSEND, J. T. Mock parallel and serial models and experimental detection of these. *Purdue Centennial Year Symposium on Information Processing*, 1969, **II**, 617–628.

TOWNSEND, J. T. Theoretical analysis of an alphabetical confusion matrix. *Perception & Psychophysics*, 1971, **9**, 40–50. (1)

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.

TOWNSEND, J. T. Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Potomac, Md.: Erlbaum, 1974.

TOWNSEND, J. T. Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology*, in press.

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., & SNODGRASS, J. G. A serial vs. parallel testing paradigm when "same" and "different" comparison rates differ. Paper presented to Psychonomic Society, Boston, November, 1974.

RECEIVED: December 10, 1974

Printed by the St Catherine Press Ltd., Tempelhof 37, Bruges, Belgium.