

## THEORETICAL AND REVIEW ARTICLES

---

# The serial–parallel dilemma: A case study in a linkage of theory and method

JAMES T. TOWNSEND

*Indiana University, Bloomington, Indiana*

and

MICHAEL J. WENGER

*Pennsylvania State University, University Park, Pennsylvania*

The question as to whether humans perceive, remember, or cognize psychological items simultaneously (i.e., in parallel) or sequentially (i.e., serially) has been of interest to philosophers and psychologists since at least the 19th century. The advent of the information-processing approach to cognition in the 1960s reopened the inquiry, initiating a flood of experiments and models in the literature. Surprisingly for so elemental an issue, persuasive experimental tests have, until recently, proven rather elusive. Several decades of theoretical, methodological, and experimental effort, propelled and shaped by a meta-theoretical perspective, are leading to powerful strategies for assessing this and related cognitive issues. The present article reviews the theoretical and empirical history of these inquiries and details situations in which decisive experimental tests are possible.

In 1969, the first author presented his initial treatment of the question of whether it is possible to distinguish serial from parallel processing (Townsend, 1969). That same year, a popular song lyric provided a prescient summary of what is now a wide-ranging literature on this issue: “*You can’t always get what you want . . .*” (Richards & Jagger, 1969). Indeed, in the years that have followed, many challenging problems associated with the theoretical and empirical distinction between serial and parallel processing have been documented; yet many significant advances have been made. The goal of the present review is to collect and showcase many of these advances, with particular attention being paid to approaches that, by intent, provide a tight coupling of theory with experimental and statistical methods.

Our treatment of these issues will begin with some context: In particular, we will provide a brief description of the intellectual perspectives (or pretheoretical commitments, as in Lachman, Lachman, & Butterfield, 1979; Murdock, 1974) that we bring to the discussion. We then will move on to a general consideration of the problem of discriminating between serial and parallel processing, an

issue that is sometimes referred to as a problem of *model mimicry* (e.g., Townsend, 1971a, 1972; Van Zandt & Ratcliff, 1995). After this, we will move on to the core of our review, structuring that review around three global assertions: (1) The question will not, and *should not*, go away; (2) the question *cannot* be understood or answered out of context; and (3) the question *can* be answered using approaches that provide a clear and principled connection between theory and method.

### THE INTELLECTUAL CONTEXT AND THE GENERAL PROBLEM

When we discuss serial and parallel information processing, we are considering hypotheses regarding the nature of the *information-processing mechanism*. This statement captures the essential elements of the theoretical perspective we bring to this review. That is, we approach the question of serial versus parallel processing explicitly in terms of psychological mechanisms (codings, transformations, etc.) that we hold to be involved in human information processing. Mechanistic concepts and explanations have existed for centuries, from the Greeks to modern philosophy, in tracts and speculations about how humans perceive, think, and act. In the mid-19th century, scientists and scholars from physics, medicine and physiology, and philosophy began to attempt behavioral experimentation with people as subjects and with inferences regarding psychological mechanisms as a goal. Many of

---

The writing of this article was supported in part by National Institutes of Mental Health Grants 1 R01 MH57717-01 (to J.T.T.) and 1 R03 MH59845 (to M.J.W.). Correspondence concerning this article should be addressed to J. T. Townsend, Indiana University, Department of Psychology, 1101 E. 10th St., Bloomington, IN 47405-7007 (e-mail: townsend@indiana.edu).

these scientists (including Hermann von Helmholtz, Gustav Fechner, Heinrich Weber, and Ernst Mach) were established and renowned in their own disciplines; yet in the work they pursued, they occupied themselves, perhaps unwittingly, with the creation of experimental psychology.

The most successful enterprises of that period, in terms of long-term impact on experimental psychology, typically involved relatively simple perceptual and mental activities. These psychological activities were studied through the use of performance measures, such as accuracy and response times (RTs). A researcher especially pertinent to this review was a Dutch physiologist by the name of Donders. Donders believed that he could uncover the durations required by various mental processes through his *method of subtraction*. The method of subtraction was based on the idea that complicated mental activities were compounded in a simple sequential fashion from less complex parts (as described in Donders, 1868/1969), with mean RT being used to estimate the duration of each of the component durations. For instance, the scientist might engage the subject in a task requiring both perception and decision and compare mean RT in that condition with mean RT in a task requiring only perception. The difference in mean RTs was used to estimate the mean duration of the decision operation.

As experimental psychology moved into the 20th century, the slowly dying introspective method in both the phenomenological (e.g., Brentano and, later, Külpe, Brown, etc.) and the structural (e.g., Wundt, Titchener, etc.) schools was replaced by a functionalist approach engendered by Darwinism and fostered by William James and, later, by such psychologists as Woodworth and Thorndike. The functionalist philosophy nourished the nascent behaviorism of Watson and, soon, Skinner, Hull, and Tolman.

After what some modern cognitive scientists regard as an inordinately long and arid infatuation with behaviorism, psychology after World War II began to return to a frank interest in mental operations. To many psychologists, some early cognitivists, particularly those whose background lay primarily in computer science, engineering, or philosophy, seemed to stereotype all behaviorism into a sterile kind of scapegoat. Certain phases of Skinner's approach may have approached this stereotype, but later behaviorism—in particular, the researches of Tolman and Hull—were quite cognitive in aspect. In fact, they gave high credence to motivational aspects of behavior, a gaping lacuna in cognitive science that has been criticized and only recently begun to be addressed, largely through the influences of neurocognition and social/personality psychology.

The notion of humans as information processors undoubtedly had its birth in the 1950s. Concepts and quantitative tools from pure and applied mathematics, electrical and communications engineering, the emerging field of computer science, and cross-disciplinary efforts in neurobiology and mathematical modeling began to have profound impacts on the social and biological sciences. Stand-out examples include information theory, cybernetics and

control theory, automata, the theory of games and decision strategies, signal detection theory, and models of perception and elementary thought based on neural-like concepts and mechanisms. All of these examples are concerned, at some level of analysis, with transformations or computations involving various forms of information.

The approaches we discuss in the present review are consistent with the *information-processing approach* to human psychology. Although that label is typically applied to a rather loose array of concepts, the ones selected for review here are elemental, fundamental, or foundational, in the following sense. The design of almost any system intended to carry out the processing of a finite number of inputs—be it biological, mechanical, electro-mechanical, electronic, and so on—would require the designers to make choices regarding the set of issues we will review here. In this sense, the issues we will discuss can be thought of as mapping to a set of *computational* issues in the analysis of any information-processing system (see Marr, 1982, chap. 1). In addition, these issues are somewhat unique in having been subjected to general theoretical, quantitative, and methodological analysis—perhaps more than any other such concepts in the field (see Townsend, 1974, for the first quantitative taxonomic treatment of the collective issues). They are also central in investigations of the serial-parallel issue. They are logically independent, in the sense that any combination of selections on each dimension (e.g., serial, unlimited capacity, dependent, self-terminating) is logically possible. Nevertheless, the behavior of a system, as captured by various statistics, can depend on interactive properties of the various dimensions. Hence, all the issues must be deliberated on together.

Over the past several decades, considerable progress has been made in the information-processing tradition by linking formal mathematical models of the concepts under study to experimental designs capable of providing the data needed to distinguish the theoretical alternatives. The general strategy involved in accomplishing this linking involves the mathematical development of what we will refer to as *meta-theories* or *meta-models*. These meta-models are developed by specifying—typically, using a single theoretical “language” (such as stochastic process theory, linear systems theory, etc.)—distinct *sets* of models, with each set obeying diverse psychological principles. Predictions are then generated for the contrasting *classes* of models. The sets of predictions are then examined to determine the experimental conditions and dependent measures that can allow one or more of these classes to be falsified.<sup>1</sup> The goals of this methodology are similar to those of *strong inference* (see, in particular, Platt, 1964; Popper, 1935). It could be considered to be a mathematically rigorous extension of the verbal logic psychologists often employ in offering experimental designs to probe or test an issue. In principle, the strengthened approach can lead to a series of theoretical and methodological sieves that allow one to develop increasingly precise and successfully predictive theories. It

should also be differentiated from a related rigorous strategy, that of model fitting, which requires parameter estimation and, where possible, statistical tests of the model fit. Of course, these two approaches are complementary, not mutually exclusive (see, e.g., R. D. Thomas, 2001). For instance, the meta-theoretical approach might be used in the early stages of a research program to falsify large classes of models and provide focus on a narrower class, which is then pared further down by way of specific model fitting. The present review centers on the meta-modeling tactics that have been developed over the past several decades for application to the parallel-serial distinction and related issues.

### The Problem of Model Mimicking

The issue that is perhaps most widely associated with the distinction between parallel and serial processing is the problem of *model mimicking*.<sup>2</sup> Model mimicking refers to the ability of one class of models of a psychological process (e.g., models that assume parallel processing) to make the same predictions as other classes of models (e.g., models that assume serial processing) on the basis of strikingly different psychological assumptions (see definitional discussions in Bamber & van Santen, 1985, 2000). Some of the early attempts to settle the question did indeed adroitly test certain types of parallel and serial models (e.g., Egeth, 1966; Sternberg, 1966). However, it turned out that the methods used in those studies also failed to test other models that might have explained the same set of data (e.g., Atkinson, Holmgren, & Juola, 1969; Murdock, 1971; Townsend, 1971a). Exploration of these possibilities—primarily by way of mathematical modeling—led to the conclusion that representations of diametrically opposed psychological principles could sometimes mimic one another, even to the point of mathematical equivalence of their defining probability laws (Townsend, 1971a, 1990a).

For example, consider models for the processing of a set of  $n$  inputs under fixed experimental conditions (i.e., all of the models will address the same experimental task under identical conditions). Assume that (1) serial processing of these  $n$  inputs is represented by probability mixtures of convolutions of the distributions of the individual processing times and (2) parallel processing is represented by joint distributions on processing times. Assume nothing else, and confine each set of assumptions so that the latitudes (e.g., allowing the joint serial density to vary with future processing order) of each are the same. Then the parallel and serial classes of models are equivalent, in the sense that mappings can be provided that homeomorphically (i.e., one to one, continuous both ways, onto all of the members of each model, in the two sets of models) carry one distribution (joint density) for total processing times for the given condition into the other (Townsend, 1971b, 1976a, 1976b; Townsend & Ashby, 1983; Vorberg, 1977).

Given classes of models for a specific experimental task that are defined in terms of families of probability

distributions, it turns out (perhaps counterintuitively) that serial models are usually more general than parallel models for a single experimental condition. That is, parallel models are mathematically contained within the class of serial models (see, e.g., Townsend, 1976a): Given a parallel RT distribution, there is always a serial distribution that is equivalent, but the converse is not true. However, situations can arise in which the parallel class of models is more general than the serial class, in the sense of predicting less invariance across certain experimental conditions (see, e.g., Townsend & Ashby, 1983, chap. 13). Despite these “nested” relationships, accumulating theoretical results have demonstrated that if one is willing to make further restrictive but still very reasonable and general assumptions about the models, the two classes of models can, in fact, be empirically distinguished (Ashby, Tein, & Balakrishnan, 1993; Townsend, 1990a; Townsend & Ashby, 1983; Townsend & Nozawa, 1995), as we will illustrate in the sections that follow. It is worth emphasizing, at this point, that it is the link between theory and experimental methodology, provided by an explicit model for the experimental task, that allows for the possibility of deciding this question.

### THREE GLOBAL ASSERTIONS

In this portion of the review, we will take up the implications of the preceding paragraph in detail. We will structure this discussion by considering three assertions that we believe capture the central conceptual issues related to distinguishing serial from parallel processing.

#### The Question Will Not, and Should Not, Go Away

The question of serial versus parallel processing is one that has come up repeatedly in the history of scientific psychology. In fact, we would argue that although its role as a focal question has varied, it has always been present in the construction and evaluation of theories of human information processing. Consequently, the issue has been considered in a number of reviews, including two by the first author (Townsend, 1974, 1990a), the latter being intended to be useful for attaining a concise view of the situation at that time. In addition, a trenchant recent review of serial and parallel processing (Logan, 2002) has approached the issues from an alternative and complementary perspective.

Deciding whether any two or more subprocesses operate simultaneously or in sequence is absolutely essential to understanding how the overall system functions, whether we think of the subprocesses as physiological or anatomical modules (as, e.g., in Kanwisher, McDermott, & Chun, 1997; Kanwisher, Stanley, & Harris, 1999) or as more abstract entities (e.g., Sternberg, 1998). The essential character of the question is easily understood if we allow for the fact that, at multiple levels of analysis, psychological experience is determined by more than one input (e.g., temporal and nasal hemifields, left and right eyes,

left and right ears, physical and semantic features, etc.). If there are two or more inputs, a critical question is whether those inputs are processed one at a time or at the same time. Any theory that attempts to address how those inputs are used (i.e., any theory that concerns itself with psychological mechanisms), if it is to be fully specified at any level (computational, representational/algorithmic, or implementation, as in Marr, 1982; Anderson, 1990, is also relevant), needs to "take a stand" on the issue of serial or parallel processing. Thus, if it is possible to find a way to solve the problem of model mimicking (and we will argue it is indeed possible), it becomes possible to falsify a theory on the basis of a central assumption.

This ability to falsify a theory on the basis of a core assumption is no small matter. It is very frequently the case that psychological inquiries rely on evidence that *appears consistent* with an assumption, without coming from a direct test of that assumption. In this sense, much of the evidence used in psychological theory building has a status equivalent to that of *indirect*, or *circumstantial*, evidence in legal theory. In contrast, the methods we describe and advocate can be thought of as providing *direct evidence*. Thus, although it is possible to build a case for an assertion on the basis of indirect or circumstantial evidence, we would suggest that psychological science has a better chance of advancing if assertions are supported by direct evidence, such as those provided by direct tests of core assumptions. In fact, we contend that direct evidence is necessary for strong-inference hypothesis testing.

The logical necessity of considering the question of serial versus parallel processing is highlighted in a number of active areas of inquiry in contemporary cognitive science. The pivotal topic of attention provides an important and natural illustration. For example, several important accounts of visual search assume that there is an early preattentive stage in which processing of inputs occurs in parallel (e.g., Duncan, 1984; Palmer, 1995, 1998; Treisman & Gormican, 1988). A secondary stage is suggested to involve serial focus and switches of attention, especially when the search is for a conjunction of dimensions (e.g., a red square among red triangles and green squares). Note that in this example, the distinction between parallel and serial processing needs to be invoked at two levels of analysis: within a stage (parallel for the first, conditionally serial in the second) and between (an implied serial or sequential arrangement).

### The Question Cannot Be Understood or Answered Out of Context

The question of serial versus parallel processing concerns the *architecture* of the information-processing system. And the question of architecture is one of four logically distinct issues that must be addressed in specifying almost any information-processing system (Townsend, 1974; Townsend & Ashby, 1983; Townsend & Nozawa, 1995): (1) the question of architecture, (2) the question of a stopping rule, (3) the question of independence, and (4) the question of capacity. Although these four questions

are logically distinct, systematic formal investigation over the past 3 decades has revealed some very critical interdependencies in the ability to empirically determine an answer to *any one* of the questions in isolation. In this section, we will consider these four questions and their interdependencies as they pertain to empirically distinguishing serial and parallel processing.

In order to illustrate a number of the issues we will be discussing, it is useful to consider an example experimental task. The task we choose for this purpose—visual and memory search—is one in which the question of serial versus parallel processing has been considered repeatedly and in some detail, beginning with several pioneering studies in the 1960s (e.g., Atkinson, Holmgren, & Juola, 1969; Egeth, 1966; Sperling, 1960; Sternberg, 1966). The version of this task that has been used in the study of brief-duration memory requires the retention of a small number of items (e.g., randomly arranged letters or digits) for anywhere from a few seconds to several minutes. In this paradigm, a varying number of items is presented to the experimental subject, auditorily or visually. This collection of items is referred to as the *memory set*. Following the presentation of the memory set, a specified amount of time (the retention interval) is allowed to elapse. At this point, the subject is presented with a *probe item* and is required to indicate, as quickly and accurately as possible, whether or not the probe was contained within the memory set. RT is the primary dependent variable of interest here, although a substantial history of work with accuracy also exists (e.g., Ashby, 1992; Ashby & Townsend, 1986; Kadlec & Townsend, 1992a, 1992b; Laming, 1973; Link & Heath, 1975; Luce, 1986; Massaro, 1987; Ratcliff, 1978; E. A. Thomas, 1971; R. D. Thomas, 1995; Townsend & Ashby, 1983; Vorberg & Ulrich, 1987; all of these contain pertinent discussions and reviews).

**The question of architecture.** The question of architecture refers to the temporal structure of the process of comparing the probe item with the memory items, with the two alternatives being serial and parallel. Two hallmark characteristics of many early information-processing models were (1) *seriality* and (2) *discrete stages* of operation. In these models, each component process (e.g., each of the  $n$  possible comparisons) could operate only when the preceding processes were completed, and only one process could be performed at a time. Thus, strict seriality in these models connoted that no overlap of processing times occurred, in addition to the sequentiality of starting times. The latter constraint is referred to as *discrete flow*. The so-called standard serial model (e.g., Townsend, 1974, 1990a) assumes, in addition to discrete flow and sequentiality, that the average processing times (i.e., for the comparison of probe with memory items) are equal over all display positions, item identities (e.g., target vs. distractor), processing positions (i.e., place in the order imposed by a processing path through the memory items), and memory set sizes (often referred to as *load*). This assumption is usually not stated explicitly. When actual mathematical models are built to satisfy

this *invariance of speed* constraint, it is usually also assumed that the individual processing times are independent of one another and that each probability distribution on processing times is, in fact, identical to every other one (thus implying the equal mean individual processing times). Two points are of interest with respect to this model. The first is that, even at this level of specification, addressing the question of architecture requires consideration of the question of independence. Second, although very interesting and natural, this model is only one instance from the array of possible serial models.

The alternative architectural possibility—that all the items might be searched simultaneously, or in parallel—was considered very soon after the serial models were first proposed. The class of parallel models investigated by Sternberg (1966) is a natural example of parallel processing (e.g., Townsend, 1974; Townsend & Ashby, 1983), just as the standard serial model is a natural example of serial processing. In addition to assuming simultaneity, the standard parallel model is defined by probabilistically independent processing times for the various simultaneous item comparisons and (even more important) unlimited capacity for the separate processing time distributions. As we shall discuss in more detail in subsequent sections, the assumption of unlimited capacity means that the probability distributions (and thus all other statistics; see Townsend, 1990b) for the processing times for each item will be invariant across any changes in the total number of items being processed. The form of a standard parallel model will differ greatly depending on whether processing can cease when the probe is located. Note here that the question of architecture is bound to the questions of independence (just as was true for the serial model) and the stopping rule.

Naturally, arrangements other than serial and parallel are possible. For example, it is possible to conceive of a system's being serial part of the time and parallel part of the time. In addition, the basic flow diagram of a system's operations might be unchanged, but the assumption of discrete stages might be relaxed. When operations can feed into other components, but with overlapping processing times, the system is said to be a *continuous flow*, rather than a *discrete flow*, system (e.g., Colonius & Townsend, 1997; McClelland, 1979; Townsend & Fikes, 1995).

**The question of a stopping rule.** The question of a stopping rule refers to the conditions under which the system can cease processing and generate a response. Depending on the task, it may or may not be necessary for the subject to process all of the items in order to make a correct response. In the memory search paradigm, for example, if the probe is present in the memory set, processing can in principle cease as soon as the probe is encountered; in this case, processing is characterized as being *self-terminating*. However, on probe-absent trials, it is necessary to process all of the memory items in order to correctly make a *no* response; in this case, *exhaustive* processing must occur. In some experimental designs, all the items are probes (not necessarily physi-

cally identical; for instance, the “probe” might be specified as “any vowel,” and all memory items might be vowels). The latter case permits the possibility of *first-terminating* or *minimum-time* processing, to be explored further momentarily.

Interestingly, mathematical characterizations of this task have demonstrated that the traditional plot of RT as a function of experimentally manipulated search set size, although problematic with regard to the issue of architecture, is quite powerful with respect to the issue of a stopping rule (more on this point below; see Townsend & Van Zandt, 1990).

In the case of the standard serial model, if processing is self-terminating with a single target, then on average, the target (i.e., the item matching the probe) will be found halfway through the list, and hence, the slope of the mean RT function of load will be half that of the probe-absent trials, which demand exhaustive processing (Sternberg, 1966; see Figure 1). Note that this prediction is one that depends on answers to the questions of a stopping rule and of architecture. However, it is also known that there are reasonable parallel models that mimic this prediction (Atkinson et al., 1969; Murdock, 1971; Townsend, 1971a). Nevertheless, many investigators, although acknowledging this fact, advise accepting the inference of serial processing (e.g., Schneider & Shiffrin, 1977; Treisman & Gormican, 1988), even in the absence of direct decisive evidence. We have argued that this strategy is flawed, since it blurs the important distinction between direct and circumstantial evidence and leads to the unquestioning acceptance of inferences based on circumstantial evidence.

Nonetheless, with regard to the stopping rule issue, a more satisfying result is obtained. Quantitative analysis of the 2:1 slope ratio phenomenon reveals that it, as well as far less stringent slope ratios, falsifies all exhaustive models, parallel as well as serial. In fact, we have demonstrated that negative-to-positive slope ratios that deviate quite small degrees from unity (and certainly those in the 2:1 range) cannot be accommodated by exhaustive models: We proved that the combination of any slope ratio higher than  $n:n - 1$  and/or strong position effects (or an approximation to these) effectively rule out exhaustive processing in all serial and parallel models. (Again, note the intertwining of the questions of a stopping rule and of architecture.) Thus, an experimental finding that provided weak evidence with regard to architecture, at best, affords powerful falsification with respect to the stopping rule. In addition, equal slopes are easy for self-terminating serial or parallel models to predict. So unequal slopes falsify exhaustive processing, but equal slopes can be predicted by both stopping rules. Proofs, discussion, and some application of this and the following theoretical and methodological results have appeared in Townsend and Van Zandt (1990), Van Zandt and Townsend (1993), and Townsend and Colonius (1997).

The other strong theoretical result begins with an observation. Consider any standard serial model operating under self-termination. Assume that processing is orga-

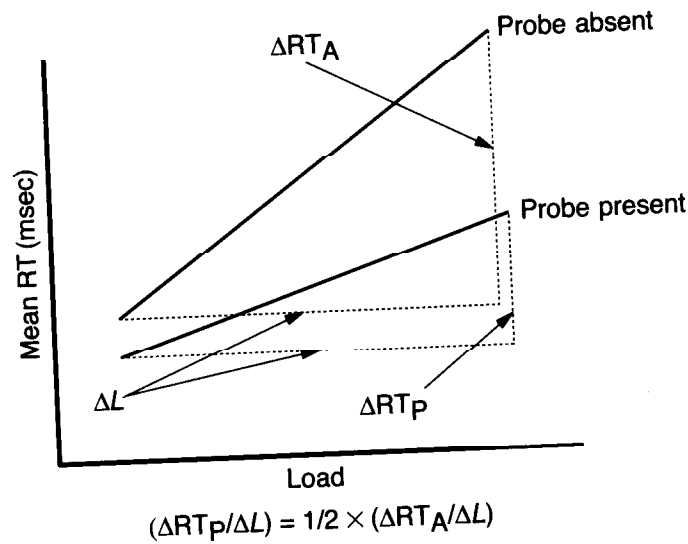


Figure 1. Typical relationship between the slope of the function relating mean response time (RT) to processing load on probe-present trials and probe-absent trials.  $\Delta RT_P$ ,  $\Delta RT_A$  = change in RT for the probe-present and probe-absent trials, respectively;  $\Delta L$  = change in processing load.

nized so that the items (indexed by, say, location) are operated on in exactly the same sequence on each trial, and assign Serial Position 1 to the first location processed, 2 to the second and so on. Then target-present RTs, plotted as a function of serial position, must fall precisely on the target-absent RTs as a function of  $n$  (e.g., Bricolo, Gianesini, Fanini, Bundesen, & Chelazzi, 2002; Townsend & Van Zandt, 1990). The reasoning is clearly that for, say, the third position, observers will always process the target third and, hence, their RT must equal that for  $n = 3$  on target-absent trials, and so on. We have referred to this as a *prediction of strong position effects* (Van Zandt & Townsend, 1993), and exhaustive models of either the serial or the parallel variety simply cannot make that prediction.<sup>3</sup>

Another approach to addressing the question of a stopping rule has to do with the minimum RT, although here again we will see that the question inevitably connects with the question of architecture. Consider first the possibility of serial (architecture) self-terminating (stopping rule) processing. Here, the idea is that in serial processing, if at least one target is present and a self-terminating stopping rule is employed, the minimum processing time in a set of trials should usually come from the probability distribution for the first item processed, since on some trials the target will be processed first. Hence, as the load,  $n$ , is increased, the intuition is that the minimum RT will stay more or less the same. Even if an observer processes item positions in the same order on each trial, if the experi-

menter randomizes the target position, the foregoing assumption will hold. On exhaustive trials, such as those in which no targets are present, it would be expected that the minimum time will increase (on average) as  $n$  grows.

This prediction undoubtedly holds in some sense, but so far we have not seen it established or applied in a rigorous manner. Thus, an oft employed tack is to simply take the smallest RT for the condition for a specific  $n$  and compare that across  $n$ . Naturally, a sample of a single number has an extraordinarily large variance and is probably difficult or impossible to test rigorously. However, we offer some ideas about how the investigator might proceed: In the case of serial exhaustive processing with load  $n$ , the formula for the survivor function for the minimum time in  $k$  trials will be

$$S_{\text{MIN,EX}}(t, k) = [1 - P(T_1 + T_2 + \dots + T_{n-1} + T_n \leq t)]^k.$$

From this, one can immediately work out the density function and the distribution function, compute the moments, and so forth. Similarly, in the case of serial self-terminating processing, the survivor function for the minimum time in  $k$  trials will be the equation at the bottom of this page. It can be seen that the probability that the target is processed first will decrease, in general, as  $n$  gets larger, thus affecting the distribution on the minimum (as was noticed, e.g., by Bricolo et al., 2002). Once the properties of the distributions are known, a number of established strategies for data analysis can be utilized. One approach would be to break the trials from a certain

$$S_{\text{MIN,ST}}(t, k) = \left[ 1 - \sum_{i=1}^n P(\text{Target is processed } i\text{th}) P(T_1 + T_2 + \dots + T_i \leq t) \right]^k.$$

$n$ -condition down into smaller sets, estimate the minimum from each set, form an overall statistic (e.g., median, mean, etc.), and compare across  $n$  values. Note that this would allow the question of a stopping rule to be adjudicated under the assumption of a specific processing architecture.

Finally, let us compare the predictions made by a standard parallel model with those made by the standard serial model, under the three major stopping rules. Recall that standard parallel processing is defined as being parallel, with unlimited capacity, and with stochastically independent channels. In the case of the presence of a single target among  $n - 1$  distractors, and assuming self-termination, the prediction is that mean RT will be a flat function of  $n$ , a prediction that requires super-capacity (to be defined precisely below) for any kind of exhaustive processing (assuming each channel possesses nonzero variance). If there are  $n$  targets on *yes* trials and no distractors, parallel models will, except when capacity is very limited, predict a decrease in RT with increasing  $n$ . A pioneering study using this reasoning was Egeth (1966). Again, any kind of exhaustive model has to be super-capacity to predict such effects (e.g., Townsend, 1974). In the instance of exhaustive processing, we remind the reader that mean RT curves as  $n$  varies always are increasing, concave down. Note here how all four questions intersect in determining empirical discriminability.

Unlike the situation with parallel predictions, recall that standard serial models naturally predict increases in the ordinary, single-target, self-terminating case, as well as with exhaustive processing in the cases above, when  $n$  increases. When all the items are targets, so that RTs can be minimum time (first terminating), the RT curve as  $n$  varies is flat, since as soon as the first item is processed, search can cease. Note, then, that even self-terminating serial models must be super-capacity to make predictions that are compatible with the predictions of the standard parallel models. Usually, super-capacity is counter to our intuitions and falsifies the super-capacity models, when they must be such in order to accommodate the data. Thus, as happens rather frequently in model testing, there is an asymmetry in how results affect our interpretation of psychological issues. In this instance, standard serial self-terminating predictions support self-termination yet say little about architecture, but standard parallel kinds of predictions are quite unlikely for serial models.

**The question of independence.** The question of independence refers specifically to the relationships among the *rates* of processing for each of the inputs. For instance, in parallel models, if processing is independent, the probability of two items being completed during an interval is the product of the probabilities of each being completed in that interval. In serial models, independence means that the processing times of each succeeding item are independent. A *violation* of the assumption of independence in a parallel system could be expressed as a positive or negative correlation of rates—for instance,

through lateral interference or mutual facilitation. Both parallel lateral inhibition and facilitation are found at all levels of the nervous system, including the retina (e.g., Bridgeman, 1971; Francis, 2000). A violation of independence in a serial system could occur if the state of the output of a preceding unit or subsystem affects the speed of the succeeding subsystem. For example, a serial subsystem might sometimes work harder on its own input, which could increase the processing time required for that input, thus taking longer but providing a more “finished product” for the next subsystem or stage, which can proceed faster. This system would evidence a negative dependence in the sequential-processing times. In both cases, the preservation or violation of independence in rate will be related to the question of *capacity*.

Independence is a term that naturally expresses the concept of a *lack* of a systematic relationship between or among things, and as such, it is probably used in science and mathematics more than any other concept and for myriad purposes. In many cases, other terms, such as *invariance*, might be preferable. Nevertheless, one very important case concerns the variable representing the completion time for the processing of an input. To consider this in detail, we must distinguish among the notions of intercompletion time, total completion time, and actual processing time. Figure 2 provides a schematic representation of these concepts.

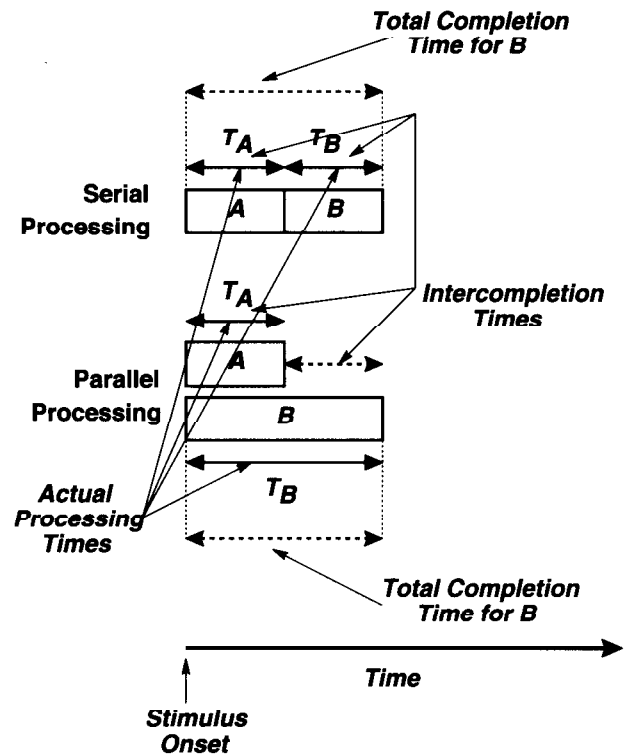


Figure 2. An illustration of the concepts of intercompletion, total completion, and actual processing times for serial and parallel processing of two inputs, A and B.

It is important to first understand that intercompletion times and total completion times are descriptive, apply to any system, and are potentially observable (although most psychological experimental observations are too coarse to permit recording these individual events), whereas the actual processing times will refer to what is occurring in a particular architecture. Figure 2 represents the serial and parallel processing of two inputs, labeled as *A* and *B*, and for one particular order of processing (for the serial model) and one possible relationship between processing times (for the parallel model). Consider first what happens in the case of serial processing. The times required to process each of the inputs (elements, items, etc.) in a serial system are  $T_A$  and  $T_B$ , respectively (these are the *actual processing times*). Since these times are also the times between the successive completions, these times can be considered as both processing durations and intercompletion times. Now consider what happens in the case of parallel processing. The times required to process each of the inputs are  $T_A$  and  $T_B$ , but these processing durations are not the intercompletion times, except for the first element finished. The second intercompletion time in parallel processing is given by the difference between  $T_B$  and  $T_A$ . The total completion time for an input is simply the total time from the start of a trial ( $t = 0$ ) until that input is finished. Note that "total" does not refer to the duration taken by everything until processing ceases; rather, it means the time from the start of any processing until the specified item is completed. Hence, the total completion time for all of the inputs is identical to the actual processing time in a parallel system, since all inputs begin processing immediately at time  $t = 0$ . However, in a serial system, the total completion time for all of the inputs is equal to the sum of serial actual processing times (and hence, also intercompletion times) from  $t = 0$  until the processing of all of the input is finished.

Now let us consider the issue of independence with respect to the intercompletion times for serial and parallel models. As might be intuited from Figure 2, this can be a little tricky if one does not first tie down the architecture. Consider first the time event represented by the duration measured from  $t = 0$  to the time point at which the processing of an input is complete. Independence on intercompletion times is logically independent of independence on total completion times. In general, models that posit independence for successive intercompletion times (the actual processing times for a serial system) do not have to yield independence on the total completion times for the various inputs, which are the actual processing times for a parallel system, and vice versa. For instance, a standard serial model, based on sequentially independent, exponentially distributed processing times, predicts positively correlated total completion times (e.g., Townsend & Ashby, 1983). Furthermore, a parallel model with independent processing times (i.e., independent total completion times) will typically not give rise to independent intercompletion times. Of course, there are

models that encompass both independent intercompletion times and independent completion times. A prime example is found when processing on all items is parallel and exponential, with rates that do not vary within a trial but can be different in distinct channels.

Now we will discuss the actual processing times for the parallel processing model, which are equivalent to the total completion times in any architecture. A major conundrum in testing for independence is that, with only one response employed in most experiments, mean RT does not typically deliver any measure of possible dependence on two or more channels. If, on the other hand, we allow each channel to carry distinct signals (i.e., we assume two distinct inputs to be processed), with one of two responses called for from both channels on each trial (i.e., the output requires the processing of both inputs), we can test for independence. However, the result could well be telling us about the response selection, rather than about more sensory or central processes (e.g., Pashler, 1994; Schweickert, 1993). Yet another obstacle is the residual time that represents all the processes before (e.g., sensory registration) and after (e.g., motor output) the ones under study (Sheu & Ratcliff, 1995; Smith, 1990). Even if the residual duration is independent of the durations of the processes under study (called into question, e.g., by Dzhafarov, 1992), its presence obscures the dependence relation between two parallel channels.

It is possible to test independence if one makes other assumptions—specifically, unlimited capacity (as might already be apparent, the questions of independence and capacity are closely related when one envisions behavioral consequences; see, e.g., Townsend & Wenger, in press). If one assumes unlimited capacity, parallel processing, and independence, one can test predictions as  $n$  is varied (e.g., Goldstone, 2000). Yet even here, it is necessary to take care of the residual time problem. For instance, Goldstone used a simple detection task to estimate the residual time distribution, then deconvolved the observed RT distribution into a convolution of the residual time distribution plus the newly estimated psychological (process of interest) distribution. Assuming independent parallel processing, unlimited capacity, and that all the channels were alike (i.e., no position effects), Goldstone was in a position to make predictions for the RT distribution as  $n$  varied.<sup>4</sup>

If we posit an architecture to begin with, independence of certain types can be tested with accuracy experiments, as in the experimental methodology connected to general recognition theory (Ashby & Townsend, 1986). However, two signals could be processed independently as far as accuracy goes, with either a serial or a parallel underlying architecture! One inference is that although independence in measures of response accuracy (e.g., probability correct on two items being equal to the product of the probabilities of being correct on either item alone) is often invoked in sensory psychophysics as implying independent parallel processing, it could just as well be a serial arrangement that gave rise to such results.



**The question of capacity.** The meaning of capacity is to be understood in the context of a system that must accomplish a task (i.e., that has to perform work). Globally and roughly, *capacity* refers to efficiency in performing the work required by the task. Our interpretation of capacity in terms of *efficiency* (Townsend & Ashby, 1978) is a construct that has a crucial place in understanding attention (e.g., Gibson & Amelio, 2000; Wenger & Gibson, in press). Of interest is the extent to which attentional capacity can vary with effort (e.g., Kahneman, 1973), with the goals of the observer (especially as they interact with stimulus characteristics; see, e.g., Folk, Remington, & Johnston, 1992; Gibson & Kelsey, 1998; Theeuwes, 1994; Theeuwes, Atchley, & Kramer, 2000), and with the state of the observer (e.g., exhaustion or lethargy).<sup>5</sup>

The issue of efficiency can be raised at a number of levels. The most important usage, and one that relates to the observable RT measures, is concerned with measures of efficiency or processing speed as the workload is varied. For instance, varying the memory set size in the Sternberg task is a manipulation of workload, and we can ask how processing efficiency is affected. In a coarse sense, if the workload increases and performance suffers (i.e., RT goes up, accuracy goes down), we should infer that the system possesses limited workload capacity. If performance does not change as workload is varied, we would infer that the system has unlimited load capacity (at least across the range of workload variations examined; see Fisher, 1982, 1984). Finally, if an increase in workload leads to improvements in performance (as might be expected, e.g., with the perception of highly configurational forms; see O'Toole, Wenger, & Townsend, 2001), the inference might be in favor of super-capacity processing.

All of this begs a critical question: How should we define and measure load capacity? One can, in principle, talk about capacity in terms of RT with, for instance, any of the stopping rules or, indeed, at the level of the individual item or even part of an item, such as a feature. Any of the earlier statistics could be used in any of these cases. And even if a level (such as the individual item) is selected for measurement of capacity, the architecture also can affect its appearance. Suppose that the memory set size  $n$  is increased with a single target among  $n-1$  distractors, and we measure efficiency at the level of the individual item. Further assume that at the individual item level, capacity is unlimited, meaning here that processing speed is unaffected at that level. Then, mean RT in, say, self-terminating processing will dramatically increase if processing is standard serial but will remain flat if processing is standard parallel. Therefore, in the interest of solving this dilemma, we will define, in the paragraphs that follow, a measure of load capacity that is intended to serve as a benchmark index; it turns out to be very useful in establishing ordinal relationships as a function of workload.

As another example, an important type of parallel system, one that can mimic serial processing, assumes that as each item is completed, its processing capacity is re-

allocated to other, remaining items (Townsend, 1972). This obviously affects the overall RT but also creates a positive dependence among the item total completion times. And different types of dependency can affect different measures of capacity (see Townsend & Ashby, 1978). A type of parallel processing where activation (information, neural spikes, etc.) from two or more parallel channels converges in the same output channel before a decision is made evidences super-capacity (Miller, 1982; Townsend & Nozawa, 1995). Note that in each of these examples, not only are we relating independence and capacity, we also are doing so in the context of a specific architecture (i.e., *none* of these questions can be understood or answered out of context).

The basis for a measure of load capacity was originally presented by Townsend and Nozawa (1995) and draws on theoretical and conceptual work by Townsend and Ashby (1978). That article provided a rationale for the use of the *hazard function* on the RT distribution as a measure of capacity. The hazard function is a conditional probability function, defined as

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} \\ = \frac{f(t)}{S(t)}$$

Since the hazard function can be interpreted as the conditional likelihood of completing processing in the next instant, given that processing is not yet complete, there is a natural mapping between this function and the *intensity* with which a system is capable of operating. In fact, there exists a rigorous interpretation of a hazard function as a measure of power in a near-physical sense (Townsend & Ashby, 1983). Although the hazard function provides a nice, fine-grained measure of efficiency, estimating this function from data has been taken to be difficult (e.g., Luce, 1986; Van Zandt, 2000).<sup>6</sup>

The measure of load capacity developed by Townsend and Nozawa (1995) is based on a moderately fine-grained statistic, one that lies between the micro-level grain of the hazard function but is much more fine-grained than the mean RT (see Townsend, 1990b). This is the integrated hazard function,

$$H(t) = \int_0^t h(t') dt'$$

The measure formed using the integrated hazard function directly compares performance for  $n = 1$  versus  $n = 2$  (easily generalized to larger  $n$ ) in an experimental task involving the processing of redundant targets (we will have much more to say about this experimental task later in the review). The goal for this statistic was that it produce a certain constant-across-time index in the case of the standard parallel system (i.e., unlimited capacity, independent, and parallel). This would then stand as a benchmark against which experimental results and other models could be calibrated.

The resulting *capacity coefficient* can be expressed as

$$C(t) = \frac{H_{1,2}(t)}{H_1(t) + H_2(t)}, \quad (1)$$

where  $H_{1,2}(t)$  is the integrated hazard function from the double-target (i.e., redundant or repeated target) condition and  $H_1(t)$  and  $H_2(t)$  are from trials in which only a single target was given.

When processing is described by the standard parallel model,  $C(t) = 1$ . If processing is what we usually think of as limited capacity, then  $C(t) < 1$ . Figure 3 illustrates the cumulative distribution functions, survivor functions, integrated hazard functions, and capacity coefficient for this benchmark model. In contrast, standard serial processing, with self-termination, will predict  $C(t) = 1/2$  in the redundant targets paradigm, as will a parallel model that divides up capacity between the two items or channels. Thus, the fixed capacity model makes the same predictions as the standard serial model on the minimum processing time. If processing is even faster than ordinary parallel processing, then  $C(t) > 1$ .

Next, consider a class of models in which information on each channel is accrued by way of counts; a Poisson model is an example, but not the only one, for this class. Then, if Miller's (e.g., 1982) notion of *coactivation* is interpreted as the pooling of the parallel channel's activations (i.e., counts) into a final common channel, it turns out that such models are always super-capacity—that is,  $C(t) > 1$  for all times  $t > 0$  (Townsend & Nozawa, 1995). Thus, although in one sense, a serial model's items are processed in an unlimited-capacity way, since the speed of individual items does not increase as  $n$  varies, when measured against the parallel benchmark, serial models are very limited capacity. And again, we see how all four questions regarding processing characteristics are critically related in measures of observable performance.

In fact, we can witness a striking example of this intertwining of processing issues in the case of parallel channel dependence and capacity. Assume that we have a perceptual task that involves processing one, two, or three inputs (e.g., one, two, or three facial features). Now assume that there is a *positive dependence* among the rates of process-

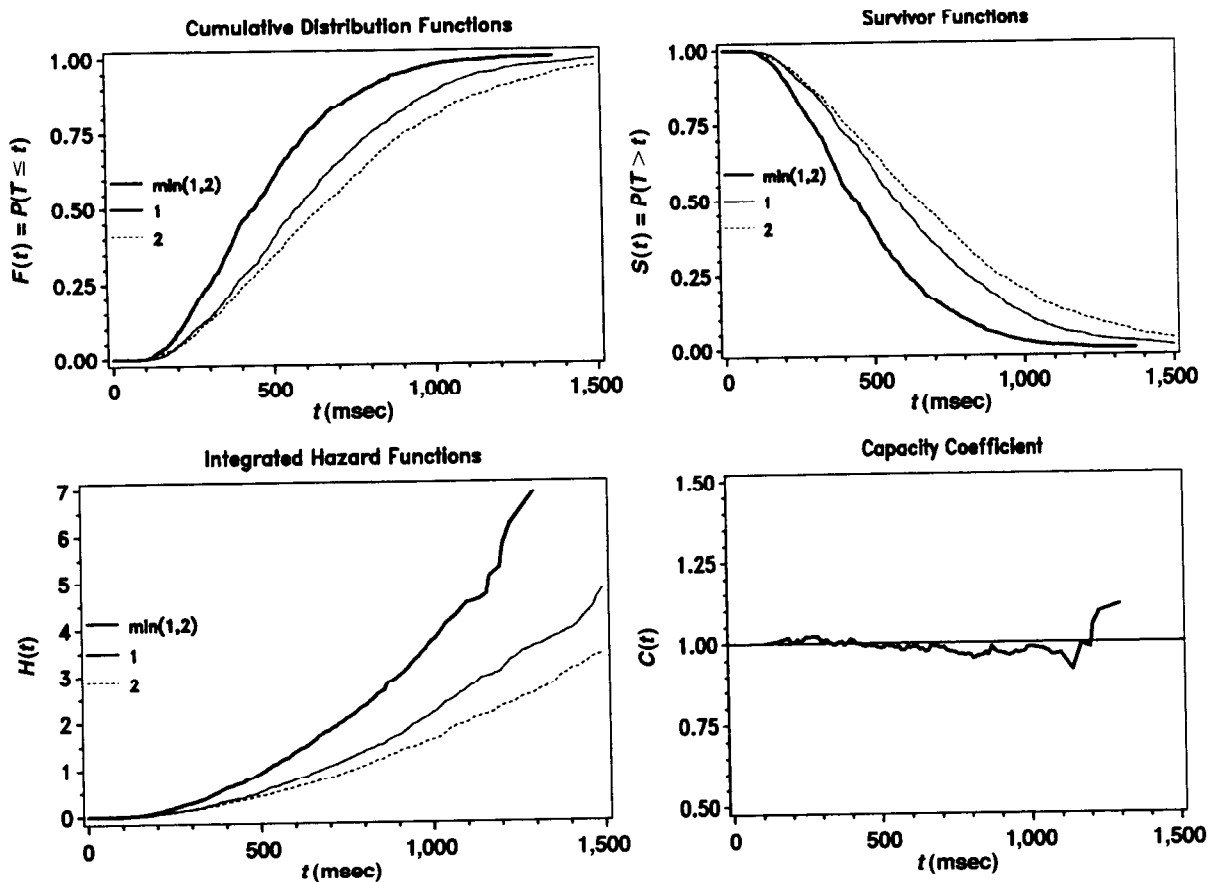


Figure 3. Example cumulative distribution functions, survivor functions, integrated hazard functions, and the resulting capacity coefficient for the two single-target (1, 2) conditions and the double-target condition [ $\min(1, 2)$ ] in a redundant targets experiment, assuming independent parallel processing with a self-terminating stopping rule.

ing the three inputs. So as more information is available for any one of the inputs, the rates of processing the other two increase, and their increase further enhances the rate at which the first input is processed. The overall effect would be to increase the amount of work that the system could accomplish in a given time period: an increase in capacity. The reverse is true for a parallel system whose channels interact in an inhibitory (negatively interactive) fashion.

Recently, Townsend and Wenger (in press), have developed a general theory of capacity measurement that includes the earlier Miller (1982) and Grice (Grice, Canham, & Boroughs, 1984; Grice, Canham, & Gwynne, 1984) inequalities, which apply to first-terminating (minimum-time) processing, as well as new inequalities for exhaustive processing (Colonus & Vorberg, 1994) and an interpretation for an interesting class of noisy dynamic systems. The theory also develops a capacity coefficient that is appropriate for exhaustive (as opposed to first-terminating) processing tasks. A key aspect of this work is a set of regularities governing the relationship of stochastic dependence in parallel systems to the notion of load capacity.

### The Parallel-Serial Dilemma Can Be Solved Using Approaches That Provide a Tight Coupling of Theory and Method

If the question cannot (and should not) be ignored and if it cannot be answered out of context, then what is possible? It turns out that quite a bit is possible, including the distinct possibility of strong inferences. Our stance, however, includes the stipulation that these possibilities exist only when theory (meaning a comprehensive, explicit, computational account of the issues) determines and directs the experimental and statistical methodology. It is of some interest that what we are proposing is at odds with modal practice, in the sense that, in the bulk of the experimental literature in cognitive psychology, experimental tasks or the confines of traditional statistical milieu (e.g., Loftus, 1991; Townsend, 1994) determine and direct the development of theory. However, we must postpone extended consideration of that issue for another time. Instead, in this section, we will provide a limited survey of attempts to answer the question of architecture and will use this survey to motivate the presentation of a theory and methodology that we have developed and used in our own work and that illustrates all of the potentials we have discussed. We have to this point postponed introduction of the modest formalisms regarding serial and parallel processing in which we indulge in this article, although a little of this was inescapable in the earlier outline of the various target issues. We now will expand on the standard serial and parallel models, in the context of the typical stopping rules, self-terminating search with a single target among  $n-1$  distractors, and exhaustive processing, as when all  $n$  items are distractors. We will include reference to a pioneering study by Egeth (1966), which permits first-terminating RTs.

Subsequently, we will outline enough of the quantitative picture of the standard parallel and serial models to

briefly evaluate the classical memory and visual search experimental findings, focusing on the popular Sternberg paradigm and the later program of Treisman, Julesz, and colleagues (e.g., Treisman & Gormican, 1988). Then we will turn our attention to a set of important paradigms, approaches, and facets of the architectural and cognate issues. Naturally, the topics we include are somewhat optional and reflect the subjective judgment of the authors. Townsend (1990a) and Townsend and Ashby (1983) include a broader, if not quite so up to date, set of references. The factorial methodologies (referred to by us in other contexts as *systems factorial technology*) culminate in the double factorial paradigm, which as will be seen, permits the simultaneous assay of architecture, stopping rule, and capacity within the same block of trials.

**A very brief intellectual history: Visual and memory search.** We begin with the task context that has given rise to a large proportion of the work addressing the question of processing architecture: the visual and memory search task. Beyond the important intellectual precedents set in the context of this task, a brief review also highlights numerous connections between theory and data. The serial and parallel classes of models for this task both contain an infinite set of possibilities. But for the present, we will focus on the standard serial model that assumes identical average processing times for each item, independent of the number of items in memory (i.e., the load) or the order of processing the items. It is also often assumed that the individual processing times are themselves stochastically independent of one another.

Let the density of processing times for each item, in this model, be designated  $f$  and that for the independent residual processing time be  $r$ . The time to find a target will naturally be the sum  $s$  of the individual item processing times until the target is processed. Of course, if processing is exhaustive, the required time will be the sum of all  $n$  constituent times. When these times are stochastically independent, the overall probability density function is the convolution of the individual densities. Hence, the density for the observable latency on a probe-absent trial (one requiring exhaustive processing) is the mathematical convolution of the  $n$  processing densities and the residual time density:

$$P(s | EX) = f * f * \dots * f * r,$$

where, as indicated by \*, the convolution is carried out on all  $n$  replicates of  $f$  and the EX stands for *exhaustive*. Note that although all the processing densities are identical, we must convolve the convolution of the processing densities with that for the extra, residual time ( $r$ ).

The expected or average theoretical RT is

$$E(RT) = n \cdot E(T) + E(R).$$

In the instance of self-terminating (ST) processing on a probe-present trial, the number of items finished before the probe is found is itself a random variable. Under usual conditions, the probability that the probe will be

found in the  $i$ th processing position is just  $1/n$ , so the average density for this case is just

$$E[P(t|ST)] = \left[ (1/n) \sum_{j=1}^n f * f * \dots * f \right] * r,$$

where the convolution is carried out  $j$  times. Similarly, the expected RT is just

$$\begin{aligned} E[RT|ST] &= (1/n) \sum_{j=1}^n jE[T] + E[R] \\ &= \left[ \frac{n+1}{2} \right] E[T] + E[R], \end{aligned}$$

which yields the time-honored result that this line has half the slope of that for the exhaustive serial RT (as in Figure 1).

As was noted above, there are models in which the concept of *first finishing time* or, equivalently, *minimum completion time* is important. *Minimum completion time* refers to the first item completed. In standard serial processing, the mean RT would always be the same, according to this kind of model, since the first item processed always consumes the same amount of time. In experimental paradigms in which the number of target items is varied with the number of distractors held constant, a standard serial model with a self-terminating stopping rule will predict a decrease in RT as a function of number of targets, since the processor can stop as soon as a target is located. The same prediction holds if  $n$  is held constant with  $k$  targets and  $n-k$  distractors, when processing is self-terminating.

The pioneering visual search experiment of Egeth (1966) definitely supported self-termination of the minimum time (first termination). He used a *same-different* paradigm in which observers were presented two sets of patterns made up of three distinct dimensions (form, orientation, and color). Observers gave a *same* response if the two patterns were identical and a *different* response if they differed on one or more dimensions. Varying the number of dimensions that are in the target state serves the same purpose as varying the number of targets, since a *different* response can occur as soon as the first mismatching value on a dimension is discovered. The data for trials involving at least one dimension that was different revealed RTs that decreased as a function of the number of differing dimensions. This result supports self-termination (Townsend & Van Zandt, 1990) but gives no information regarding the question of architecture. The reason, as was noted above, is that both serial and parallel models predict a decrease when a varying number of different dimensions is embedded in a fixed number of same dimensions (or analogously, the frequency of targets is varied within a fixed total number of items). However, when the experimenter collects data from sets of trials in which only targets are present and varies that number, decreasing RTs are strongly supportive of parallel processing (depending, of course, on the requisite assumptions regarding independence and capacity).

The concept of parallel processing, like that of serial processing, is usually confined to a very special kind of parallel processing. Processing is often said to be parallel if average RT is invariant of load—that is, the mean RT curve is a flat line as a function of load. When exhaustive processing is required, as in the probe-absent trials of memory search, this stipulation actually implies a highly unusual kind of parallel processing. This error (at least in inference) in the experimental literature is committed so often that we will pause to illustrate it, after introducing the standard parallel processing model.

Consider a set of parallel models with independent processing times, with the additional provision of unlimited capacity, in the sense that the probability distribution on completion time for each item does not change as  $n$  is increased. This class of models—the standard parallel model introduced earlier—serves as a prototype of parallel processing in a way analogous to that of the standard serial model just above. Every standard parallel model, irrespective of the underlying family of probability distributions on processing times, has to predict increasing concave (down) mean RT curves, not flat curves in exhaustive processing (Townsend & Ashby, 1983, p. 92; see also Sternberg, 1966). Let  $g(t)$  be the fixed density function on processing time for each of the  $n$  items.<sup>7</sup> Then, for exhaustive processing,

$$E[RT|EX] = E\{\max[T_1, T_2, \dots, T_n] * r\},$$

where  $r$  is as above and  $T_i$  is the processing time of the item in position  $i$ . We drop reference to  $r$ , which plays no role in the function shape, in order to lessen cumbersome notation (keeping in mind the fact that we are ignoring the component of residual time).

Then,

$$E[RT|EX] = \int_{t=0}^{\infty} \{1 - [G(t')]^n\} dt',$$

where  $G(t)$  is the cumulative distribution function associated with  $g$ . It is straightforward to show from this formula that it is indeed an increasing concave function of  $n$  (Townsend & Ashby, 1983, chap. 4). Note that  $g(t)$  is assumed to be invariant across values of  $n$ . Hence, many investigators use a very restrictive and unrealistic criterion for parallelism, since they demand flat, rather than increasing, mean functions of  $n$ , even in exhaustive processing conditions. It can also be observed that if mean exhaustive RTs are flat functions of  $n$ , mean self-terminating times would actually be decreasing within the same model—a strong prediction that apparently has never been checked in studies using this logic. Thus, in the rare cases in which flat exhaustive mean RT functions are actually found with exhaustive processing, the implications are very strong.

In order to make the situation more concrete, consider a parallel processing model in which all of the component durations possess an exponential distribution. Further assume exhaustive processing and equal rates on each item, which remain constant across changes in load.

Call this rate  $v$ . Thus, this model is the exponential case of the standard parallel model. The average time to exhaustively process all  $n$  items is then

$$E(\text{RT} | \text{EX}) = \frac{1}{v} \sum_{i=1}^n \frac{1}{i},$$

which is, as was predicted, increasing but concave down as  $n$  increases. Now if the exhaustive processing time were to be constant, as is often assumed about parallel processing, then we clearly must make the processing rate on each item,  $v$ , depend on load,  $n$ —that is, the rate must be written as  $v(n)$ . If the mean exhaustive time is flat, then

$$E(\text{RT} | \text{EX}) = \frac{1}{v(n)} \sum_{i=1}^n \frac{1}{ci} = 1,$$

where  $c$  is a constant indicating the constant overall exhaustive processing duration. This in turn, implies that

$$v(n) = \frac{1}{c} \sum_{i=1}^n \frac{1}{i};$$

that is, the individual processing rate has to increase like something close to the logarithm of  $n$ . Since unlimited capacity means that  $v(n)$  is constant, the latter result implies *super-capacity* and considerable speeding up of processing items as the load grows larger.

In contrast, in the case of self-terminating processing when a single position contains the probe, standard parallel models predict a flat mean RT function. This can be readily intuited, since only the channel processing the probe matters in self-terminating processing and it is independent of all the rest (and is unlimited capacity as well). The processing time density in this case is just  $P(\text{RT} = t | \text{ST}) = g(t)$ .

Finally, in this standard parallel model, mean first-terminating times should decrease with  $n$  (e.g., Townsend, 1974). Here, the first-terminating density is

$$P(\text{RT} = t | \text{FT}) = ng(t)[1 - G(t)]^{n-1}.$$

In the above exponential version of the standard parallel model, it is interesting to see that  $E(\text{RT} | \text{ST}) = 1/v$  and  $E(\text{RT} | \text{FT}) = 1/nv$ , fulfilling the expectations for their behavior. Interestingly, a subset of Egeth's (1966) data, where all of the dimensions differed and the number of targets varied from one to three, would have allowed a test of serial versus parallel processing, since a standard parallel model would predict a decline in RT, whereas a standard serial model would predict a flat RT function. Those data were not plotted, but in the work that followed, a plethora of such data, much of it from Egeth's laboratory, supported parallel race (first-terminating) processing.

**The earliest studies.** Now we can sketch in the conclusions reached by Sternberg (1966) and Atkinson et al. (1969). In both studies, mean RTs increased in a linear fashion, in line with standard serial processing. The probe-absent trials evidenced a slope that was about the same as that for the probe-present trials, compatible with the standard serial model with exhaustive processing,

even on probe-present trials. The fact that the mean RT functions of  $n$  were linear, rather than bending over, falsifies the standard parallel model and is in favor of standard serial processing. In the closest match to this paradigm, but in the visual domain, we find Atkinson et al.'s data, which basically replicate the essential findings of Sternberg.

Although both these sets of data are compatible with standard serial processing, Atkinson et al. (1969) and Townsend (1969, 1971a) offered a parallel model that could mimic these findings. In addition, Townsend (1969, 1971a) developed a serial model that mathematically mimicked the standard parallel model. More general results followed that indicated wide-ranging potential for mimicking (e.g., Townsend, 1972, 1976a). Although the architecture question remained of great interest in many domains, the frequency of experimental studies attempting to test the concepts dropped off in the 1970s, perhaps partly due to concern with mimicking issues. In fact, as we suggested earlier, some investigators apparently concluded—*incorrectly!*—that the question was a nonissue and that the two types of processing were not distinct, even in reality. This was an unfortunate overreaction. Theorists exploring relations between parallel and serial processing and striving to use their mathematical results to build experimental methods to broaden the venue of testability were not so dire. The “wiring” of parallel and serial systems that one could actually build with physical components are distinct entities—the trick is to find ways to test as many parallel models against as many serial models as possible (a strategy consistent with the commitment to strong inference and multiple hypotheses; see, e.g., Platt, 1964; Popper, 1935).

**Treisman and Julesz.** Treisman and Julesz and their colleagues reinvigorated consideration of the question of architecture through a wide range of experiments and models. Their theories were structurally almost identical to each other, and we shall follow Cave and Wolfe (1990) in referring to both under the *feature integration theory* label. Their models were constructed to account for intriguing data they had gathered on the processing of visual features under differing conditions and instructions. The central finding, replicated many times, was that RTs in search for a single feature value (e.g., a red square) among a set of differing values (e.g., a set of green squares) increased with the number of distractors at a very slow rate, sometimes with almost zero slope. In contrast, search RTs for a conjunction of features (e.g., a red square among a set of red circles and green squares) increased quite dramatically with  $n$ .

The inference was that the single-feature search took place in an early visual-processing stage at which processing was parallel. The more complex conjunction search was forced to be carried out in a subsequent, serial fashion in which attention could be placed on the several items one at a time. Although this basic idea had been suggested much earlier (dating at least to Neisser, 1964), the Treisman and Julesz groups greatly expanded the experimental body of data that were compatible with

it. Despite this wealth of consistent *circumstantial* evidence, this approach did not avoid the problem of model mimicking. Observe that standard parallel models do indeed predict a flat mean RT function on target-present trials. And of course, standard serial models make the prediction of linearly increasing mean RT functions even if processing is self-terminating. Nevertheless, the same limited-capacity parallel models that mimicked Sternberg's (1966) and Atkinson et al.'s (1969) data perfectly mimic the serial model espoused by Treisman and Julesz, as they in fact recognized (e.g., Treisman & Gormican, 1988).

An arbitrary serial model demands only that processing be sequential with discrete flow. In the general case, we need to stipulate both the locations of items (e.g., serial position in a visual display or memory storage repository) and the identities of items. Thus, the stimulus letter V might be placed in Array Location 1, which could be the leftmost location, whereas the stimulus letter S might be placed just to the right, which we could call Location 2, and so on. Then, any permutation of, say, the various locations could be used as an order of processing according to some probability distribution that is either prefigured by the system or possibly imposed from without. A complete description, therefore, requires notation for identity, location, and processing order.

Of course, a permutation in terms of location immediately determines the permutation in terms of identity, given information about the placement by the subject or the experimenter of the items in the array. However, array location and item identity possess, in our opinion, quite distinct implications for model identifiability. Little experimental work has been done using identity properties, but obvious exceptions include the coarse but important notions of *pop-out* stimuli and the conjunctive and disjunctive dimensional constituents of search objects. A manipulation of the similarity of distractors to each other and to the target can also be cast in this mode (e.g., Duncan, 1984).

Consideration of such issues (extended treatment of which goes beyond the purposes of the present article) reveals that the set of aspects on which processing depends is not equivalent in parallel and serial processing. The serial models can make processing times depend on the entire order with regard to location (not items!) all the way through the evolving history of processing, whereas the parallel models can only make their processing times depend on the previously finished order of locations. On the other hand, serial models cannot make current processing times depend on the identity of items not yet completed, but parallel models can. Both kinds of models can exhibit dependence of current times on the past items processed and their order.

Nevertheless, serial models can still be more general for a fixed experimental set-up in that at stage *i*, the processing time can be made dependent jointly on the previous items processed, their order, the previous order with respect to location, and the future order of process-

ing the various unfinished locations. In contrast, although parallel models can make the finishing time at stage *i* depend on the previous items processed, their order and the previous order with respect to location, they cannot make the processing time depend on the future order with respect to location.<sup>8</sup>

**Whole-report tasks and mean RT.** The whole- (or *full-*) report paradigm goes back to the 19th century and centers on presentation of a number of visual items, all of which the observer must try to report. It came into renewed prominence with Sperling's (1960) experiments on whole and partial report (where a signal indicates the part of the display to report). Basically, when the number of displayed items was held constant but the postmasked display duration was varied, Sperling's early results (e.g., Sperling, 1960, 1963) seemed to suggest serial processing; that is, when the average number of correct items was plotted against display duration, the result was linear, with a slope of about 1 item/10 msec. However, later data with a greater range of display durations, yielded curves that appeared to be more compatible with parallel processing (Sperling, 1967). Many (but not all) serial models predict S-shaped curves, with perhaps the curve in the lower part of the S becoming more pronounced in the later-processed positions. And the successive curves may appear to shift along the time axis. Some parallel models (for instance, those based on exponential distributions) can predict non-S-shaped proportion correct curves; however, they need not. Independent parallel models tend to predict curves that fan out from a starting point, rather than exhibiting the shift property. To the eyeball, the curves seem to be more in accord with independent parallel processing.<sup>9</sup> In general, in these circumstances, the serial-parallel mimicking dilemma is a significant obstacle. However, certain helpful statistics that could help adjudicate the issue have been overlooked.

Interestingly, almost all research on whole report has centered on the marginal accuracies of the various positions, with no attention to correlations. The first author has long called for research on the across-position dependencies (Townsend, 1981), since such statistics could be highly useful in testing models that might be untestable or difficult to test on the marginal accuracies. Most models of whole report assume independence (e.g., Busey & Loftus, 1994), but some predict negative dependencies (e.g., Bundesen & Shibuya, 1999; Townsend, 1981). Although Townsend's (1981) study showed the positional correlations to be nil, a recent whole-report investigation, with somewhat different stimuli and experimental conditions, discovered significant positive across-position correlations (Busey & Townsend, 2001), falsifying all current models for those data. Although such dependencies must be predicted by any serious candidate for a whole-report process theory, they do not, strictly speaking, test between serial and parallel models (see Busey & Townsend, 2001, Appendix and related discussion). Nevertheless, in general, many standard serial models tend to predict positive correlations of total completion times (not actual

processing times for serial models), whereas independent parallel models (including the standard parallel models) obviously predict independence. The best-fitting models in Busey and Townsend's study were, in fact, a standard serial model and a variable attention parallel model. The latter assumed independent parallel processing on each trial, but since attentional capacity fluctuated across trials, overall, a positive dependency was induced.

Are there any whole-report designs that focus specifically on the parallel-serial question? The answer is yes and involves a paradigm that continues to be exploited in various contexts. Let us illustrate the idea with  $n = 2$  items, although the concepts clearly generalize. Let us call it the *split-presentation* paradigm. The basic idea is that the experimenter compares performance in a condition in which all the items are presented simultaneously with one in which they are presented one at a time, usually with poststimulus masks. There are two basic variations on the basic design. In the first, the sequential presentation times for each item are determined by estimates (garnered from regular whole-report trials) of how long it should take, if processing is serial, to complete the first and then the second items. For instance, the times for each item might be equal if the average-number-correct curve is approximately linear as stimulus duration increases. Then the simultaneous condition will have a duration equal to the sum of those in the sequential condition. If processing is parallel, the intuition is that performance should be better in the simultaneous condition, since each item will receive the full duration for its processing, whereas in the sequential condition each item receives only about one half the duration to finish processing. However, in strict serial processing, it should make no difference, since attention has to abandon the first item after about one half the total time. This is the model used by Townsend (1981). His results were strongly supportive of parallel processing, and an explicit parallel model provided good fits to the regular whole report, both the marginal and the dependence statistics, and to the split-presentation condition data. The fact that a parallel class of models encompassed both the marginal and the dependency data may help support the parallel alternative model in Busey and Townsend's (2001) whole-report study.

The second variation reverses the logic, in that it has each of the individual sequential presentation times fully equal to the total presentation time in the simultaneous condition. Here, the serial model should predict superiority in the sequential condition, but the parallel model should predict that there will be no difference between the conditions, since each item will receive exactly the same processing time in both conditions. Interestingly, as far as we know, this version has never been employed in whole-report studies. It has been employed a number of times in search or detection paradigms, where most results appear to support parallelism (e.g., Eriksen & Spencer, 1969; Shiffrin & Gardner, 1972). Although the split-presentation paradigms can be very useful for distinguishing some competing models, for models making

the correct predictions, there are models from the alternative class—that is, parallel or serial—that can mimic the apparently winning model. Thus, serial models that mimic independent parallel models are indistinguishable in these data from true parallel processing, as are parallel models that mimic standard serial processing from true serial processing (e.g., Townsend & Ashby, 1983, especially chap. 11).

**Varying target repetitions.** Manipulating the number of targets in a display in a search task can help identify architecture and the stopping rule under specified conditions. Here, we will discuss a fairly specific architecture assay that can be applied when  $n$  is held constant and the number of redundant targets is varied. This method was first developed and implemented by Taylor, Townsend, and Sudevan (1978) and is described in detail in Townsend and Ashby (1983, chap. 6). The classes of models tested are not as general as is possible in the factorial technology discussed later; however, the method of testing is intriguing in its own right.

The test works for self-terminating models, but exhaustive predictions for either serial or parallel models include the strong prediction of invariance as a function of number of redundant targets. The parallel models assume independence and an absence of position effects, on the basis of Weibull distributions (1951, which includes the exponential). The independence assumption is reasonable, since it is a characteristic of what we call standard parallel processing, but the absence of position effects is often wrong. We think this violation of the assumption would not likely damage the fundamental aspects of predictions, but this has not been proven.

As has been described in Townsend and Ashby (1983, pp. 157–163), the predictions for mean RT as a function of  $n$  are probably too similar between the model classes to be useful in deciding the question of architecture. However, if one plots mean RT against the expected number of item completions before stopping, the predictions for the serial and the parallel models turn out to be dramat-

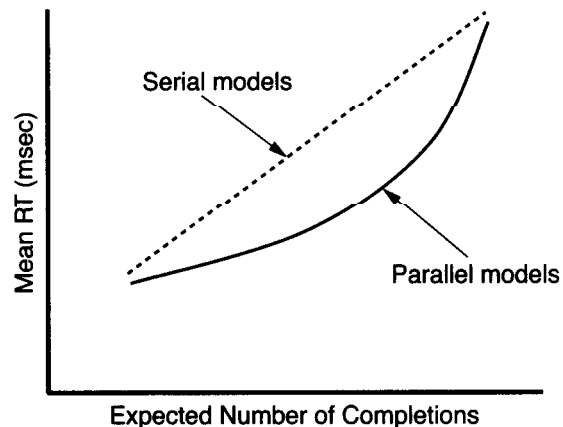


Figure 4. Predictions for mean response time (RT) as a function of architecture and total expected completions before stopping.

ically distinct, as is illustrated in Figure 4. Interestingly, the serial models predict linearity of mean RT as a function of expected number of completions. The parallel models, on the other hand, always predict curvature of the mean RT function, with the exact form of the curvature depending on the specifics of the parallel distributions; they cannot predict linearity. The results, in an admittedly first implementation, were strikingly supportive of serial processing, with firm adherence to linearity (see Townsend & Ashby, 1983, Figure 6.5, p. 162).

**The parallel-serial testing paradigm.** An example of a complete linkage of theory and method, in the context of visual and memory search, is the parallel-serial testing (PST) paradigm, described in detail in Townsend and Ashby (1983, chap. 13). This paradigm relies on three basic search conditions (see Table 13.1, p. 416, in Townsend & Ashby, 1983) involving two targets, A and B. The first condition requires an observer to detect whether A is on the right with B on the left or vice versa. Since detection of either A on the right or B on the left determines the response, a first-terminating (or minimum-time) rule is allowed. In the second condition, the decision rule requires that A be in both positions for a positive response—that is, processing must be exhaustive for an AA trial. On the other hand, a first-terminating rule can work on BB trials. In an AB trial, processing can cease just as soon as the B is done, but not before. This is ordinary self-termination; the same goes in reverse on a BA trial. The entire situation is mirrored by reversing the roles of A and B in the third condition.

This paradigm was inspired partly by earlier methodology described by Snodgrass (1972) and partly by theoretical work emphasizing certain fundamental distinctions between parallel and serial processing. The PST paradigm simplifies aspects of Snodgrass's work in such a way that very specific assumptions about the processing path need not be made. In addition, it allows for mathematical proof, using Townsend's (1976a) results, of the experimental discriminability of parallel and serial processing (see also the discussion in Luce, 1986, pp. 464–472). The fundamental postulate of the methodology is that positive, or *same*, matches can take different amounts of time (usually shorter) than negative, or *different*, matches. This assumption has been verified many times (for a relatively early landmark study, see Bamber, 1969; Bricolo et al., 2002, also is pertinent). It is also necessary to postulate that processing can self-terminate when that is allowed for a definite correct response, but the data can also be assayed to test this assumption (and see Townsend & Colonius, 1997; Townsend & Van Zandt, 1990; Van Zandt & Townsend, 1993).

Like most model-testing situations, it is also necessary to assume that the same processing distributions, whether serial or parallel, are employed across the three conditions. This assumption might raise problems (an almost universal challenge in most attempts at providing converging evidence on a topic) if the three conditions are separately blocked but is more acceptable if the three

conditions, cued on each trial, are run within the same blocks. Although it is not possible to go into detail regarding the predictions of the serial and parallel models for the relations among the data from the three conditions of this paradigm, it is possible to state that the two model classes make distinct, testable predictions (again, see chap. 13 of Townsend & Ashby, 1983, for a thorough treatment of the details).

Snodgrass and Townsend (1980) carried out a PST design that suggested a strong form of parallel processing. However, other applications of PST, as well as related and more complex designs, sometimes have shown support for serial processing (Snodgrass, 1972; Snodgrass & Townsend, 1980). The fact that different stimuli and conditions can lead to different conclusions is in accord with our most recent work suggesting strong differences between and within individuals (Townsend & Fific, in press).

**RT variance.** The variance of the RT distribution is probably more useful in testing hypotheses regarding the stopping rule, rather than hypotheses regarding architecture. Nonetheless, it certainly can be useful in probing certain kinds of parallel and serial models. As has often been the case, specific properties associated with standard serial processing (in this case, the variances or standard deviations under self-termination vs. exhaustive processing) are (1) used as if they were indicative of all serial models and (2) assumed to always be different in parallel processing from those in serial processing. Neither of these assumptions is, of course, true (Townsend, 2001).

Standard serial models typically assume that the processing time distributions are the same on all positions and items and irrespective of the order of processing. In addition, these models usually assume that each successive item-processing time is independent of all the rest. It then follows that if  $k$  things are processed, the variance is simply  $k\sigma^2$  (i.e., standard deviation is  $\sqrt{k} \cdot \sigma$ ), where  $\sigma^2$  is the variance for the distribution on processing time for any one item. For exhaustive processing, obviously, the variance is just  $\text{VAR}_{\text{EX}} = n\sigma^2$ .

Interestingly, Schneider and Shiffrin (1977) showed that the variance for correct positive trials in the case of self-termination with a single target is

$$\text{VAR}_{\text{ST}} = \frac{n+1}{2}\sigma^2 + \frac{(n-1)(n+1)}{12}\mu^2,$$

where  $\sigma^2$  is as above and  $\mu$  is the mean of the single-item processing times. They then pointed out that it is possible for the self-terminating variance to start out, for small  $n$ , less than the exhaustive variance but ultimately to cross over and become larger, as is illustrated in Figure 5. Townsend and Ashby (1983, pp. 192–201) explored the conditions under which the crossover will occur. Of course, any parallel model that is equivalent to the standard serial model and also obeys self-termination on target-present trials will make the identical prediction, as Schneider and Shiffrin pointed out. As with mean RT slopes (e.g., Townsend & Van Zandt, 1990), it is unlikely



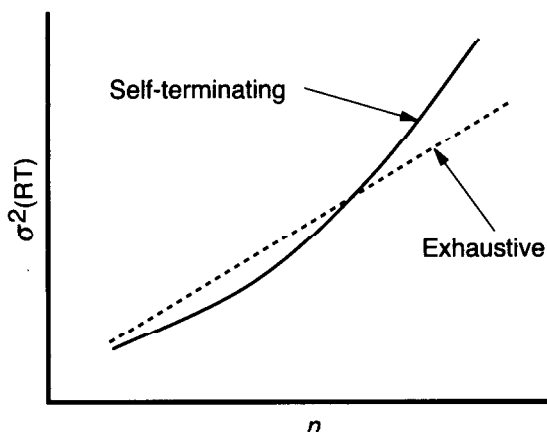


Figure 5. Predictions for the variance of response time (RT),  $\sigma^2(\text{RT})$ , as a function of processing load and stopping rule.

(although not yet proven) that exhaustive models can make predictions like this, even assuming different processing distributions for targets versus nontargets.

Recently, Donnelly, Found, and Müller (1999) employed this result to try to discriminate serial from parallel processing. A pivotal concept was that even with self-termination on target-present trials, limited-capacity parallel models would predict equal slopes of the target-present and target-absent variances, when plotted against  $n$ . Although praising the use of RT distributions to test hypotheses about process models, Townsend (2001) showed that many limited-capacity parallel models do not satisfy Donnelly et al.'s prediction (just as not all serial models need predict Schneider & Shiffrin's, 1977, variance prediction). Needless to say, there exist serial models that can also make the equal-sloped variance prediction.

Furthermore, unlike standard serial models, whose variance characteristics are very well known, the behavior of the variance of standard parallel models, even with independence and unlimited capacity and without position effects, is far from understood. We do know that for exhaustive processing (or for minimum processing times for that matter), the variance can either increase or decrease as  $n$  increases. There are some ideas about what might govern the actual behavior within a specific model, and it has been shown that for going from  $n = 1$  to  $n = 2$ , the variance will always decrease (Townsend, 1990b)! Unfortunately, Townsend's (1990b) proof does not appear to easily generalize to arbitrary values of  $n$ .

More recently, Townsend and Colonius (2003) have taken a new tack, using a generalization of the range that they call *quantile spread*, to the understanding of variability as a function of  $n$ , for parallel models with independence and unlimited capacity (i.e., standard parallel models). With this approach, they have studied the way in which range measures act for parallel models with self-terminating or exhaustive processing as  $n$  varies. It turns out that the precise nature of the change in variability as a function of  $n$  (i.e., increases or decreases) depends on the shape of the single-item processing time

distributions. Naturally, this kind of prediction can be captured by certain serial models too. However, the finding of decreasing variances in exhaustive processing would definitely falsify not only all standard serial models, but also a substantial class of limited-capacity parallel models.

Finally, with regard to issues about RT variances, Egeth and Dagenbach (1991) have suggested that when factorial methods are used to test serial versus parallel processing, the less variance in the individual processing time distributions, the better—that is, the larger—the mean interaction contrast will be. Townsend and Diederich (1996) demonstrated that this was indeed true under some conditions, but not in others. In some models, the larger the item variance, the bigger the effective mean interaction contrast.

**Monotonic hazard functions.** Most serial models, including the standard class, predict that as  $n$  grows, the hazard function will be ordered in terms of  $n$  (e.g., Ashby et al., 1993; Townsend & Ashby, 1983). This, in turn, implies that the distribution function, means, and medians will also be so ordered (Townsend, 1990b). Limited-capacity models can make this prediction, although unlimited-capacity independent ones do not.

More interesting, many serial models (although not all) predict that the hazard function will be an increasing function of  $t$  (Ashby et al., 1993) for any fixed value of  $n$ . Parallel models do not generally make this prediction (Ashby et al., 1993). In a number of instances—for instance, the extensive study by Ashby et al. (1993)—the first prediction has been shown to hold, but not the second. Assuming that the choice is between serial and parallel processing, such results are supportive of the parallel alternative.

**Factorial methods: Mean RT.** One of the most promising and general approaches to the identification of processing architecture is that based on the notion of *selective influence* of experimental factors on perceptual and cognitive processes, a notion first employed in tests of strict seriality by Sternberg (1969a, 1969b, 1998) in his additive factors method. All factorial methodologies, like the original Sternberg strategy, depend on the notion that distinct experimental factors affect distinct processing components (i.e., subsystems). This is the assumption of selective influence (Dzhafarov, 1997). For example, in the case of two experimental factors  $X$  and  $Y$ , assume that there exist manipulations that can separately influence the processing associated with factor  $X$  ( $X + \Delta X$ ) and factor  $Y$  ( $Y + \Delta Y$ ). Then  $\text{RT}(X + \Delta X, Y)$  refers to the case in which the  $X$  factor has prolonged RT but  $Y$  is at base level, and so on. The complete combination of initial and prolonged cases for each of the factors is a standard  $2 \times 2$  factorial design, and the fundamental statistic for the original method and most extensions is the interaction contrast for the mean, or ICM (see Figure 6):

$$\text{ICM} = \overline{\text{RT}}(X + \Delta X, Y + \Delta Y) - \overline{\text{RT}}(X + \Delta X, Y) - \overline{\text{RT}}(X, Y + \Delta Y) + \overline{\text{RT}}(X, Y). \quad (2)$$

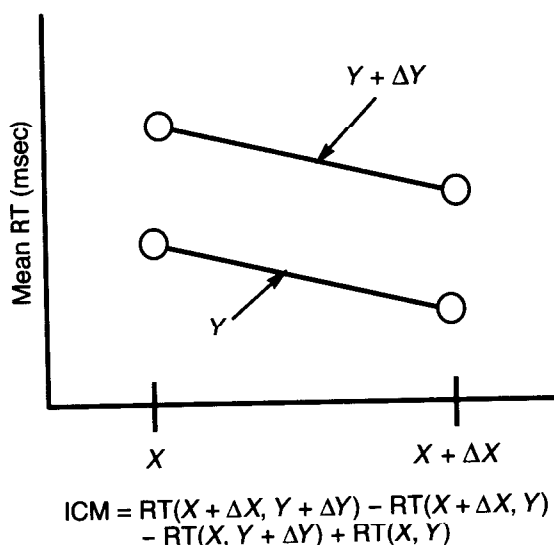


Figure 6. Data from a typical  $2 \times 2$  factorial design, represented in terms of the components of the interaction contrast for the mean.

Sternberg (1969b) indicated that serial models acting under selective influence with independent processing times would exhibit  $ICM = 0$ —that is, factorial additivity. It was intimated and was intuitively obvious (although not proven analytically at the time) that it would be possible for serial models with correlated processing times to satisfy selective influence and mean RT factorial additivity. A prime example is found with normally distributed RTs: Because the variances and correlation are parameters that are mathematically independent of the means, the latter need play no role in ICM, which is a function of the means only.

In the original method, if  $ICM \neq 0$ , the logical conclusion was to be either that selective influence had failed or that serial processing was not in force. Literally hundreds of studies have been carried out using the additive factors method, as originally conceived. Notably, Sanders (e.g., 1990; Sanders, Wijnen, & van Arkel, 1982) has employed the method for decades, accumulating a broad range of evidence regarding serial architecture and other aspects of processing. Recently, Sternberg has contributed a deep interpretation of systems open to the additive factors method, in terms of neurophysiological modules (e.g., Sternberg, 1998, 2001).

To a number of theorists, the appearance of the original method in 1969 raised a vital question: What kinds of predictions would nonserial architectures make? Schweickert (1978, 1983), in his latent mental network theory, contributed the first major extension of the additive factors method, involving more complex architectures under the assumption of selective influence. This theory was very general, including serial and parallel systems as special cases. A key concept in Schweickert's theory

was that of *slack*. Slack between two processes A and B indicates the amount of time the A process can be prolonged by experimental means without delaying the time that B begins. Total slack measures the amount of time a certain subprocess may be prolonged without increasing the overall response time (Schweickert, 1978). Coupled slack for two processes, A and B, measures the difference in the total slack of a process and the regular slack.

In addition to many theorems regarding deterministic forward-flow networks, based on the slack concept, the major formulae regarding effects of factor manipulation were shown to be valid for stochastic networks if the effects were large and the two factor-affected subprocesses were on a path together. Schweickert (1982) subsequently laid out a set of stochastic inequalities under broader assumptions. Taking a different approach, Townsend and Ashby (1983) found that the mean interaction contrast distinguished parallel and serial stochastic models when selective influence was assumed. Interestingly, the sign of the mean interaction contrast depends not only on the architecture, but also on the stopping rule. Thus, parallel exhaustive processing exhibits a negative contrast, whereas a parallel race model will evidence a positive contrast (see Townsend & Nozawa, 1995).

Schweickert and Townsend (1989) produced general theorems for Schweickert's latent networks, within a stochastic setting, and assuming exhaustive processing. The fact that the sign of the mean interaction contrast (negative, zero, or positive) was employed to distinguish the varieties of architectures and their functioning, led to the methodology (in Schweickert & Townsend, 1989) being referred to as the *trichotomy method* (see also Townsend & Schweickert, 1989). Although the early theorems in all this work were proven in the context of exhaustive processing, analogous results can be found in the case of self-terminating and first-terminating processing times (e.g., Schweickert & Giorgini, 1999; Schweickert & Wang, 1993; Townsend & Nozawa, 1988, 1995).

Figure 7 illustrates the mean RT predictions for parallel and serial processing for single-target self-terminating times, first-terminating times, and exhaustive processing, assuming independence in the rates of processing each of the items. It can be seen that all serial models predict an absence of interaction—that is, mean RT additivity. However, in the case of parallel processing, each stopping rule is associated with a distinct result: (1) First-terminating times result in over-additivity, (2) single-target self-termination is associated with additivity, and (3) exhaustive parallel processing predicts under-additivity. Hence, additivity in the context of single-target conditions is unhelpful. Otherwise, a positive or negative interaction or no interaction all reveal the underlying architecture.

**Factorial methodology: RT distributions.** As has been suggested by Townsend (1990a, 1990b), certain aspects of probability distributions are more powerful than others. That is, knowledge of some aspects always implies knowledge of others, but not vice versa. In particular, the

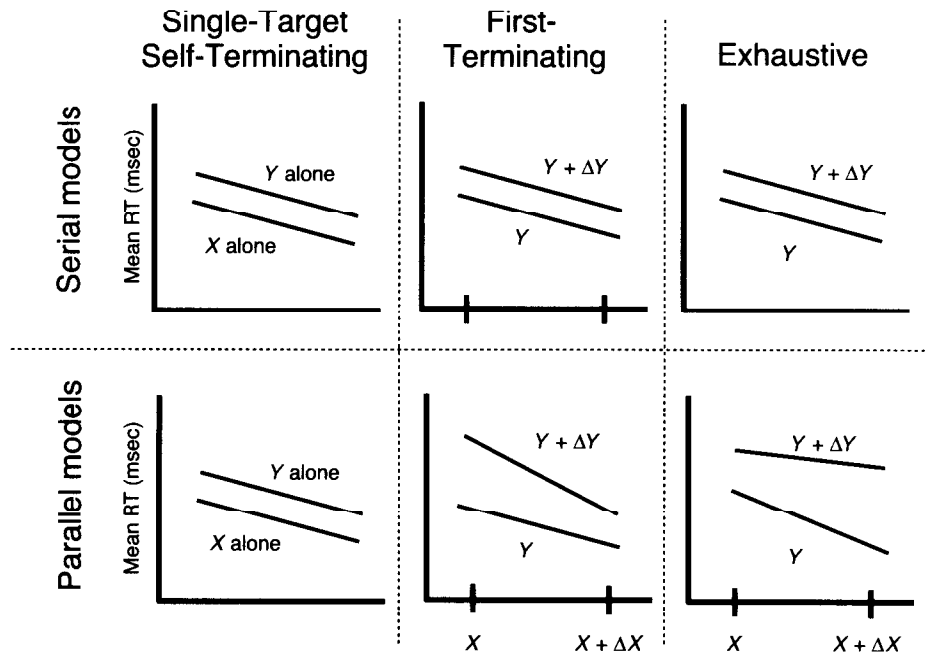


Figure 7. Predictions for mean response times (RTs) and mean interaction contrasts as a function of architecture, stopping rule, and number of targets (one target, in the panels of the leftmost column; two targets, in the panels of the second and third columns).

entire cumulative probability distribution function on RT is more powerful than the means alone. Figure 8 illustrates a critical subset of all of the relations described in Townsend (1990b).

An apt example to motivate this discussion can be found in the consideration of a special class of parallel models known as *coactive processing models* (e.g., Miller, 1982; Townsend & Nozawa, 1995). In ordinary parallel processing, each channel does its job, and these individual channel decisions (e.g., present vs. absent) are followed by a logical decision gate. Thus, in an exhaustive (conjunctive) condition, the logical decision for a designated response would be an AND gate, but in a disjunctive condition, the logical decision would require an OR gate. By contrast, in coactive parallel processing, the multiple channels all feed into a common channel that itself streams into a single decision threshold function (i.e., stop and respond if and only if the criterion is reached). This kind of parallel machinery is not well equipped to make diverse logical decisions but may be in operation when, for instance, an OR decision is needed (Colonius & Townsend, 1997; Miller, 1982; Townsend & Nozawa, 1995).

A relevant question is whether this kind of parallel processing can be distinguished, using factorial methods, from ordinary parallel processing. It turns out that at the level of mean RTs, and within an OR design, it *cannot* be distinguished from ordinary parallel processing with an OR stopping rule (i.e., with a race between the two channels determining when the process is completed). Specifically, the ICM is positive just as in a parallel

horse race. However, if the factorial interaction concept is extended to the entire RT distribution (as in Townsend & Nozawa, 1995), it turns out that it is possible to distinguish a coactive model from a standard parallel model with an OR gate. In principle, either the cumulative probability distribution [ $P(T \leq t) = F(t)$ ] or the survivor function [ $S(t) = 1 - F(t)$ ] can be used (e.g., Schweickert, Giorgini, & Dzharov, 2000). Since the original derivations were

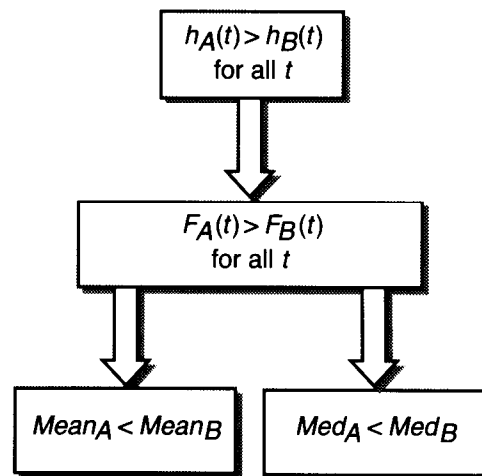


Figure 8. Dominance relations among three levels of statistical description for a set of response time data. Arrows indicate the allowable direction of implication. Note:  $h_i(t)$  = hazard function,  $F_i(t)$  = cumulative distribution function. Med, = median.

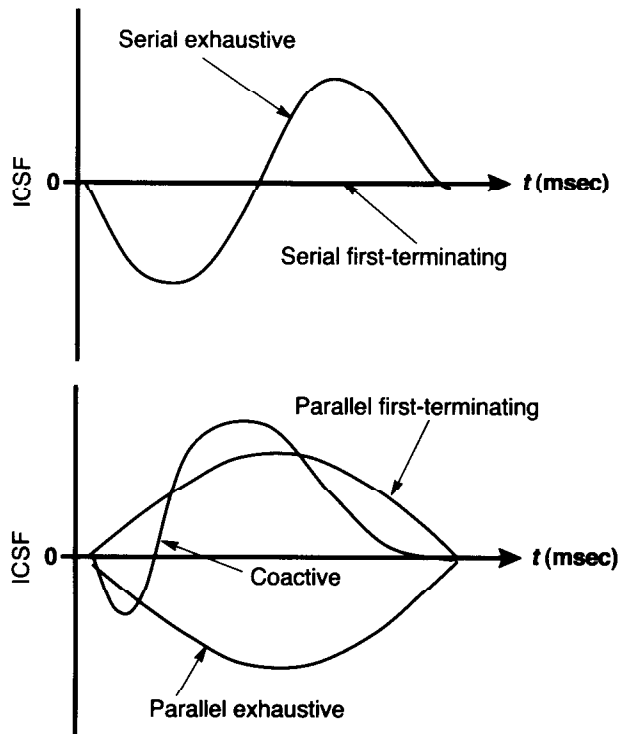


Figure 9. Predictions for the interaction contrasts on the survivor function (ICSFs) for each combination of processing architecture and stopping rule.

in terms of the survivor function (Townsend & Nozawa, 1995), we employ that one here, with the interaction contrast for the survivor function (ICSF) being

$$\text{ICSF} = S(t | X + \Delta X, Y + \Delta Y) - S(t | X + \Delta X, Y) - S(t | X, Y + \Delta Y) + S(t | X, Y). \quad (3)$$

Figure 9 displays the predictions for the various models. Observe that ordinary parallel-processing ICSFs reveal total positivity in the case of OR conditions but total negativity in the case of AND conditions. Furthermore, OR parallel and coactive parallel processing now are distinguished by their respective ICSFs: The contrast for OR parallel processing is consistently positive, whereas the contrast for the coactive model possesses a small negative "blip" at the earliest times, before going positive. Since the ICM must be positive in coactivation, it can be shown that the positive portion of the ICSF always has to exceed the negative portion.<sup>10</sup>

The advantages associated with use of *both* the ICSF and the ICM go beyond the ability to distinguish coactive from parallel processing, as we will discuss in the next section. It is also intriguing that the OR and the AND serial stopping rules are now experimentally distinguishable, since in the OR case  $\text{ICSF} = 0$  always, but in the AND case there is a large negative portion of the ICSF, followed by an equally large positive portion. Thus, both the architecture and the stopping rule are experimentally determinable by the factorial tests carried out at the dis-

tributional level. The general applicability of the distributional approach has benefited from theoretical extensions by Schweickert and colleagues (e.g., Schweickert et al., 2000) to general feed-forward architectures, which contain parallel and serial subsystems, and from advances in methods of estimating entire RT distributions (see, in particular, Van Zandt, 2000, 2002).

**Doing it all: The double factorial paradigm.** As was just described, the application of factorial methods at the level of the RT distributions and the means allows strong tests of architecture and stopping rule. In fact, assuming selective influence at the distributional level, Townsend and Nozawa (1995) were able to test architecture and stopping rule, as well as capacity, within the same block of experimental trials.

This method, called the *double factorial paradigm*, exploits two kinds of relatively independent information and yields converging evidence about architecture, capacity, and stopping rule within the same experimental trials. As might be intuited from the name, the double factorial paradigm consists of two  $2 \times 2$  factorial designs, one "nested" within the other. Figure 10 illustrates this design for a task involving the detection of simple stimulus features.

At the outer level is a  $2 \times 2$  factorial design involving the state of each of the stimulus features (target and nontarget). The inner level involves a manipulation of the processing of each of these features, so that each state is processed at either a slow or a fast rate. This is the manipulation that instantiates the assumption of selective influence, discussed earlier. Potentially, any response assignment can be given to the observers, for instance, OR, AND, XOR, and so on. So far, OR and AND have been implemented in experiments. In the first, observers give a positive response when either or both of the features are in the target state. This corresponds to the logical OR and

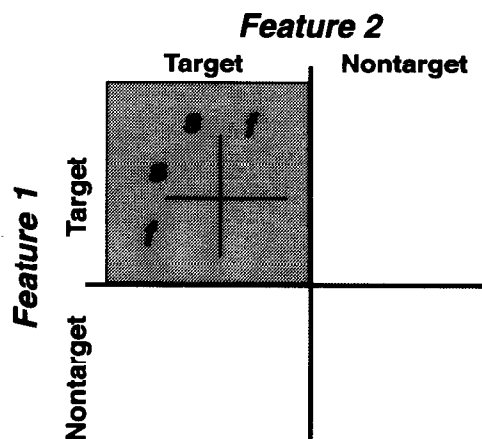


Figure 10. The double factorial paradigm for an experiment involving two stimulus features, each of which can be in a target or a nontarget state. It is assumed that there are experimental factors that can exert selective influence on the rate at which each of the features can be processed, so that each can be processed at either a fast (*f*) or a slow (*s*) rate.

is a condition that allows minimum-time processing on positive trials. In the second, observers give a positive response only when both features are in the target state. This corresponds to the logical AND and is a condition that requires exhaustive processing on positive trials. We focus here on the OR design.

The overall design provides two sources of evidence for testing hypotheses regarding characteristics of processing. The first pertains to hypotheses regarding architecture and involves the four cells in the upper left quadrant of the design (shaded gray in Figure 10). Here, processing load (number of features in the target state) is held constant, while processing speed is varied. The data from these cells allow the calculation of the two interaction contrasts (ICM and ICSF). The second pertains to hypotheses regarding capacity and involves those cells in which processing load is varied while processing rate is held constant. The data from these cells allow the calculation of the load capacity coefficients (see Townsend & Nozawa, 1995; Townsend & Wenger, in press).

Assuming the OR task design, Townsend and Nozawa (1995) demonstrated that coactivation models defined by general countersuperposition kinds of processing (which included the Poisson models mentioned above) were al-

ways super-capacity when individual channels did not slow down in the presence of additional (i.e., redundant) signals on other channels. In fact, it was shown that such models could be expected to produce violations of Miller's inequality. Applications of this theory and methodology have provided tests of configural hypotheses in the detection of facial features (Wenger & Townsend, 2001) and tests of the processing of multiple types of information in facial perception (Ingvalson & Wenger, in press).

**A guide to using the double factorial paradigm.** Although there is a small but growing set of papers that have described the use of the double factorial paradigm in some detail (see, in particular, Townsend & Nozawa, 1995; Wenger & Townsend, 2001), the potential allowed by the double factorial (in terms of strong-inference hypothesis testing) suggests that at least a brief procedural discussion is in order. Toward that end, and in order to encourage others to consider the use of the double factorial paradigm in their work, Figures 11-13 present *flow charts* for the application of the double factorial paradigm. Each of these flow charts begins in the lower left-hand corner, with the specific architecture under consideration. One can then work through a series of questions regarding the requisite characteristics of the experimental

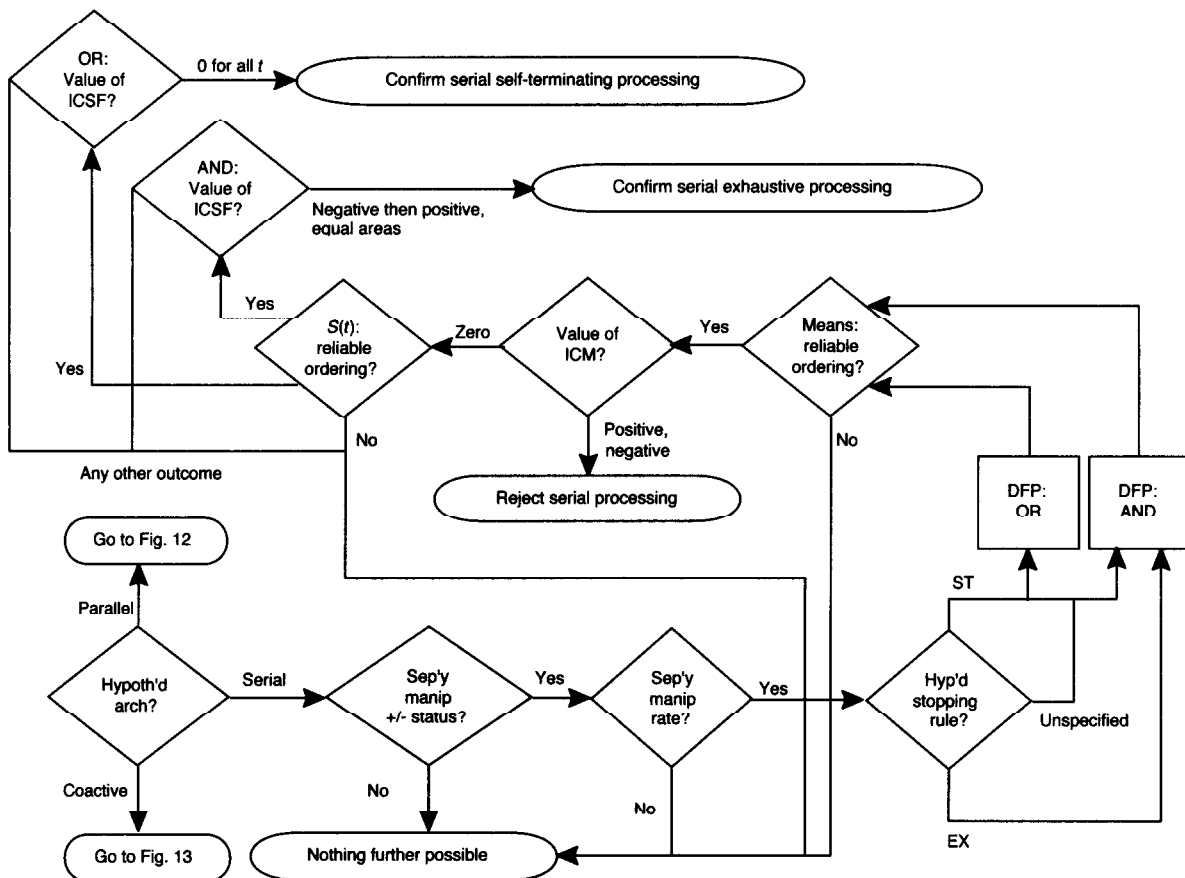


Figure 11. Flow chart indicating the procedural logic associated with applying the double factorial paradigm to a hypothesis regarding serial processing.

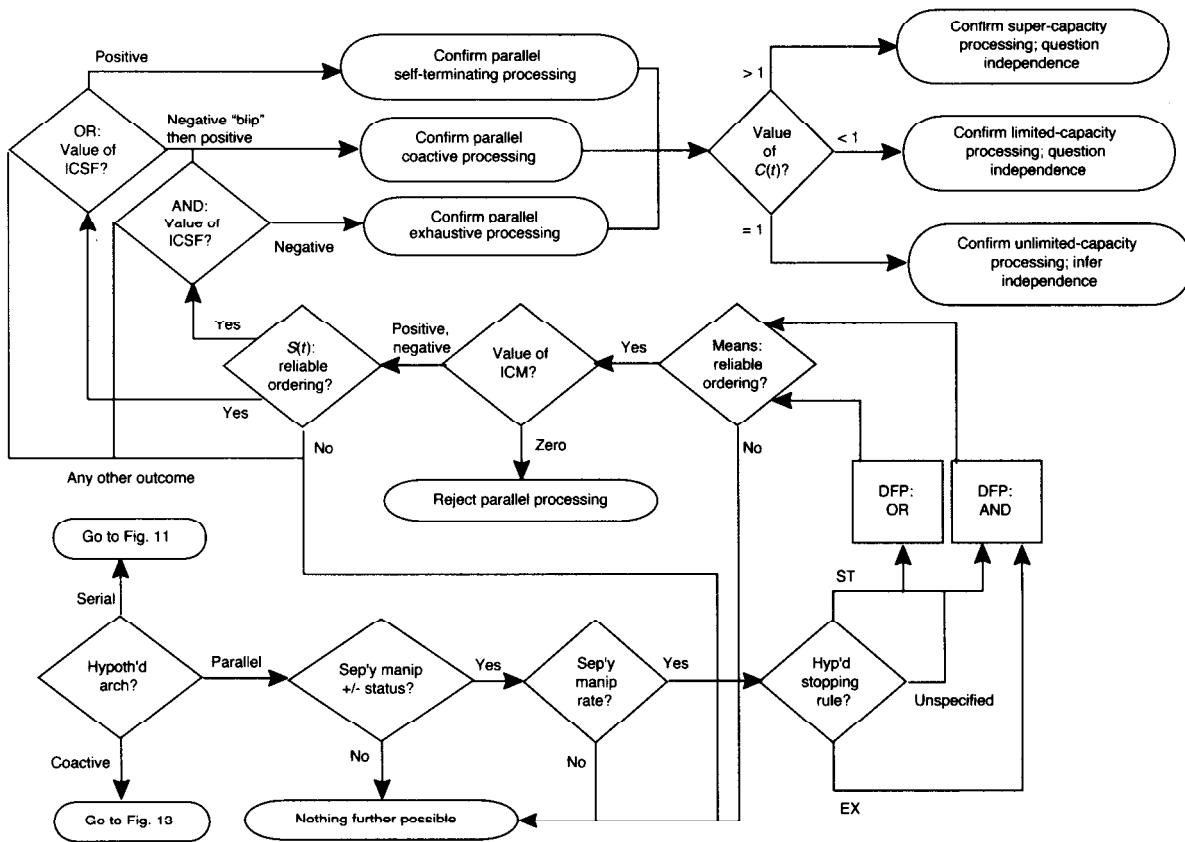


Figure 12. Flow chart indicating the procedural logic associated with applying the double factorial paradigm to a hypothesis regarding parallel processing.

manipulations and the characteristics of the data to the allowable conclusions (with one possibility being that no conclusion is possible).

To illustrate the use of these flowcharts, imagine that the hypothesis under consideration is serial processing (see Figure 11). The first two questions to be addressed are methodological. First (corresponding to the outer factorial in Figure 10), one must be able to independently manipulate the status of each of the elements to be processed with respect to the type of response required for each (positive or negative). Second (corresponding to the gray set of cells in the upper left quadrant of Figure 10), one must be able to independently manipulate the rate at which each of the elements can be processed. If either one or both of these questions is answered in the negative, nothing further is possible. Of course, in most experimental designs that the investigator would contemplate at the outset, these criteria are usually satisfied.

The next question to consider is the hypothesized stopping rule. If the hypothesis specifies a self-terminating stopping rule or is agnostic with respect to the stopping rule, one should design the experimental task so that the instructions to subjects include both OR and AND stopping rules. On the other hand, if the hypothesis specifies an exhaustive stopping rule, it is only necessary to instruct subjects to use an AND stopping rule.

Once the data have been collected, one begins the analysis by examining the means for the double-target trials. It is important that these be reliably ordered in a manner that is consistent with the rate manipulation. If this ordering is not reliable or an alternative ordering is observed, nothing further is possible, since the data are inconsistent with the critical assumption of selective influence.<sup>11</sup> On the other hand, if the ordering is reliable, it is possible to check the value of the interaction contrast at the level of the mean. If this interaction contrast is reliably different from 0 in either direction, it is necessary to reject the hypothesis of serial processing. Again, pilot work or previous literature usually has already assessed this criterion.

If the interaction contrast on the means is not reliably different from 0, one has preliminary evidence in favor of serial processing and can proceed to the distributional analyses. At this point, one examines the survivor functions on the double-target trials to determine whether they are reliably ordered (remembering here that an ordering on the means does not ensure an ordering on the survivor functions). If these functions are not reliably ordered, no further comparisons are possible. Alternatively, if the functions are reliably ordered in a manner that is consistent with the manipulation of rate, one can calculate the interaction contrast on the survivor func-

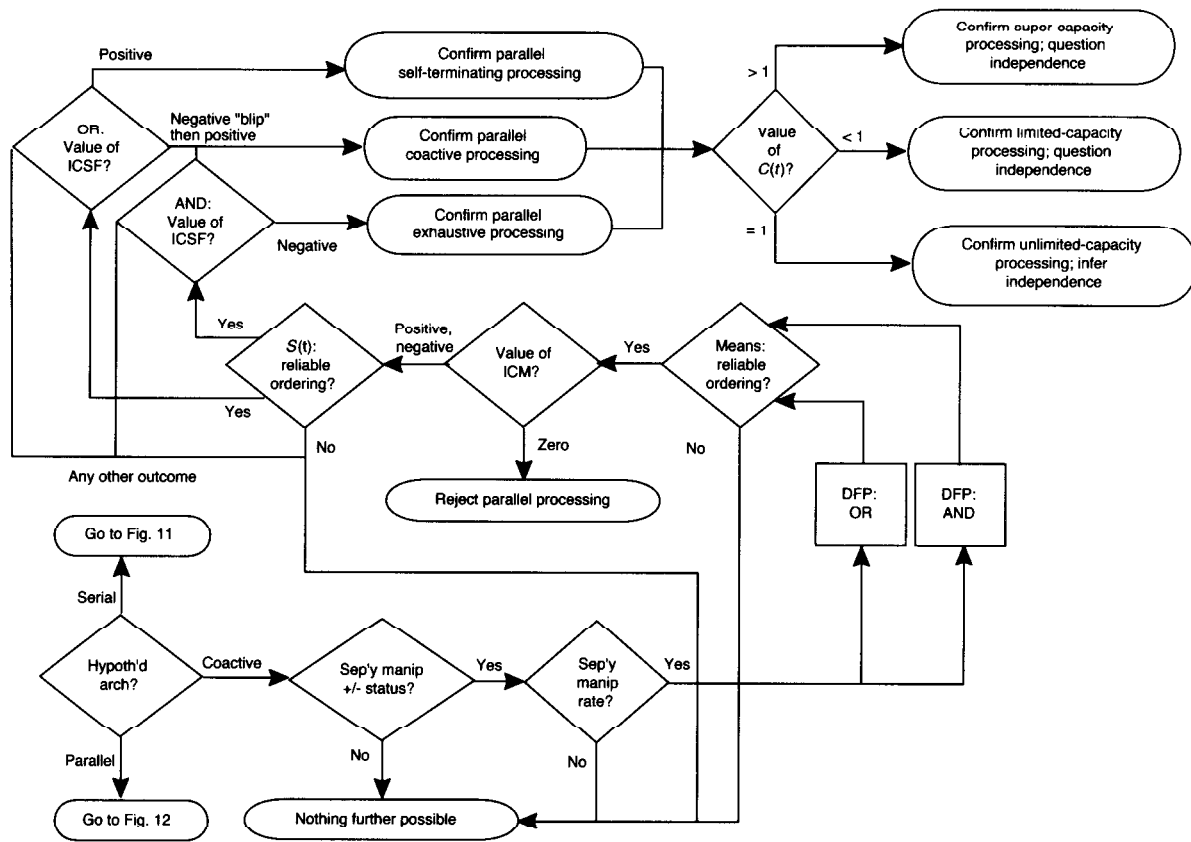


Figure 13. Flow chart indicating the procedural logic associated with applying the double factorial paradigm to a hypothesis regarding coactive processing.

tions. The critical outcomes are interaction contrasts that are 0 for all values of time or begin negative and go positive, with approximately equal areas associated with the negative and the positive values of the interaction contrast function. The first of these two is evidence for serial processing with a self-terminating stopping rule, and the second is evidence for serial processing with an exhaustive processing rule.

**Challenges for the Future**

The bringing together of mathematical representations of psychological hypotheses and experimental methodology has led to a useful stratagem that we call *meta-modeling*. It facilitates the development of experimental methodologies that are capable of testing broad classes of models, rather than highly specific individual models, against one another. This approach is especially helpful when attempting to settle well-defined and relatively simple issues, including the diametrically opposed concepts of parallel and serial processing. Meta-modeling has arguably led to striking advances in the ability to experimentally decide a number of elementary but major issues in human information processing. Meta-modeling may be difficult to impossible when the aim is to construct a test of an elaborate nexus of perceptual, cognitive, and

motor operations. In fact, some of our most esteemed investigators, such as Newell (1973), have questioned the benefits of attempting to settle simple binary issues and, instead, have striven to build mighty universal theories of cognitive behavior.<sup>12</sup>

We see much value in building general theories of cognition, as in Anderson's (e.g., Anderson & Labiere, 1998) evolving theoretical structures. Nonetheless, we wish to proselytize with regard to the considerable worth in testing fundamental issues, sometimes in a way harking to the strong-inference approach espoused by Platt (1964) indeed, as exhibited in Ross and Anderson (1981). Massaro (e.g., 1987, 1998), for one, seems to have made substantial progress proceeding in this vein (see also Townsend & Spencer-Smith, 2000). Although very complex unified theories, such as Newell's (1990) SOAR, have led to substantial progress in various domains, it may be debatable to what extent they have solved starkly simple, but highly strategic, issues such as the parallel-serial dilemma. However, many challenges remain even in the case of testing elementary issues. Some of these attend to general methodology and usage of RTs, as we now will briefly discuss.

One immediate challenge relates to parallel models whose channels interact but that otherwise preserve their

individual decision criteria. It has become clear that such parallel models can make predictions qualitatively in line with coactive models (Colonius & Townsend, 1997; Mordkoff & Yantis, 1991; Townsend & Wenger, in press). Thus, they can evince super-capacity behavior and predict violations of Miller's inequality. It is far from obvious, at this point in time, how to test true coactivation against interactively facilitative parallel processing with separate decision criteria.

Other looming but more long-standing challenges relate to the somewhat pesky residual time component introduced earlier. For instance, as was noted earlier, residual processes are usually assumed to be in series with, and stochastically independent of, the other processing components (Luce, 1986; Townsend & Ashby, 1983). However, this assumption may be and has been challenged and compared, for instance, with the extreme of perfect correlation, but still in a series arrangement of the residual component with the other stages of processing (Dzhafarov & Rouder, 1996).

The most time-honored approach has been to attempt to separate the residual time component from those under study, employing Fourier analysis, assuming stochastic independence of these components. This has proven to be a tricky and arduous strategy. Luce (1986), Smith (1990), and Sheu and Ratcliff (1995) have provided useful reviews and commentaries on this approach. Another kind of powerful, if often difficult, approach regarding the residual time component has been to prove that experimental predictions hold true regardless of its presence. This has been accomplished with reference to Miller's inequality (Ulrich & Giray, 1986) and distributional factorial strategies of testing parallel and serial processing (Ashby & Townsend, 1980; Dzhafarov & Schweickert, 1995; Roberts & Sternberg, 1993).

The residual time component also renders the testing of stochastic independence more problematic, due to the artifactual dependence introduced by its presence. In addition, dependence in perceptual or central processes is difficult in any case because, typically, a single response is all that is attained in experimental trials. One could, in principle, require two responses per trial and, thus, produce joint probability distributions, but there would be a cost of likely introducing unwanted contamination from more peripheral response selection and motor processes. It may be that accuracy measures, which seem more amenable to independence tests (e.g., Ashby & Townsend, 1986), can be allied with RT measures to better probe this question.

The residual time problem also abuts a more general possibility: that processes in the same forward sequence may not satisfy discrete flow but, rather, obey continuous flow properties. For instance, almost all systems based on differential and integral equations would fall into the continuous flow class. Some intriguing progress, both theoretical as well as empirical, has been made on certain subclasses of such systems (e.g., Ashby, 1982a, 1982b; McClelland, 1979; Meyer, Yantis, Osman, & Smith, 1985; Miller, 1993; Schweickert, 1989; Schweickert & Mounts, 1998), but much more remains to be done.

## CONCLUSION

At the outset of this article, we made three assertions: (1) The question of architecture will not, and *should not*, go away; (2) the question of architecture *cannot* be understood or answered out of context; and (3) the question of architecture *can* be answered using approaches that provide a tight coupling of theory and method. We hope to have persuaded those who may have been tempted to ignore the issue (hoping, perhaps, that psychological science might be susceptible to strategies for extinguishing unwanted behaviors) that avoiding the question of architecture is not going to make it go away. We hope to have convinced those who have focused on the question out of context that much can be gained by adopting a broader view. Finally, we hope to have convinced those who have assumed that there simply is no hope (at least with respect to the question of architecture) that there exist powerful and tractable solutions.

Simply disregarding the issue will not, of course, solve it. Simply acting "as if" processing is serial or parallel just because behavior is compatible with a certain, even preferred, type of serial or parallel processing is not likely to get very far. If there are plausible models from the alternative class that are equivalent to the preferred model, obviously we do not know much more than when we started. And for other alternatives, we have seen that the interactions of models according to their position on the various dimensions (capacity, stopping rule, dependence, etc.) can be extremely diverse, depending on the exact interaction among these dimensions. Even in global or unified theories, what guarantee do we have that assumptions about architecture, say, in one section of the theory will not affect predicted behavior and, therefore, the identifiability and testability of the entire enterprise?

Some of the most exciting prospects for the study of architecture and process models in general, with regard to the broad field of psychology, are encountered in emerging applications and interactions with other domains in cognition and psychology. For example, fascinating results have been forthcoming in applications of mathematical models and their associated experimental paradigms to clinical studies of psychopathology (e.g., McFall & Townsend, 1998; McFall & Treat, in press; Neufeld & McCarty, 1994). Theories in decision making, long the sole province of modified versions of subjective expected utility theory, are being influenced by quantitative information-processing theory (e.g., Busemeyer & Townsend, 1993), and certain developments from formal decision making are beginning to impact theory about process models (e.g., Marley & Colonius, 1992). Certain lines in this vein begin to employ processing concepts along with enriched notions of motivation and emotion (e.g., Mellers, Weber, Ordóñez, & Cooke, 1995). One avenue that is likely to be particularly fruitful with regard to providing converging evidence for model assumptions involves recent developments in cognitive neuroscience (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1999; Chelazzi, Miller, Duncan, & Desimone,



2001; Luck & Hillyard, 1995; Rahman, Sommer, & Schweinberger, 2002; Woodman & Luck, 1999). To be sure, it seems fair to say that most of the leads presently being followed by cognitive neuroscientists were first opened by behavioral cognitive scientists. On the other hand, there seems little doubt that the kinds of postulates made by each side in order to provide definitive tests are at least partially distinct, not to mention the promise of "peering at the hardwiring" offered by the neurosciences. This outlook proffers hope and optimism concerning synthesis of the two major avenues.

In closing, and in tribute to the successes associated with linking mathematical models to specific experimental designs, we return to the coincidence we noted in our initial paragraph. Certainly it is the case, with respect to the question of architecture, that "you can't always get what you want"—that is, if what you want is a strong and definite answer. Still, the powerful insights and experimental leverage that are possible with the theoretical and empirical tools we have reviewed suggests that, for the question of serial versus parallel processing, "if you try sometime, you just might find . . . you get what you need" (Richards & Jagger, 1969).

#### REFERENCES

- ANDERSON, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- ANDERSON, J. R., & LABIERE, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- ASHBY, F. G. (1982a). Deriving exact predictions from the cascade model. *Psychological Review*, **89**, 599-607.
- ASHBY, F. G. (1982b). Testing the assumptions of exponential, additive reaction time models. *Memory & Cognition*, **10**, 125-134.
- ASHBY, F. G. (1992). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G., ALFONSO-REESE, L. A., TURKEN, A. U., & WALDRON, E. M. (1999). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, **105**, 442-481.
- ASHBY, F. G., TEIN, J. Y., & BALAKRISHNAN, J. D. (1993). Response time distributions in memory scanning. *Journal of Mathematical Psychology*, **37**, 526-555.
- ASHBY, F. G., & TOWNSEND, J. T. (1980). Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology*, **21**, 93-123.
- ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.
- ATKINSON, R. C., HOLMGREN, J. E., & JUOLA, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, **6**, 321-326.
- BAMBER, D. (1969). Reaction times and error rates for "same"-"different" judgments of multidimensional stimuli. *Perception & Psychophysics*, **6**, 169-174.
- BAMBER, D., & VAN SANTEN, J. P. H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, **29**, 443-473.
- BAMBER, D., & VAN SANTEN, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, **44**, 20-40.
- BRICOLO, E., GIANESINI, T., FANINI, A., BUNDESEN, C., & CHELAZZI, L. (2002). Serial attention mechanisms in visual search: A direct behavioral demonstration. *Journal of Cognitive Neuroscience*, **14**, 980-993.
- BRIDGEMAN, B. (1971). Metacontrast and lateral inhibition. *Psychological Review*, **78**, 528-539.
- BUNDESEN, C., & SHIBUYA, H. (Eds.) (1999). *Visual selective attention*. Mahwah, NJ: Erlbaum.
- BUSEMEYER, J. B., & TOWNSEND, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, **100**, 432-459.
- BUSEY, T. A., & LOFTUS, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review*, **101**, 446-469.
- BUSEY, T. A., & TOWNSEND, J. T. (2001). Independent sampling vs inter-item dependencies in whole report processing: Contributions of processing architecture and variable attention. *Journal of Mathematical Psychology*, **45**, 283-323.
- CAVE, K. R., & WOLFE, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, **22**, 225-271.
- CHELAZZI, L., MILLER, E. K., DUNCAN, J., & DESIMONE, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral Cortex*, **11**, 761-772.
- COLONIUS, H., & TOWNSEND, J. T. (1997). Activation-state representation of models for the redundant-signals-effect. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 245-254). Mahwah, NJ: Erlbaum.
- COLONIUS, H., & VORBERG, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology*, **38**, 35-58.
- COX, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, **B34**, 187-220.
- DONDERS, F. C. (1969). Over die snelheid van psychische processen [On the rapidity of mental processes] (W. G. Koster, Trans.). *Acta Psychologica*, **30**, 412-431. (Original work published 1868)
- DONNELLY, N., FOUND, A., & MÜLLER, H. J. (1999). Searching for impossible objects: Processing form and attributes in early vision. *Perception & Psychophysics*, **61**, 675-690.
- DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, **113**, 501-517.
- DZHAFAROV, E. N. (1992). The structure of simple reaction time to step-function signals. *Journal of Mathematical Psychology*, **36**, 235-268.
- DZHAFAROV, E. N. (1997). Process representations and decompositions of response times. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 255-277). Mahwah, NJ: Erlbaum.
- DZHAFAROV, E. N., & ROUDER, J. N. (1996). Empirical discriminability of two models for stochastic relationship between additive components of response time. *Journal of Mathematical Psychology*, **40**, 48-63.
- DZHAFAROV, E. N., & SCHWEICKERT, R. (1995). Decompositions of response times: An almost general theory. *Journal of Mathematical Psychology*, **39**, 285-314.
- EGETH, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, **1**, 245-252.
- EGETH, H. E., & DAGENBACH, D. (1991). Parallel versus serial processing in visual search: Further evidence from subadditive effects of visual quality. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 550-560.
- ERIKSEN, C. W., & SPENCER, T. (1969). Rate of information processing in visual perception: Some results and some methodological considerations. *Journal of Experimental Psychology: General*, **79**, 1-16.
- FISHER, D. L. (1982). Limited-channel models of automatic detection: Capacity in scanning in visual search. *Psychological Review*, **89**, 662-692.
- FISHER, D. L. (1984). Central capacity limits in consistent mapping, visual search tasks: Four channels or more? *Cognitive Psychology*, **16**, 449-484.
- FOLK, C. L., REMINGTON, R. W., & JOHNSTON, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 1030-1044.
- FRANCIS, G. (2000). Quantitative theories of metacontrast masking. *Psychological Review*, **107**, 768-785.
- GIBSON, B. S., & AMELIO, J. (2000). Inhibition of return and attentional control settings. *Perception & Psychophysics*, **62**, 496-504.

- GIBSON, B. S., & KELSEY, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for displaywide visual features. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 699-706.
- GOLDSTONE, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 86-112.
- GRICE, G. R., CANHAM, L., & BOROUGHS, J. M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics*, **35**, 451-463.
- GRICE, G. R., CANHAM, L., & GWYNNE, J. W. (1984). Absence of a redundant-signals effect in a reaction time task with divided attention. *Perception & Psychophysics*, **36**, 565-570.
- INGVALSON, E. M., & WENGER, M. J. (in press). A strong test of the dual-mode hypothesis. *Perception & Psychophysics*.
- KADLEC, H., & TOWNSEND, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology*, **36**, 325-374.
- KADLEC, H., & TOWNSEND, J. T. (1992b). Signal detection analysis of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181-228). Hillsdale, NJ: Erlbaum.
- KAHNEMAN, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- KANWISHER, N., McDERMOTT, J., & CHUN, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, **17**, 4302-4311.
- KANWISHER, N., STANLEY, D., & HARRIS, A. (1999). The fusiform face area is selective for faces not animals. *NeuroReport*, **10**, 183-187.
- LACHMAN, R., LACHMAN, J. L., & BUTTERFIELD, E. C. (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Erlbaum.
- LAMING, D. (1973). *Mathematical psychology*. New York: Academic Press.
- LINK, S. W., & HEATH, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77-105.
- LOFTUS, G. R. (1991). On the tyranny of hypothesis testing in the social sciences. *Contemporary Psychology*, **36**, 102-105.
- LOGAN, G. D. (2002). Parallel and serial processing. In J. Wixted (Ed.), *Stevens' handbook of experimental psychology: Methodology in experimental psychology* (3rd ed., Vol. 4, pp. 271-300). New York: Wiley.
- LUCE, R. D. (1986). *Reaction times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- LUCK, S. J., & HILLYARD, S. A. (1995). The role of attention in feature detection and conjunction discrimination: An electrophysiological analysis. *International Journal of Neuroscience*, **80**, 281-297.
- MARLEY, A. A. J., & COLONIUS, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *Journal of Mathematical Psychology*, **36**, 1-20.
- MARR, D. (1982). *Vision*. New York: Freeman.
- MASSARO, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.
- MASSARO, D. W. (1998). *Perceiving talking faces: From speech perception to a general principle*. Cambridge, MA: MIT Press, Bradford Books.
- MAXWELL, S. E., & DELANEY, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.
- McCLELLAND, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, **86**, 287-330.
- McFALL, R. M., & TOWNSEND, J. T. (1998). Foundations of psychological assessment: Implications for cognitive assessment in clinical science. *Psychological Assessment*, **10**, 316-330.
- McFALL, R. M., & TREAT, T. (in press). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology*.
- MELLERS, B. A., WEBER, E. U., ORDÓÑEZ, L. D., & COOKE, A. D. J. (1995). Utility invariance despite labile preferences. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Decision making from a cognitive perspective: Psychology of learning and motivation* (Vol. 32, pp. 221-246). San Diego: Academic Press.
- MEYER, D. E., YANTIS, S., OSMAN, A. M., & SMITH, J. E. K. (1985). Temporal properties of human information processing: Tests of discrete versus continuous models. *Cognitive Psychology*, **17**, 445-518.
- MILLER, J. O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, **14**, 247-279.
- MILLER, J. O. (1993). A queue-series model for reaction time, with discrete-stage and continuous-flow models as special cases. *Psychological Review*, **100**, 702-715.
- MORDKOFF, R. T., & YANTIS, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 520-538.
- MURDOCK, B. B., JR. (1971). A parallel-processing model for scanning. *Perception & Psychophysics*, **10**, 289-291.
- MURDOCK, B. B., JR. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.
- NEISSER, U. (1964). Visual search. *Scientific American*, **210**, 94-102.
- NEUFELD, R. W. J., & McCARTY, T. S. (1994). A formal analysis of stressor and stress-proneness effects on simple information processing. *British Journal of Mathematical & Statistical Psychology*, **47**, 193-226.
- NEWELL, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press.
- NEWELL, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- O'TOOLE, A. J., WENGER, M. J., & TOWNSEND, J. T. (2001). Quantitative models of perceiving and remembering faces: Precedents and possibilities. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 1-38). Mahwah, NJ: Erlbaum.
- PALMER, J. (1995). Attention and visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, **4**, 118-123.
- PALMER, J. (1998). Attentional effects in visual search: Relating search accuracy and search time. In R. D. Wright (Ed.), *Visual attention* (pp. 348-388). New York: Oxford University Press.
- PASHLER, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, **116**, 220-244.
- PLATT, J. R. (1964). Strong inference. *Science*, **146**, 347-353.
- POPPER, K. R. (1935). *Logik der Forschung* [The logic of research]. Berlin: Springer-Verlag.
- RAHMAN, R. A., SOMMER, W., & SCHWEINBERGER, S. R. (2002). Brain potential evidence for the time course of access to biographical facts and names of familiar persons. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 366-373.
- RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
- RICHARDS, K., & JAGGER, M. (1969). You can't always get what you want. In *Let it bleed*. London: London/Decca Records.
- ROBERTS, S., & STERNBERG, S. (1993). The meaning of additive reaction time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611-654). Cambridge, MA: MIT Press.
- ROSS, B. H., & ANDERSON, J. R. (1981). A test of parallel versus serial processing applied to memory retrieval. *Journal of Mathematical Psychology*, **24**, 183-223.
- SANDERS, A. F. (1990). Issues and trends in the debate on discrete versus continuous processing of information. *Acta Psychologica*, **74**, 123-167.
- SANDERS, A. F., WIJNEN, J. L. C., & VAN ARKEL, A. E. (1982). An additive factor analysis of the effects of sleep loss on reaction processes. *Acta Psychologica*, **51**, 41-59.
- SCHNEIDER, W., & SHIFFRIN, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, **84**, 1-66.
- SCHWEICKERT, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology*, **18**, 105-139.

- SCHWEICKERT, R. (1982). The bias of an estimate of coupled slack in stochastic PERT networks. *Journal of Mathematical Psychology*, **26**, 1-12.
- SCHWEICKERT, R. (1983). Latent network theory: Scheduling of processes in sentence verification and the Stroop effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 353-383.
- SCHWEICKERT, R. (1989). Separable effects of factors on activation functions in discrete and continuous models:  $d'$  and evoked potentials. *Psychological Bulletin*, **106**, 318-328.
- SCHWEICKERT, R. (1993). Information, time, and the structure of mental events: A twenty-five-year review. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 535-566). Cambridge, MA: Academic Press.
- SCHWEICKERT, R., & GIORGINI, M. (1999). Response time distributions: Some simple effects of factors selectively influencing mental processes. *Psychonomic Bulletin & Review*, **6**, 269-288.
- SCHWEICKERT, R., GIORGINI, M., & DJZHAFAROV, E. (2000). Selective influence and response time cumulative distribution functions in serial-parallel task networks. *Journal of Mathematical Psychology*, **44**, 504-535.
- SCHWEICKERT, R., & MOUNTS, J. (1998). Additive effects of factors on reaction time and evoked potentials in continuous-flow models. In C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement* (pp. 311-327). Mahwah, NJ: Erlbaum.
- SCHWEICKERT, R., & TOWNSEND, J. T. (1989). A trichotomy method: Interactions of factors prolonging sequential and concurrent mental processes in stochastic PERT networks. *Journal of Mathematical Psychology*, **33**, 328-347.
- SCHWEICKERT, R., & WANG, Z. (1993). Effects of response time on factors selectively influencing processes in acyclic task networks with OR gates. *British Journal of Mathematical & Statistical Psychology*, **46**, 1-30.
- SHEU, C. F., & RATCLIFF, R. (1995). The application of Fourier deconvolution to reaction time data: A cautionary note. *Psychological Bulletin*, **118**, 285-299.
- SHIFFRIN, R. M., & GARDNER, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology*, **93**, 72-82.
- SMITH, P. L. (1990). Obtaining meaningful results from Fourier deconvolution of reaction time data. *Psychological Bulletin*, **108**, 533-550.
- SNODGRASS, J. G. (1972). Reaction times for comparisons of successively presented visual patterns: Evidence for serial self-terminating search. *Perception & Psychophysics*, **12**, 364-372.
- SNODGRASS, J. G., & TOWNSEND, J. T. (1980). Comparing parallel and serial models: Theory and implementation. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 330-354.
- SPELBERG, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, **74**(Whole No. 498).
- SPELBERG, G. (1963). A model for visual memory tasks. *Human Factors*, **5**, 19-31.
- SPELBERG, G. (1967). Successive approximations to a model for short-term memory. *Acta Psychologica*, **27**, 285-292.
- STERNBERG, S. (1966). High-speed scanning in human memory. *Science*, **153**, 652-654.
- STERNBERG, S. (1969a). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276-315). Amsterdam: North-Holland.
- STERNBERG, S. (1969b). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, **4**, 421-457.
- STERNBERG, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Vol. 4. Methods, models and conceptual issues* (pp. 739-811). Cambridge, MA: MIT Press.
- STERNBERG, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica*, **106**, 147-246.
- TAYLOR, D. A., TOWNSEND, J. T., & SUDEVAN, P. (1978, November). *Analysis of intercompletion times in multi-element processing*. Paper presented at the 19th Annual Meeting of the Psychonomic Society, San Antonio, TX.
- THEEUWES, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 799-806.
- THEEUWES, J., ATCHLEY, P., & KRAMER, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 105-124). Cambridge, MA: MIT Press.
- THOMAS, E. A. (1971). Sufficient conditions for monotone hazard rate: An application to latency-probability curves. *Journal of Mathematical Psychology*, **8**, 303-332.
- THOMAS, R. D. (1995). Gaussian general recognition theory and perceptual independence. *Psychological Review*, **102**, 192-200.
- THOMAS, R. D. (2001). Characterizing perceptual interactions in face identification using multidimensional signal detection theory. In M. J. Wenger & J. I. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contests and challenges* (pp. 193-228). Mahwah, NJ: Erlbaum.
- TOWNSEND, J. T. (1969, April). *Mock parallel and serial models and experimental detection of these*. Paper presented at the Purdue Centennial Symposium on Information Processing, West Lafayette, IN.
- TOWNSEND, J. T. (1971a). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, **10**, 161-163.
- TOWNSEND, J. T. (1971b). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, **9**, 40-50.
- TOWNSEND, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical & Statistical Psychology*, **25**, 168-199.
- TOWNSEND, J. T. (1974). 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* (pp. 133-168). Hillsdale, NJ: Erlbaum.
- TOWNSEND, J. T. (1976a). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology*, **14**, 219-238.
- TOWNSEND, J. T. (1976b). A stochastic theory of matching processes. *Journal of Mathematical Psychology*, **14**, 1-52.
- TOWNSEND, J. T. (1981). Some characteristics of visual whole report behavior. *Acta Psychologica*, **47**, 149-173.
- TOWNSEND, J. T. (1990a). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should be) distinguished. *Psychological Sciences*, **1**, 46-54.
- TOWNSEND, J. T. (1990b). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, **108**, 551-567.
- TOWNSEND, J. T. (1994). Methodology and statistics in the behavioral sciences: The old and the new. *Psychological Science*, **5**, 321-325.
- TOWNSEND, J. T. (2001). A clarification of self-terminating versus exhaustive variances in serial and parallel models. *Perception & Psychophysics*, **63**, 1101-1106.
- TOWNSEND, J. T., & ASHBY, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3, pp. 200-239). Hillsdale, NJ: Erlbaum.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- TOWNSEND, J. T., & COLONIUS, H. (1997). Parallel processing response times and experimental determination of the stopping rule. *Journal of Mathematical Psychology*, **41**, 392-397.
- TOWNSEND, J. T., & COLONIUS, H. (2003). *Variability of the MAX and MIN statistic: A theory of the quantile spread as a function of sample size*. Manuscript submitted for publication.
- TOWNSEND, J. T., & DIEDERICH, A. (1996). The relationship of variance to interaction contrast in parallel systems factorial technology. *British Journal of Mathematical & Statistical Psychology*, **49**, 211-223.
- TOWNSEND, J. T., & FIFIC, M. (in press). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception & Psychophysics*.
- TOWNSEND, J. T., & FIKES, T. (1995). *A beginning quantitative taxonomy of cognitive activation systems and application to continuous flow processes* (Tech. Rep. 131). Bloomington: Indiana University, Cognitive Science Program.
- TOWNSEND, J. T., & NOZAWA, G. (1988, November). *Strong evidence for parallel processing with dot stimuli*. Paper presented at the 29th Annual Meeting of the Psychonomic Society, Chicago.

- TOWNSEND, J. T., & NOZAWA, G. (1995). On the spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*, 321-359.
- TOWNSEND, J. T., & SCHWEICKERT, R. (1989). Toward the trichotomy method: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology*, *33*, 309-327.
- TOWNSEND, J. T., & SPENCER-SMITH, J. (2000). A tome on tests and extensions of the FLMP. *American Journal of Psychology*, *112*, 449-489.
- TOWNSEND, J. T., & VAN ZANDT, T. (1990). New theoretical results on testing self-terminating vs. exhaustive processing in rapid search experiments. In H. Geissler (Ed.), *Psychological explorations of mental structures* (pp. 469-489). Toronto: Hogrefe & Huber.
- TOWNSEND, J. T., & WENGER, M. J. (in press). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*.
- TREISMAN, A. M., & GORMICAN, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*, 15-48.
- ULRICH, R., & GIRAY, M. (1986). Separate-activation models with variable base times: Testability and checking of cross-channel dependency. *Perception & Psychophysics*, *39*, 248-254.
- VAN ZANDT, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*, 424-465.
- VAN ZANDT, T. (2002). Analysis of response time distributions. In J. T. Wixted (Vol. Ed.) & H. Pashler (Series Ed.), *Stevens' Handbook of experimental psychology: Vol. 4. Methodology in experimental psychology* (3rd ed., pp. 461-516). New York: Wiley.
- VAN ZANDT, T., & RATCLIFF, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin & Review*, *2*, 50-54.
- VAN ZANDT, T., & TOWNSEND, J. T. (1993). Self-terminating versus exhaustive processes in rapid visual and memory search: An evaluative review. *Perception & Psychophysics*, *53*, 563-580.
- VORBERG, D. (1977, August). *On the equivalence of parallel and serial models of information processing*. Paper presented at the 10th Annual Meeting of the Society for Mathematical Psychology, Los Angeles.
- VORBERG, D. (1989). *Within-stage independence, future-order independence, and realizability of parallel processing systems*. Unpublished manuscript.
- VORBERG, D., & ULRICH, R. (1987). Random search with unequal search rates: Serial and parallel generalizations of McGill's model. *Journal of Mathematical Psychology*, *31*, 1-23.
- WEIBULL, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, *18*, 293-297.
- WENGER, M. J., & GIBSON, B. S. (in press). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology: Human Perception & Performance*.
- WENGER, M. J., SCHUSTER, C., PETERSEN, L. E., & PETERSEN, R. C. (in press). Application of proportional hazards models to reaction time data. In C. Bergeman & S. Boker (Eds.), *Notre Dame series on quantitative methodologies*. Mahwah, NJ: Erlbaum.
- WENGER, M. J., SCHUSTER, C., & TOWNSEND, J. T. (2001, November). *The role of capacity in organized perception: Tools for hypothesis testing*. Paper presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL.
- WENGER, M. J., SCHUSTER, C., & TOWNSEND, J. T. (2002). *Testing capacity hypotheses: Applying proportional hazards and frailty models*. Paper presented at the 35th Annual Meeting of the Society for Mathematical Psychology, Miami University of Ohio, Oxford.
- WENGER, M. J., & TOWNSEND, J. T. (2001). Faces as gestalt stimuli: Process characteristics. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition* (pp. 229-284). Mahwah, NJ: Erlbaum.
- WOODMAN, G. F., & LUCK, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual research. *Nature*, *400*, 867-869.

## NOTES

- Note that we are making a distinction between the falsification of a class of models (e.g., a set of models that share a particular computational characteristic) and the falsification of one particular model. Note also that we are making a distinction between the testing and development of alternative classes of models and the testing of hierarchically related versions of a general model, an approach that has been quite successful in statistical modeling (e.g., Maxwell & Delaney, 1990).
- In addition, we would contend that the most widely held interpretation of the model-mimicking problem, as it pertains to the distinction between serial and parallel processing, is that it is a problem without a solution. One intent of this review is to offer a decisive and detailed refutation of that interpretation.
- Interestingly, an occurrence of strong position effects could be taken as being generally consistent with serial processing, since it could be argued that there is no good reason for a parallel model, even under environmental and contextual, instructional or inherent asymmetries, to produce that result.
- Interestingly, he found that the rather arduous procedure of estimating the residual time distribution did not affect the results significantly over that found in disregarding its contribution. Unfortunately, at this point in time, there is no a priori way to make this determination in advance of actually trying it!
- The general concept of efficiency is naturally related to speed, and we could measure speed with mean RT, the cumulative distribution function, the hazard function, the likelihood function, and so on. Any of these might be appropriate for different purposes. Nevertheless, the foregoing list is in terms of increasing detail and inferential power, since an ordering of the latter statistics (i.e., those toward the end of the list) force an ordering of the earlier statistics, but not vice versa (Townsend, 1990b). Hence, the power of the scientific conclusion will be greater when hazard functions are used than when mean RTs are used.
- It should be noted that these difficulties are ones associated with estimating the parametric form of an individual hazard function. There is a literature on methods of comparisons among hazard functions (e.g., from different experimental conditions) that do not require parametric assumptions (dating to Cox, 1972) that are showing some promise in application to RT data (e.g., Wenger & Gibson, in press; Wenger, Schuster, Petersen, & Petersen, in press; Wenger, Schuster, & Townsend, 2001, 2002).
- We here follow the convention adopted by the first author in a number of earlier publications, using  $f$  when discussing serial models and  $g$  when discussing parallel models.
- There is some question as to the extent to which the order dependence can be employed to distinguish serial and parallel models (Vorberg, 1989). This issue is intriguing because of the potential implications for our psychological and philosophical notions of causality, as well as for pragmatic testing methods, and work continues along this avenue.
- Of course, even the serial curves are not typically true shifts of one another; usually, the variance is changing across positions too.
- This is because of the relationship between the mean and the survivor function:

$$E[T] = \int_0^{\infty} S(t) dt.$$

- We should emphasize that a statistical test for a reliable ordering indicates only that a manipulation was effective. It is not direct evidence for selective influence being in force.
- To be sure, Newell (personal communication, circa 1985) later stated that he was more interested in generating some discussion of this important issue, rather than necessarily simply "shooting it down."

(Manuscript received March 5, 2003;  
accepted for publication June 26, 2003.)