Chapter 11

Experimental Discrimination of the World's Simplest and Most Antipodal Models: The Parallel-Serial Issue

James T. Townsend, Haiyuan Yang, and Devin M. Burns

Indiana University

In general cognitive systems are comprised of more than a single subprocess. The arrangement and linkages of these subprocesses are known as "mental architecture". The simplest non-trivial systems or manner of carrying out multi-tasking on discrete items is found in two diametrically opposite models that have been classically used to describe the architecture of this processing. *Serial* systems allow only one subsystem at a time to operate. The antithesis of serial processing assumes that all subsystems operate simultaneously, or in *parallel*. Mathematical characterizations of these and other architectures have been developed. Models of serial vs. parallel systems, seemingly so distinct, are shockingly hard to tell apart with experimental methodology. Here we review the history of psychological work on the subject, both theoretical and experimental. We also describe three additional properties of systems that must be considered simultaneously (and have often been conflated with architecture): stopping rule, dependency, and capacity. We present formal mathematical descriptions that can allow us to examine in which cases the two will be indistinguishable, and summarize successful paradigms for differentiating the models in the context of various psychological tasks.

11.1. Brief Review

The question of whether people can perform multiple perceptual or mental operations simultaneously, referred to as parallel processing, has intrigued psychologists at least since the late 19th century (Sternberg, 1969; Schweickert, 1978, 1982, 1985; Schweickert & Wang, 1993; Townsend, 1984; Townsend & Ashby, 1983; Townsend & Schweickert, 1989). This innocuous sounding question has turned out to be far more complicated than one might expect, however, since it is not possible (at least in the foreseeable future) to peer inside and count the number of operative channels in our brain at any point in time. A major branch of research resides within Reaction Time (hereafter RT) methodology (e.g., Egeth, 1966; Eriksen & Spencer, 1969; Estes & Taylor, 1964; Garner, 1974; Hick, 1952; Luce, 1986; Neisser, 1967; Posner, 1978; Smith, 1968; Sternberg, 1966; Welford, 1980). The reader is referred to the older tomes, (Welford, 1980; Laming, 1968), as well as newer treatments, (Luce, 1986; Townsend & Ashby, 1983; Zandt, 2000), for general RT surveys and tutorials.

Perhaps the first empirical attempt to answer the parallel-serial question was Sir William Hamilton (Hamilton, 1859). In his informal "experiment" several dice were tossed onto a desk, and he set himself the task of instantaneously assessing the total number of dots showing. From this operation, he attempted to determine the number of objects that could be apprehended simultaneously (in parallel) by a human observer.

Next we find a true pioneer in using the measure of reaction time to explore internal psychological processes. F. C. Donders, a Dutch psychologist, was one of the first to use differences in human reaction time to infer distinct subsystems involved in cognitive processing. The concept of using subtraction to decompose whole reaction times into component parts is still one of the most common tools used for making inferences about mental processes such as learning, memory and attention. Donders (1868) developed the method of subtraction, in which mean reaction times from two different tasks are compared. The second task is designed to require all the stages of the first plus an additional stage. Thus, the difference between mean RTs is taken to be an estimate of the mean duration time of the interpolated stage. The work of Donders was based on the idea that various psychological subprocesses are carried out in a serial fashion, that is, in a series with no overlap in processing time. Such serially arranged psychological systems are now referred to as *Dondersian Systems* (See Fig. 11.1).

This method, however, later became the subject of criticism for several reasons. First, the assumption, known as *pure insertion*, that when the task is changed to insert an additional stage, the other stages will remain unchanged is questionable. If this condition is not met, then the difference between RTs cannot be identified as the duration of the inserted stage. Second, if the underlying subprocesses are carried out with any overlap in processing time (as happens in parallel processing), then the method of subtraction will introduce significant error and will typically underestimate the contributions of the subprocess to the overall RT (Taylor, 1976; Townsend, 1984).

Although the method of subtraction was not wholeheartedly accepted,

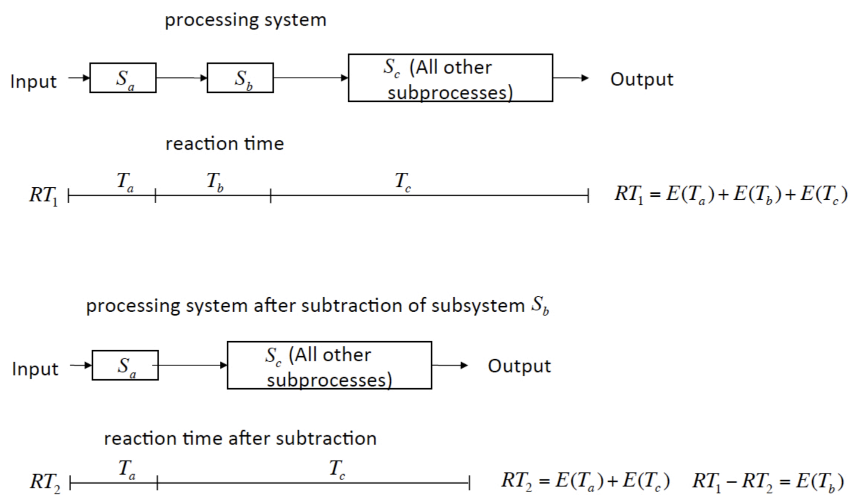

Fig. 11.1. A schematic illustrating the assumptions of Donders' subtractive method.

the attempt to analyze RT into components went on. During the early advent of cognitive science and in particular, the information processing approach, Sternberg (1966) and Egeth (1966) were instrumental in beginning to assay predictions associated with serial vs. parallel systems. Sternberg (1966) examined how reaction times in a short-term memory search experiment changed with an increasing number of items in memory. When Sternberg found that reaction time increased linearly with the number of targets, he inferred that this pattern could only be expected from a serial model. Later work has shown that this pattern is more indicative of capacity limitations than of architecture, but we will get to that later. Interestingly, in a slightly different visual search task, Egeth (1966) found evidence more in favor of parallel processing.

Three years later Sternberg developed a new tool which would subsequently usher in a more potent and universal set of methodologies, called the *additive factors method* (e.g., Sternberg, 1969; see also Ashby & Townsend, 1980; Pachella, 1974; Pieters, 1983; Townsend, 1984). These studies also provide general caveats with regard to factorial methods. As opposed to Donders' method, the additive factors method does not require the postulate that one can insert or take away the contribution of complete stages through experimental changes while leaving the other stages unaltered. Rather, it requires only the weaker assumption of selective influence: that there can be found two (or more) experimental factors that influence disjoint subsets of the stages. These factors can then be manipulated to answer questions about processing architecture and independence. This framework heralded the beginning of work attempting to answer these questions from a more formal, theoretical, information-processing approach. The cognitive revolution and dawn of computing brought a new perspective on cognitive modeling.

Investigators soon began to notice that serial/parallel identification is not a simple isolated question; there are other important facets that need to be considered. Four critical issues or theoretical dimensions that serve to help determine the nature of the processing system are: 1. *Architecture*; 2. *Stopping rule*; 3. *Dependency*; 4.*Capacity*. It is to be emphasized that these issues are all logically independent of one another (Townsend, 1974), in the sense that any combinations of values of the above four dimensions might be found in physically realizable systems. However, certain combinations may be more intuitively or psychologically acceptable than others. Although our focus here is on architecture (serial, parallel, and others), we must provide a brief summary of what these other dimensions are for proper understanding of the models.

Stopping rule: No predictions can be made about processing times until the model designer has a rule for when processing stops. Two main cases of interest are: first-terminating processing and last-terminating (= exhaustive) processing. A first-terminating rule means that processing can be completed as soon as any one of the channels is finished. When paired with a parallel architecture, this produces the familiar race model. This case is often called an OR design because completion of any item from a set of presented items is sufficient to stop processing and ensure a correct response. (e.g., Egeth, 1966).

If all items or channels must be processed to ensure a correct response, then exhaustive or AND processing is entailed. For instance, in a target absent trial for a visual search task, every item must be examined to guarantee no targets are present. Figure 11.2 provides schematics of these stopping rules in a serial/parallel system.

Dependency: Models have classically assumed independent information channels, but there are many situations where this presumption does not make sense. There are myriad different ways to allow "cross-talk", or interaction between the channels. In other words, there are a number of types of independence that might be of interest. One type of independence we emphasize is within-stage independence (see Section 11.2 for detailed explanation). Additionally, there are two other types of independence which are

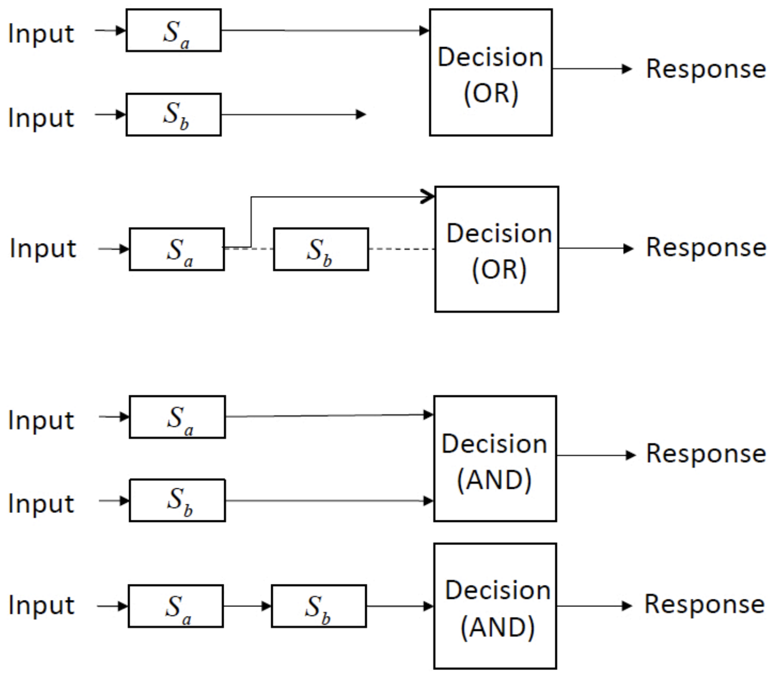

Fig. 11.2. Schematics of different stopping rules and architectures.

related: across-stage independence and independence of total completion time. The former case indicates that knowing the element in position a finishes first and at time t_{a1} tells us nothing about how much longer it will take the element in b to finish. The latter case indicates that the overall time to process a and the overall time to process b are independent. In all cases, we shall investigate the independence in a strict probabilistic sense: two events are independent if and only if the chance that they both happen simultaneously is the product of the chances that each occurs individually, that is, $P(AB) = P(A)P(B)$.

Capacity: This is a generalized notion of the amount of resources available to a given channel. If we say a system has *unlimited capacity* we mean that the overall time to process a single item does not vary with the total number of items undergoing simultaneous processing. The implication is that as a whole, the individual channels are neither slowing down or speeding up as the workload increases or decreases. *Limited capacity*, on the other hand, implies that as the number of items to process goes up, the overall rate at which each item is processed will go down. Figure 11.3 provides schematics of different type of capacities in a parallel system.

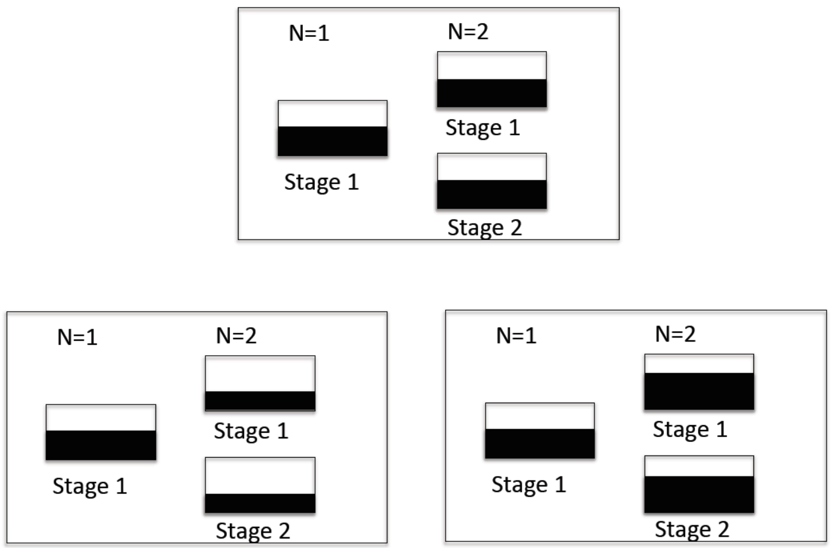

Fig. 11.3. Schematics of different types of capacity in a parallel system.

To return to the history, we can see that Sternberg's (1966) assumption about the performance of a parallel model implies an unlimited capacity system. A limited capacity parallel system can predict linear increasing reaction times as a function of item number just as does a serial system. Egeth (1966) (see also Townsend & Ashby, 1983; Snodgrass & Townsend, 1980) pointed out that even unlimited capacity parallel processors can yield increasing, but not necessarily linear, mean reaction time functions in a natural way. A number of authors also started using limited capacity parallel processes that could mimic the straight line predictions of standard serial models. Atkinson, Holmgren, and Juola (1969) offered a simple stochastic model which mimicked ordinary serial processing predictions and Townsend (1971, 1972) showed that each type of model could mimic the set size function of the other (see also Murdock, 1971).

Due to these (and other) studies, the phenomenon of linear increasing reaction time curves was no longer considered a fundamental parallel/serial distinction, because it simply indicates first and foremost a limitation in capacity. At the same time, people started paying more attention to analyzing entire distribution functions of reaction time, rather than just the mean. In the next section we will discuss more on serial/parallel mimicking based on the analysis of reaction time distribution functions (Townsend & Ashby, 1983).

Substantial advances in modeling of diverse phenomena using broad classes of serial models have appeared in the work of Treisman and colleagues (e.g., Treisman & Gormican, 1988) and Wolfe and colleagues (e.g., Wolfe, 1994). However, the serial models have not been tested against parallel alternatives.

In the third section we will talk about the empirical side of the parallel/serial issue. There are now several experimental strategies based on mathematical proofs available to help distinguish whether processing is serial or parallel. Most of these methods are based on reaction time but some are based on accuracy.

11.2. Theoretical Treatment

11.2.1. *Stochastic models*

Townsend and others (Townsend, 1972, 1976; Townsend & Ashby, 1983; Townsend & Schweickert, 1989; Schweickert, 1978, 1982, 1989) have developed precise definitions of stochastic parallel and serial processes at the distributional level. For the remainder of this chapter, discussions on natural properties of parallel/serial systems are based on those rigorous definitions.

Many of the results in this chapter will be derived fully only for the case when there are two elements to be processed, but in most cases they could be generalized to any number of elements.

To provide a fine grained analysis of serial and parallel processing, we first introduce some temporal concepts (illustrated in Fig. 11.4). Suppose a and b are the two positions of the elements to be processed. Let letter t denote the *intercompletion time* (ICT), which is defined as the interval between successive completions. As can be seen from the figure, the ICT is also the *actual processing time* (duration spent by the system on an element) for elements in a serial system. However, the ICT will not be an actual processing time in parallel processing except for the first element completed. We then denote τ as the *total completion time* on a particular element, which refers to the duration from $t = 0$ until the element is completed. Thus, as is apparent in Figure 5, the total completion time of an element in a parallel system is equal to its actual processing time. Also, it should be noticed that the total completion time in either model is always the sum of the number of ICTs preceding the completion of the considered

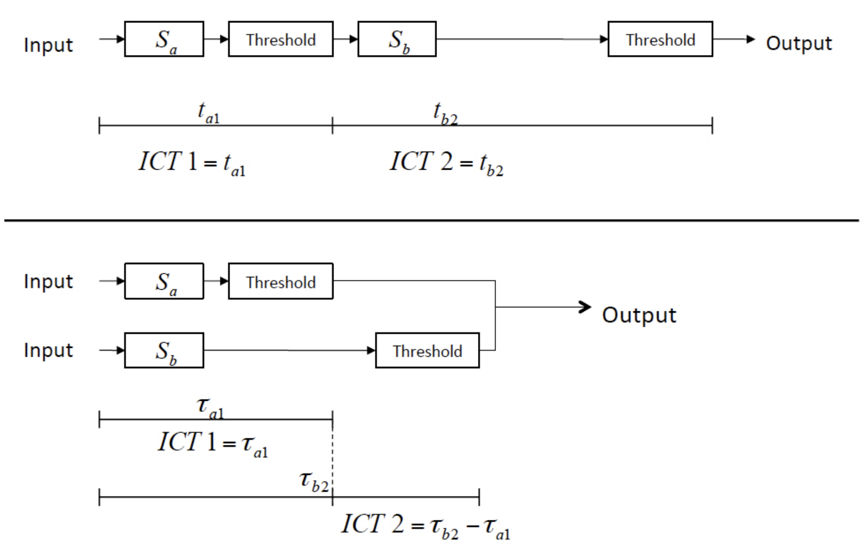

Fig. 11.4. Illustrations of intercompletion times, actual processing times and total completion times.

element. Finally, the concept of *stage* is needed to decompose the model of processing. By stage we mean the interval between the instants of two successive element completions. Thus "stage k" refers to the interval occupied by the kth intercompletion time. It is easily seen that in a serial model, the length of a stage is equal to the time required for an element to be processed (i.e., the actual completion time), while in a parallel model, the actual completion time of a single element will be equal to the sum of the stage durations up until the element is complete.

Now we are ready to introduce the serial and parallel definitions. Let $f_{ai}(t_{ai})$ be defined as the density function describing the ith intercompletion time when the element in position a is completed ith and processing is serial. Analogous expressions hold for position b. To emphasize the potential difference between the probability densities associate with serial and parallel models, we similarly define $g_{ai}(t_{ai})$ for $i = 1$ or 2 as the density function describing the ith intercompletion time when a is completed ith and processing is parallel. Let $G_{ai}(t_{ai})$ be the parallel distribution function and $G_{ai}(t_{ai}) = 1 - G_{ai}(t_{ai})$ the survivor function for i= 1 or 2 on position a. We now state our definitions.

Definition 11.1. A stochastic model of a system for processing positions

a and b is *serial* if and only if

$$f_{a1,b2}(t_{a1}, t_{b2}; \langle a, b\rangle) = pf_{a1}(t_{a1})f_{b2}(t_{b2}|t_{a1})$$

and

$$f_{b1,a2}(t_{b1}, t_{a2}; \langle b, a\rangle) = (1-p)f_{b1}(t_{b1})f_{a2}(t_{a2}|t_{b1}).$$

The quantity p is the probability that a is processed first. $f_{a1,b2}(t_{a1}, t_{b2}; \langle a, b\rangle)$ is an expression of the probability density of the joint occurrence of the completion order $\langle a, b\rangle$, that a consumes t_{a1} time units of processing and that b completes processing t_{b2} time units after a.

This definition is an intuitive description of the behavior of serial models within each stage. For example, the first equation states that with probability p, position a is processed first, in which case the first-stage density is $f_{a1}(t_{ai})$. Now the second-stage density must allow for the length of time taken for a to affect the processing time of b, thus the second-stage density is the conditional density function $f_{b2}(t_{b2}|t_{a1})$.

We can also define the parallel model in the similar way.

Definition 11.2. A stochastic model of a system for processing positions a and b is *within-stage independent and parallel* if and only if

$$g_{a1,b2}(t_{a1}, t_{b2}; \langle a, b\rangle) = g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})g_{b2}(t_{b2}|t_{a1})$$

and

$$g_{b1,a2}(t_{b1}, t_{a2}; \langle b, a\rangle) = g_{b1}(t_{b1})\bar{G}_{a1}(t_{b1})g_{a2}(t_{a2}|t_{b1}).$$

Here $g_{a1,b2}(t_{a1}, t_{b2}; \langle a, b\rangle)$ is completely analogous to $f_{a1,b2}(t_{a1}, t_{b2}; \langle a, b\rangle)$ of the serial model. $\bar{G}_{b1}(t_{a1})$ is the survivor function of t_{b1}, that is, $\bar{G}_{b1}(t_{a1}) = p\{T_{b1} > t_{a1}\}$. This survivor function along with $g_{a1}(t_{a1})$ describes the first stage of processing when the completion order is $\langle a, b\rangle$. The product gives the probability density that a is completed first at time t_{a1} and that b is not yet completed by this time. This component of parallel models represents the fact that a and b begin processing simultaneously. That it can be written as a product of functions of the individual elements is a result of our assumption of within-stage independence.

Thus looking at the two definitions, it is easy to notice that the components in the serial and parallel model that describe the second stage of processing are structurally identical. This indicates the fact that on the second stage, whether processing is parallel or serial, only one element remains to be completed, and with only one element left to be processed, serial and parallel models must be equivalent. So what makes the serial model differ from the parallel model? A fundamental difference is in determining processing order. If the system is parallel, from the definition we see that the determination of processing order inherently depends on the rates with which a and b are processed. In contrast, if the system operates serially, then the decision as to whether the processing order will be $\langle a, b\rangle$ or $\langle b, a\rangle$ is made a priori. This difference in how the systems select processing order is a fundamental difference between parallel and serial processing models.

An important aspect in need of clarification is the assumption of within-stage independence. This postulate states that during any single stage (i.e., the time between the completion of two successive elements), the processing of all unfinished elements is independent. This assumption does rule out some interesting parallel models, but it still allows many different kinds of dependency to occur. For example, parallel capacity reallocation models cause a dependency to occur between stages and even on the level of overall completion times, but within-stage independence is still possible.

11.2.2. *Serial-parallel equivalence*

Now that we have explicitly stated the form of serial and parallel models, we are in a position to ask the question of when there exists a parameter mapping such that the two models are equivalent. As suggested by Townsend and Ashby (1976), given any within-stage independent parallel model, we can always construct a serial model that is completely equivalent to it by setting

$$p = \int_0^\infty g_{a1}(t)\overline{G}_{b1}(t)dt,$$

$$f_{a1}(t_{a1}) = \frac{1}{p} g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1}),$$

$$f_{b1}(t_{b1}) = \frac{1}{1-p} g_{b1}(t_{b1})\bar{G}_{a1}(t_{b1}),$$

$$f_{b2}(t_{b2}|t_{a1}) = g_{b2}(t_{b2}|t_{a1}),$$

$$f_{a2}(t_{a2}|t_{b1}) = g_{a2}(t_{a2}|t_{b1}).$$

Inversely, given a serial model, *if* there exists a within-stage independent parallel model that is completely equivalent to it, it can be found by setting

$$\bar{G}_{a1}(t) = exp[-\int_0^t \frac{pf_{a1}(t')}{p\bar{F}_{a1}(t') + (1-p)\bar{F}_{b1}(t')} dt'],$$

$$\bar{G}_{b1}(t) = exp[-\int_0^t \frac{(1-p)f_{b1}(t')}{p\bar{F}_{a1}(t') + (1-p)\bar{F}_{b1}(t')} dt'],$$

$$g_{b2}(t_{b2}|t_{a1}) = f_{b2}(t_{b2}|t_{a1}),$$

$$g_{a2}(t_{a2}|t_{b1}) = f_{a2}(t_{a2}|t_{b1}).$$

It turns out that either of the parallel survivor functions for stage 1 (indexed by the capital "G's") may not be true survivor functions in that one or more may not approach 0 as t becomes large, as it must for item completion to occur.

Hence, these results prove that serial processes are more general than within-stage parallel processes, at least in the sense that the class of parallel models defined above is contained within the class of serial models. This should not be altogether unexpected, since from the definitions of the models we can see that serial models have one more parameter than parallel models for manipulating processing order. This result was employed in an experimental test of serial vs. within-stage independent parallel memory retrieval (Ross & Anderson, 1981).

Thus the previous discussion has illustrated that many serial and parallel models can mimic each other in the sense of distributional equivalence, which demonstrated the complexity of this serial/parallel issue for any empirical reaction time data set, we can never easily draw the conclusion that the processing is serial or parallel, because either model may give rise to the same RT distribution. However, it is critical to understand that when we say serial models are more general than parallel models, it does not mean that a parallel machine of this type works in real time like a serial machine, only that the mathematical description of the parallel class of machine is contained within the mathematical description of the serial class for a particular paradigm. Armed with the foregoing mathematics and intuition, we now exhibit an even more general mathematical equivalence function between the classes of parallel and serial models.

11.2.3. *Further results on serial-parallel equivalence*

Notice that up to now, with-in stage independence plays an important role in the analysis of serial-parallel equivalence problem. To loosen this condition, we need to consider the above stochastic models in a more general way. As suggested in Townsend and Ashby (1983), following up on Vorberg (1977), if we treat the problem from the point of view of foundational probability theory, we find that the two general classes of serial and parallel model are equivalent to each other.

Formally, our goal is to establish that for any model $m_s \in \mathrm{M}_s$ (the class of serial models), there exists a model $m_p \in \mathrm{M}_p$ (the class of parallel models) such that m_p gives the same probability measure on all the corresponding possible events. Then the class M_s is said to imply the class M_p. If the implication goes both ways, then we can say that these two classes are equivalent to each other. To achieve this goal, we need to employ the concepts of the *Borel field* and *Borel functions* from *measure theory*. We note that the real line R with its usual topology is a locally compact *Hausdorff space*, hence we can define a Borel field on it. To be specific, regarding our present case, the sets in Borel field B will associate with intervals of time in which elements might complete processing, and the Borel function J establishes the relation between the two Borel fields, thus associating the probability spaces. We may now state the formal definitions of parallel and serial models in terms of measure spaces.

Definition 11.3. A model of a parallel system is defined by $m_p = \langle \mathrm{W}_p, \mathrm{B}_p, \mathrm{P}_p \rangle$, where B_p is the Borel field associated with W_p, our space, and P_p is some particular probability measure on B_p. For every point w belonging to W_p, the probability density function assigned by m_p is given by

$$g_p(w) = g(\tau_1, ..., \tau_n; a_1, ..., a_n)$$

with

$$w \in \mathrm{W}_p = \{(\tau_1, ..., \tau_n; a_1, ..., a_n) | \tau_i \in R^+,$$

$$1 \leq i \leq n; \tau_1 \leq \tau_2 \leq ... \leq \tau_n, (a_1, ..., a_n) \in permutation(n)\}$$

where τ_i is the total completion time (also the actual processing time in parallel case) at which the ith element to be finished is completed, and a_i is the serial position of the element completed ith.

Definition 11.4. A model of a serial system is defined by $m_s = \langle \mathrm{W}_s, \mathrm{B}_s, \mathrm{P}_s \rangle$, where B_s is the Borel field on W_s. Thus for every point w belonging to W_s, the probability density function assigned by M_s is

$$f_s(w) = p(a_1, ...a_n)f(t_1, ...t_n|a_1, ...a_n)$$

with

$$w \in \mathrm{W}_s = \{(t_1, ..., t_n; a_1, ..., a_n)|t_i \in R^+,$$

$$1 \leq i \leq n; (a_1, ..., a_n) \in permutation(n)\}$$

where t_i is the intercompletion time (also the actual processing time in serial case). $p(a_1, ...a_n)$ gives the probability of the order shown.

As indicated in Townsend and Ashby (1983), it is always possible to provide a mapping between the parallel and serial spaces that is one-to-one and onto (i.e., the Borel function J), to show that for any model of one type, there exists a model of the other type which is equivalent. Thus, from the point of view of pure mathematics, serial models and parallel models could be actually totally equivalent at the distributional level, once we loosen the condition of within-stage independence.

To be sure, these mappings obscure vital distinctions that may exist between the two types of processing. Another issue that needs to be emphasized is that even when two theories or models make predictions at some level that are equivalent, to the extent that the two theoretical structures are not truly identical, there must exist aspects of the models or theories that relate to different possibilities in the real world. Here, to get a deeper understanding of the mechanisms of serial and parallel processing that can give us guidance in terms of experimental strategies, we list several divergent properties of serial and parallel systems.

Fundamental Distinctions between Serial and Parallel Systems

(i) In the case of serial systems, at any point in time, at most one element can be in a state of partial completion, whereas parallel systems can have any number of partially completed elements.

(ii) In the case of serial systems, only one element can change its processing state at a certain point in time, whereas in parallel systems, any element can increase its processing status at any point in time.

(iii) In the case of serial systems, the time taken to process an element cannot depend on the identity of any element that is not completed until after the given element is finished, whereas in parallel systems such dependence can exist.

(iv) In the case of serial systems, it is possible to build a system so that a processing order of the elements is preselected and then to make the processing times depend on the particular preordained order on any given trial, whereas a parallel system cannot be built so that a processing order can be preselected.

11.2.4. *The principle of correspondent change*

We now introduce an important principle of mathematical models and theory testing as the summary of this section. The name we give to this principle is the *principle of correspondent change* (Townsend & Ashby, 1983, Chapter 15). This principle is rather obvious and employed by scientists on a day-to-day basis in an implicit way, but not always fully appreciated.

(i) For any given empirical milieu and for any given class of models, there will exist a set of subclasses where models are indistinguishable within their subclass and in that specific milieu.

(ii) Empirical changes in the environment or stimulating situation should be reflected in a nonvacuous theory or model by corresponding changes or invariances in the model or theory. Such changes or invariances should be predictable and in consonance in the correct model.

It is not surprising that two theories or models may make predictions at some level that are equivalent. However, the second point of this principle points out that we do not need to be too upset about this fact, because to the extent that the two theoretical structures are not truly identical, there must exist aspects of the models or theories that relate to different possibilities in the real world. Let us return to the context of serial and parallel model discrimination. If we examine the matter from a static view, it has been demonstrated that for any parallel model, there must exist a serial model which under certain conditions yields the same processing time distribution. However, if we examine the matter from a dynamic view, then due to the structural differences inherent in the models, there must exist one or more factors such that by controlling those factors, the outcome of the two different models will change in different ways. To make the discussion a little more concrete, suppose there are two positions a and b, and suppose we are trying to decide whether the processing behavior is serial or parallel. Each model will certainly possess structure, which will be in the form of functions of parameters. Let the experimenter manipulate the complexity of the element placed in b while leaving the complexity of the element in a the same. Then surely the structure associated with position a should be invariant whereas the structure connected with position b should vary in an appropriate fashion. We definitely allow that interactions or limitations in capacity affect the rate in position a, but again they should be reflected in regular changes which show different patterns in different models, which then allow us to distinguish between the two models.

The principle of correspondent change could be considered in a more abstract way and could be applied to general model discrimination problems rather than just serial and parallel issues. If we denote δ as the variable indicating the experiment conditions, and denote $F(\delta)$ and $G(\delta)$ as the observed outcomes of the experiment conditions under two different models, then it is highly possible that model equivalence occurs in some sense. This would mean functions F and G have the same range, that is, for any δ' , there exists a δ'' to make $G(\delta'') = F(\delta')$. However, the signature of F and G must not be the same since the structures of models are essentially different. Thus, if we are lucky enough, by moving the value of δ, we can tell the difference of the two models since the direction of the changes of the outcome are different in the different models.

The principle of correspondent change establishes the connection between a static view and a more dynamic view, and provides guidance for experimental design. As stated before, the principle of correspondent change has been employed by scientists for many years, although in an implicit form much of the time. In the next section we will introduce some experimental paradigms aiming to distinguish serial and parallel processing. Note that, for most experimental paradigms, the principle of correspondent change is a necessary condition.

11.3. Experimental Paradigms

In this section, we discuss several experimental paradigms which aim to distinguish serial versus parallel processing.

11.3.1. *Tests using capacity*

In general, as discussed extensively above, limited capacity parallel models can mathematically and intuitively perfectly mimic serial models, so we are generally averse to architectural tests based on that notion. However, it has been long known, that unlimited or super capacity parallel models can make predictions that reasonable serial models simply cannot duplicate (e.g., Townsend, 1972, 1974).

Lately, methodologies associated with capacity distinctions have been proposed that might be able to distinguish serial models from parallel processes (Townsend & Nozawa, 1995; Townsend & Wenger, 2004). Thornton and Gilden (2007) is such an example. Their technology includes a multiple target search (MTS) method and utilizes computational models of serial and parallel processing. Although they do not deal directly with our capacity measure or explicitly consider it, both the information of reaction time and accuracy are used in their method, allowing them to consider both the capacity qua efficiency, issue as well as speed-accuracy trade-offs.

They employed a random walk model to simulate the multiple-target visual search. To address the capacity issue in the parallel model, they proposed employing a limitation parameter τ that attenuates the drift rate of each random walk as a function of set size. With regard to the serial model, they introduced two parameters to relax decision criteria as a function of both set size and accumulation time, which allows the serial model to incorporate a rational decision strategy consistent with subjects psychological motivations.

In this way, whether serial or parallel processing was performed in a task could be assessed by comparing goodness of fit metrics for the two models. The authors provide 29 sets of data that cover different difficulty levels of visual search. The results of applying their algorithm to each of the 29 tasks demonstrated that the majority of tasks were performed in parallel, while some difficult searches (e.g., rotation and mirror inversion tasks) were more likely to use serial processing.

Little research has been done in using random walk simulations to model either interactions between channels or the capacity limitation issue in parallel processing. However, Thornton and Gilden (2007) do provide a method of model selection in which at least the capacity issue is considered.

In order to account for the influence of the other three previously mentioned processing characteristics on architecture, Wenger and Townsend (2006) undertook perhaps the first study to simultaneously consider all four properties in a factorial design without having to resort to averaged data. In order to compare serial and parallel models in their task, they employed linear dynamic systems (see Townsend & Wenger, 2004). Unlike Thornton and Gilden (2007), however, this approach did not rely on quantitative model fits of parameterized models, but rather looked for distinguishing qualitative behaviors.

In this study, both meaningful (face-like) and meaningless (scrambled face) stimuli were tested, and in both cases the number of targets and distractors were manipulated. Rather than finding what had been previously assumed in these cases, that the former would be processed in parallel and the latter in serial fashion, it was demonstrated that parallel processing was used in both cases. The primary difference between the two was found via the capacity measure, with much higher capacity for meaningful stimuli.

These findings argue against the theory that there exist separate, distinct processing systems for gestalt-like stimuli, and instead that processing may differ only in the degree of efficiency. Other interesting findings concerning stopping rule and independence are put forth, but are less germane to the current discussion.

11.3.2. *The parallel-serial tester*

PST can be viewed as a descendant of Snodgrass (1972) pattern matching paradigm, combined with mathematical proof (Townsend, 1976) to distinguish parallel vs. serial models. One of the nice properties is that the tester functions at the level of the mean reaction time, which obviates dilemmas with the exact form of the distribution. On the other hand, when it is desirable to employ specific classes of distributions, the sensitivity of the test may be enhanced.

The PST paradigm requires an observer to search through a list of two items for one or more targets. The paradigm consists of the three experimental conditions. Condition I is composed of two types of trials, and requires response R1 if the target is on the left, and response R2 if the target is on the right. Condition II and III are composed of four types of trials each. Condition II requires that both comparison items match the target in order for response R1, whereas in condition III only one comparison is required to match the target in order for R1 to be the correct response. This is outlined in Table 1.1.

The underlying assumptions of PST are: the processing is self - terminating, the distribution on processing time is somehow different when two matching items are being compared than it is when two mismatching items are compared, accuracy is high enough and does not covary in an important way with reaction time. PST is a distribution-free theory. There are many examples of experimental applications in the work of Townsend and Snodgrass (Townsend & Snodgrass, 1974; Snodgrass & Townsend, 1980; Snodgrass, 1972). PST is founded on the principle that all (and any) items being processing in a parallel system can be in partial

Table 11.1. Parallel-serial Testing paradigm

Target = A		
Condition	Comparison items	Response
2*C I	AB	R2
	BA	R1
4*C II	AA	R1
	AB	R2
	BA	R2
	BB	R2
4*C III	AA	R1
	AB	R1
	BA	R1
	BB	R2

states of completion at any point in time, whereas only one item can be in such a state in a serial system.

The results from Snodgrass and Townsend (1980) suggest that PST results depend on the complexity of the matching required of the subject. With more complex patterns and processing requirements, subjects appear to be forced to resort to serial processing, whereas in the simpler versions evidence showed that subjects can operate in parallel.

11.3.3. *Tests by time delimitation*

Eriksen and Spencer (1969), Shiffrin and Garner (1972), and Travers (1973) compared the effect of presenting one symbol or part of the display for t msec followed by another symbol for t msec and so on until the entire k symbols had been shown after $k * t$ msec, with the effect of displaying the entire array simultaneously for t msec. They found that accuracy in the latter condition was roughly equal to that in the first condition, which supported parallel processing as opposed to serial processing.

Townsend and his colleagues (Townsend, 1981; Townsend & Ashby, 1983; Townsend & Fial, 1968) developed a counterpart to that paradigm, in which the first condition is identical to the above. In the latter condition, k successive time intervals, each of duration t/k are permitted for processing each item. Hence if processing is serial, performance should be about equal in the two cases, whereas if processing was parallel, the sequential type of presentation should seriously degrade performance. The experimental results were in favor of parallel processing (Townsend, 1981).

Although a potentially useful tool in studying mental architecture such as serial vs. parallel processing, it appears that some degree of model mimicking still might occur (see, e.g., the discussion in Townsend, 1981;

and Townsend & Ashby, 1983).

11.3.4. *The second response paradigm*

From the previous section and the principle of correspondence change, we see that parallel models can predict that if processing is stopped at an arbitrary point in time, any number of items may be in a state of partial processing. For instance, if the cognitive system is processing features in parallel on several items, then cessation of processing can leave each item with some features completed. In contrast, in serial processing, one item is completed at a time, with the succeeding item not being started until the last is finished (Townsend, 1974). Therefore, if processing is sharply terminated, at most one item should be in a state of partial processing. This will not ordinarily show up in the overall accuracy results of a typical experiment.

In order to exploit this distinction, Townsend and Evans (1983) developed a technique based on a second response on each item to be processed, which was later extended by Zandt (1988). It was demonstrated that the pattern of accuracy on the second responses differed for serial and parallel models. Null hypotheses for serial processing within several levels of constraints on responding in the serial models were introduced and the results applied to a pilot experiment. Within the study, the data passed the tests for the most lenient serial hypothesis but ran into trouble with the more restrictive criteria. A potential vulnerability of this strategy is that in some applications, the second response might be based more on the first response than on the cognitive or perceptual processing associated with the first response.

The next section reviews the theoretical and methodological legacy of the additive factors method (Sternberg, 1969) mentioned earlier. It forms by far the most broadly and deeply theoretically developed theory-driven methodology for serial-parallel assessment. It is also the most widely employed and its popularity seems to be growing.

11.3.5. *The method of factorial interactions with selective influence of cognitive subprocesses*

The concept of *selective influence* is an important notion in perceptual and cognitive processes, which is also one of the most widely used assumptions in approaches to the identification of processing architecture. This notion was first employed by Sternberg (1969) in his additive factors method, in which he stated that there exist manipulations that can separately influence the processing associated with each factor, say X and Y. Thus in a 2×2 factorial design, we can write the Mean Interaction Contrast (MIC) as

$$\begin{aligned} MIC = &[RT(X + \Delta X, Y + \Delta Y) - RT(X + \Delta X, Y)] \\ &-[RT(X, Y + \Delta Y) - RT(X, Y)] \end{aligned}$$

Sternberg (1969) suggested that based on selective influence, serial models with independent processing times would exhibit $MIC = 0$. The theory underlying selective influence was subsequently endowed with a rigorous mathematical foundation (see, e.g., Schweickert, 1978; Schweickert & Townsend, 1989; Townsend, 1984; Townsend & Ashby, 1983; Townsend & Nozawa, 1988; Townsend & Schweickert, 1985, 1989; Townsend & Thomas, 1994) also contributed to this question, and found that the sign of the MIC could be an indicator of not only processing architecture but also stopping rules. The results could be summarized as follow:

If processing is serial, then regardless of whether a self-terminating or exhaustive stopping rule is used, the MIC always equals zero. If processing is parallel, then the MIC value will depend on the stopping rule. That is, parallel exhaustive processing will lead to an underadditivity of the mean RT, thus a negative MIC, whereas parallel self-terminating processing will lead to an overadditivity of the mean RT, thus a positive MIC. Figure 11.5 illustrates the MIC prediction for parallel and serial processing under different types of stopping rules.

Thus we can see that the MIC provides us with important clues to the mental architecture and stopping rules. However, it still has limitations. For example, the MIC cannot distinguish between serial, self-terminating processing and serial, exhaustive processing, and also cannot distinguish between parallel and coactive processing (Miller, 1982; Townsend & Nozawa, 1995). Due to these limitations, it was determined that ordering the means through selective influence is not sufficient to make discriminative predictions for varying architectures and stopping rules. It is necessary that a selective influence variable exert influence on distributions at a higher level. (Check Townsend, 1990a; Townsend, 1990b for the stochastic dominance hierarchy.)

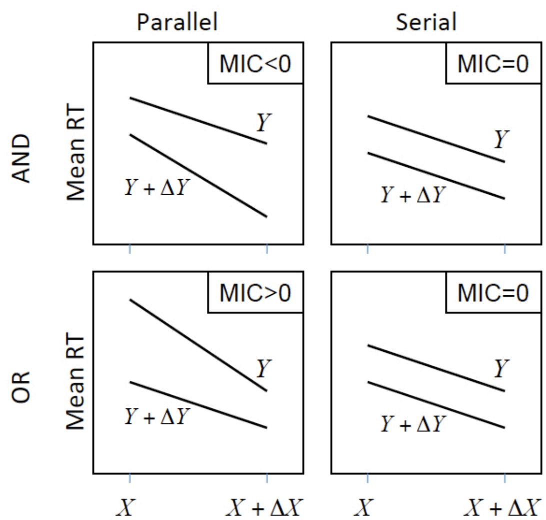

Fig. 11.5. Predictions for mean reaction times and mean interaction contrasts as a function of architecture and stopping rule.

11.3.6. *Systems factorial technology (SFT)*

Townsend and Nozawa (1995) developed the Systems Factorial Technology (SFT) which simultaneously tested parallel vs. serial processing for various stopping rules at the level of distribution functions as well as measuring capacity at a distributional level.

SFT includes an experimental paradigm (DFP) which requires the manipulation of two factors: The activation of sub-channel (target presence/absence) and the speed of processing (different salience levels). To use SFT to predict architecture and stopping rule, we assume that salience manipulation could produce selective influence at the distributional level. See Fig. 11.6 for double factorial design.

These data are analyzed using Survivor Interaction Contrast (SIC),

$$SIC(t) = [S_{ll}(t) - S_{lh}(t)] - [S_{hl}(t) - S_{hh}(t)],$$

where the subscript denotes the salience level of each process.

Figure 1.7 displays the predictions for the variety of models.

Observe that parallel-processing SICs 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 SICs: The contrast for OR parallel processing is con-

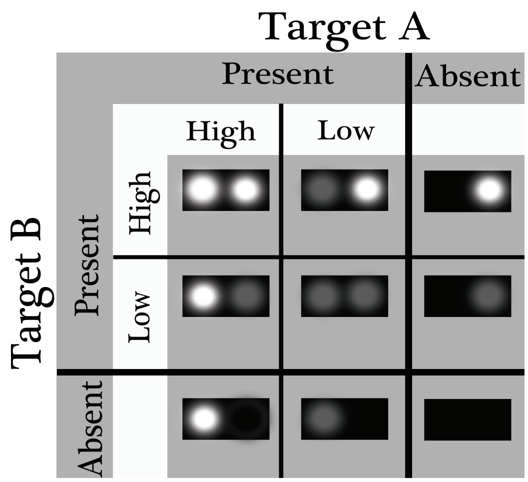

Fig. 11.6. Illustration of the Double Factorial Design.

sistently positive, whereas the contrast for the coactive model possesses a small negative "blip" at the earliest times, before going positive. Since the MIC must be positive in coactivation, it can be shown that the positive portion of the SIC always has to exceed the negative portion.

The advantage associated with use of both the SIC and the MIC goes beyond the ability to distinguish coactive from parallel processing. It is also intriguing that the OR and the AND serial stopping rules are now experimentally distinguishable, since in the OR case $SIC = 0$ always, but in the AND case there is a large negative portion of SIC, 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 distributional level. The general applicability of the distributional approach has benefited from theoretical extensions by Schweickert and colleagues (e.g., Schweickert, Giorgini, & Dzhafarov, 2000) to general feed-forward architectures, which contain parallel and serial subsystems, and from advances in methods of estimating entire RT distributions (see Zandt, 2000, 2002).

SFT can also help us predict workload capacity. Townsend and Nozawa (1995) proposed a measure called the capacity coefficient. For OR processes, the capacity coefficient is computed as the ratio between the integrated hazard function of the double target condition and the sum of the integrated hazard functions of the single target conditions, based on RT data.

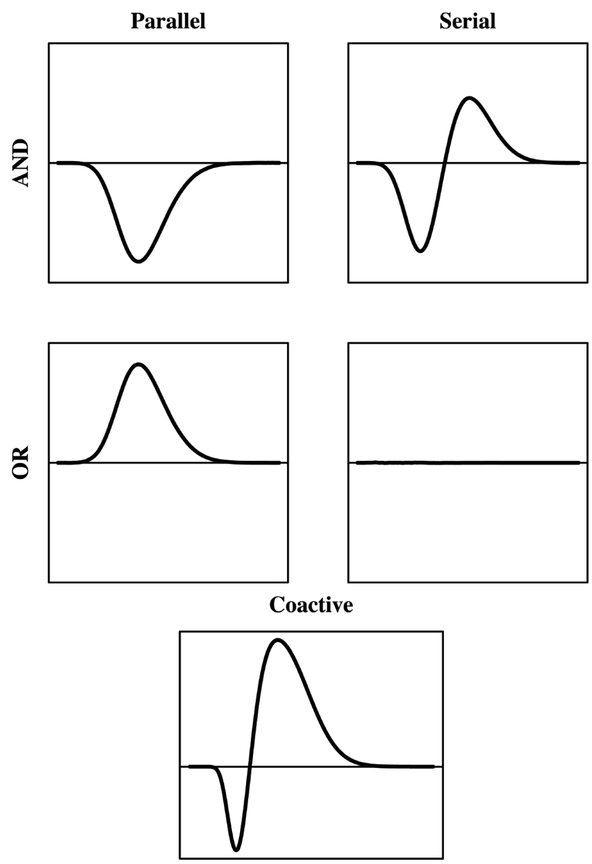

Fig. 11.7. SIC predictions for different information processing models.

$$C_{OR} = \frac{H_{AB}(t)}{H_A(t) + H_B(t)}$$

Here A and B signify the two processes. The hazard function is the probability density function over the survivor function, $h(t) = (f(t))/(S(t))$, $H(t)$ is the integral of the hazard function from zero to t. Townsend and Wenger (2004) developed a comparable capacity coefficient for the AND task.

$$C_{AND} = \frac{K_A(t) + K_B(t)}{K_{AB}(t)}$$

Analogous to the integrated hazard function, $K(t)$ is defined as the integral of $k(t)$ from zero to t, in which $k(t) = (f(t))/(F(t))$.

The interpretation of the two capacity coefficients for OR and AND conditions is the same: $C(t)$ should be equal to 1 in a system with *unlimited capacity*, i.e., the processing in a given channel is not affected by the increase of the workload. $C(t)$ above one indicates that processing efficiency of individual channels actually increases as we increase the workload, and $C(t)$ below 1 indicates that the processing efficiency of individual channels decreases with increased workload. We call the former case super capacity and the latter case *limited capacity.* The SIC function can distinguish independent models with different combinations of architecture and stopping rules. If, however, at least one model has high levels of interaction between channels, the SIC signatures of the different models can resemble each other (Eidels, Houpt, Altieri, Pei, & Townsend, 2010). In this case the capacity coefficient provides information about architecture and channel dependence indirectly, and can allow us to disambiguate the models. Thus, examining both the SIC and the capacity coefficient provides a decisive test for the architecture and possible dependencies between the processing channels.

Beyond its original use in a simple detection task (Townsend & Nozawa, 1995), SFT has extended our understanding of performance in a wide variety of cognitive tasks (e.g., Sung, 2008). Successful applications include classic phenomena such as the Stroop task (Eidels, Townsend, & Algom, 2010), Navon arrows (Blaha, Johnson, & Townsend, 2007), and various Gestalt principles (Eidels, Townsend, & Pomerantz, 2008). Another area in which the SIC has been particularly informative is in search tasks, both visual search (Fific, Townsend, & Eidels, 2008) and short term memory search (Townsend & Fific, 2004). The SIC has also been successful with more complicated stimuli, and faces in particular (Fific & Townsend, 2010; Wenger & Townsend, 2001). There have even been successful applications of the SIC in the clinical domain (Johnson, Blaha, Houpt, & Townsend, 2010; see also Townsend, Fific, & Neufeld, 2007).

Recent theoretical work has opened up the possibility of using the SIC in a wider range of tasks. Eidels et al. (2010) have extended the predictions of the SIC to a wide class of parallel models with processing dependencies. On another front, Houpt and Townsend (2010) have developed a tool for making statistical inferences with the SIC. Fific, Nosofsky, and Townsend (2008) have applied SFT to the domain of multidimensional classification. In their experiment they contrasted SIC signatures for stimuli that varied along highly separable dimensions to the case where they varied along integral dimensions. In the separable case, some subjects were found to be using serial processing with an exhaustive stopping rule, while other subjects were clearly parallel, but also used an exhaustive stopping rule. In the experiment with integral dimensioned stimuli, however, they found clear evidence that subjects were using a coactive architecture, rather than serial or parallel. They assert that this strong distinction in architecture could be used as another method for distinguishing between separable and integral stimulus dimensions (Garner, 1974).

11.4. Discussion

Psychology, and particularly cognitive psychology, aim to uncover the mysteries of the human mind. To do so as a science, it must perforce evolve rigorous theories and methods which are quite analogous, but not homologous, to the systems identification theories found primarily in electrical engineering and computer science curricula. The mind, apparently being run by a machine with 10^{14} or so primary parts (the neurons) and with a miasma of little understood graded chemical and potential fields thrown into the bargain, is an astronomically more complicated system than any discrete or continuous system open to such strategies. Hence, the question must arise not as to whether it is possible to build a computer program, not to mention verbal models–that imitate certain aspects of cognition–such tasks are common homework problems in cognitive psychology or cognitive science–but whether said models or theories that are *unique* are achievable. Would not the assiduous theorist wish, in her fondest dreams, to assert that such a model is both *necessary* and *sufficient* to explain a certain phenomenon?

The parallel vs. serial processing issue stands as a kind of icon of cognitive psychology from this point of view. Why? Because one could hardly conceive of an issue of mental function where the opposing concepts are so antipodal as these two modes of processing. Not only of entirely opposite character, but so doggone elementary and straightforward almost to the point of seeming insipidity. And yet, not: 1. Parallel and serial processes form the basic components of well-defined systems of almost arbitrary complexity (a special interesting case is found in systems based on connected graphs; see Schweickert, 1978; Schweickert & Townsend, 1989). 2. Perhaps somewhat astoundingly, given the expected complexity of mental activity, either one or the other type of process has actually been identified to a very high probability employing the approach of Systems Factorial Technology.

A few examples were cited earlier.

So, one might have forecast a rather early and decisive method of identification of serial vs. parallel processing. Alas, that has been far from the case as outlined in this essay. First raised in an experimental fashion approximately 140 years ago, and resurrected in the 1960s, the issue has proven itself to be a rather sticky wicket to say the least. The intransigence of this 'simple' issue, primarily due to mutual model mimicking in a strong mathematical sense, should give pause to the most confirmed Pollyannish psychologist or cognitive scientist.

Yet there is clearly room for the industrious optimist, in the fact that a number of years of arduous mathematical effort combined with appraisal of theory-driven methodologies in the laboratory crucible, have led to a number of highly promising strategies for assessing whether processing is serial, parallel, or possibly hybrid. The most general and, so far, explored of these, has been those approaches based on systems factorial principles. It is now possible to provide quite persuasive evidence of serial or parallel processing, which was not the case a few decades ago.

There may be an overall message for psychology and cognitive science here. Perhaps in the domain of 'taking the lid off the human black box', as opposed to other venues, and in paraphrasing Ronald Reagan: "There ain't no free lunch, my friend."

Acknowledgments

This work was supported by NIH-NIMH MH 057717-07 and AFOSR FA9550-07-1-0078.

References

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.

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

Blaha, L. M., Johnson, S. A., & Townsend, J. T. (2007). An information processing inverstigation of hierarchical form perception: Evidence for parallel processing. *Visual Cognition: Object Perception, Attention and Memory 2006 Conference Report, 15*, 73-77.

Donders, F. C. (1868). Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Uterchtsche Hoogeschool, 1868-1869. *Tweeds Reeks, 2*, 92–120.

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

Eidels, A., Houpt, J. W., Altieri, N., Pei, L., & Townsend, J. T. (2010). Nice guys finish fast, bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology, 55*, 176–190.

Eidels, A., Townsend, J. T., & Algom, D. (2010). Comparing perception of Stroop stimuli in focused versus divided attention paradigms: Evidence for dramatic processing differences. *Cognition, 114*, 129–150.

Eidels, A., Townsend, J. T., & Pomerantz, J. R. (2008). Where similarity beats redundancy: The importance of context, higher order similarity, and response assignment. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1441–1463.

Eriksen, C. W., & Spencer, T. R. (1969). Rate of information processing in visual perception: Some results and methodological considerations. *Journal of Experimental Psychology Monographs, 79*, 1–16.

Estes, W., & Taylor, H. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Science, 52*, 446–454.

Fific, M., Nosofsky, R. M., & Townsend, J. T. (2008). Information-processing architectures in multidimensional classification: A validation test of the systems factorial technology. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 356–375.

Fific, M., & Townsend, J. T. (2010). Information-processing alternatives to holistic perception: Identifying the mechanisms of secondary-level holism within a categorization paradigm. *Journal of Experimental Psychology: Learning, 36*, 1290–1313.

Fific, M., Townsend, J. T., & Eidels, A. (2008). Studying visual search using systems factorial methodology with target–distractor similarity as the factor. *Attention, Perception, & Psychophysics, 70*, 583–603.

Garnes, W. R. (1974). *The Processing of Information and Structure.* NY: John Wiley & Sons.

Hamilton, J. (1859). *Lectures on Metaphysics and Logic.* Edinburgh: Blackwood.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11–36.

Houpt, J. W., & Townsend, J. T. (2010). The statistical properties of the survivor interaction contrast. *Journal of Mathematical Psychology, 54*, 446–453.

Johnson, S. A., Blaha, L. M., Houpt, J. W., & Townsend, J. T. (2010). Systems factorial technology provides new insights on global–local information processing in autism spectrum disorders. *Journal of Mathematical Psychology, 54*, 53–72.

Krueger, L. E. (1978). A theory of perceptual matching. *Psychological Review, 85*, 278–304.

Laming, D. R. J. (1968). *Information Theory of Choice-Reaction Times.* London: Academic Press.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Attention, Perception, & Psychophysics, 39*, 155–158.

Miller, J. O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*, 247–279.

Murdock, B. B. (1971). A parallel-processing model for scanning. *Attention, Perception, & Psychophysics, 10*, 289–291.

Neisser, U. (1967). *Cognitive Psychology.* NY: Appleton-Century-Crofts.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human Information Processing.* Hillsdale, NJ: Erlbaum Press.

Pieters, J. P. (1983). Sternberg's additive factor method and underlying psychological processes: Some theoretical considerations. *Psychological Bulletin, 93*, 411–426.

Posner, M. I. (1978). *Chronometric Explorations of Mind.* Hillsdale, NJ: Erlbaum 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.

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. (1985). Separable effects of factors on speed and accuracy: Memory scanning, lexical decision, and choice tasks. *Psychological Bul-*

letin, 97, 530–546.

Schweickert, R. (1989). Separable effects on factors on activation functions in discrete and continuous models. *Psychological Bulletin, 106*, 318–328.

Schweickert, R., Giorgini, M., & Dzhafarov, E. N. (2000). Selective influence and response time cumulative distribution functions in serial-parallel task networks. *Journal of Mathematical Psychology, 44*, 504–535.

Schweickert, R., & Townsend, J. T. (1989). A trichotomy: Interactions of factors prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *Journal of Mathematical Psychology, 33*, 328–347.

Schweickert, R., & Wang, Z. (1993). Effects on response time of factors selectively influencing processing in acyclic task networks with OR gates. *British Journal of Mathematical and Statistical Psychology, 46*, 1–30.

Shiffin, R. M., & Garner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology, 93*, 72–82.

Smith, J. E. K. (1968). Models of confusion. Paper presented to the Psychonomic Society, St. Louis, MO.

Snodgrass, J. G. (1972). Reaction times for comparisons of successively presented visual patterns: Evidence for serial self-terminating search. *Perception and 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 and Performance, 6*, 330–354.

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

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 276–315.

Sung, K. (2008). Serial and parallel attentive visual searches: Evidence from cumulative distribution functions of response times. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1372–1388.

Taylor, D. A. (1976). Stage analysis of reaction time. *Psychological Bulletin, 83*, 161–191.

Thorton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review, 114*, 71–103.

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

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

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. Kantowitz (Ed.), *Human Information Processing*. Hillsdale, NJ: Erlbaum Press.

Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology, 14*, 219–238.

Townsend, J. T. (1981). Some characteristics of visual whole report behavior. *Acta Psychologica, 47*, 149–173.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology, 28*, 363–400.

Townsend, J. T. (1990a). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin, 108*, 551–567.

Townsend, J. T. (1990b). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1*, 46–54.

Townsend, J. T., & Ashby, F. G. (1976). Toward a theory of letter recognition: Testing contemporary mathematical models. Paper presented to Midwestern Psychological Association, Chicago, IL.

Townsend, J. T., & Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes.* Cambridge, MA: Cambridge University Press.

Townsend, J. T., & Evans, R. (1983). A systems approach to parallel-serial testability and visual processing. *Advances in Psychology, 11*, 166–191.

Townsend, J. T., & Fial, R. (1968). Spatiotemporal characteristics of multisymbol perception. Paper presented to the First Annual Mathematical Psychology Meetings, Stanford, CA.

Townsend, J. T., & Fific, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Attention, Perception, & Psychophysics, 66*, 953–962.

Townsend, J. T., Fific, M., & Neufeld, R. W. J. (2007). Assessment of mental architecture in clinical/cognitive research. In T. A. Treat, R. R. Bootzin, and T. B. Baker (Eds.), *Psychological Clinical Science: Papers in Honor of Richard M. Mcfall.* Mahwah, NJ: Lawrence Erlbaum Associates.

Townsend, J. T., & Nozawa, G. (1988). Strong evidence for parallel processing with simple dot stimuli. Paper presented to the Twenty-Ninth Annual Meeting of Psychonomics Society, Chicago, IL.

Townsend, J. T., & Nozawa, G. (1995). 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. (1985). Interactive effects of factors prolonging processes in latent mental networks. In G. d'Ydewalle (Ed.), *Cognition, Information Processing, and Motivation: Proceedings of the XXIII International Congress of Psychology*, vol. 3 (pp. 255-276). Amsterdam: North Holland.

Townsend, J. T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology, 33*, 309–327.

Townsend, J. T., & Snodgrass, J. G. (1974). A serial vs. parallel testing paradigm when same and different comparison rates differ. Paper presented to the Psychonomic Society, Bosten, MA.

Townsend, J. T., & Thomas, R. D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *Journal of Mathematical Psychology, 38*, 1–34.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 111*, 1003–1035.

Travers, J. R. (1973). The effects of forced serial processing on identification of words and random letter strings. *Cognitive Psychology, 5*, 109–137.

Treisman, A., & Gormican, S. (1988). Feature analysis in early version: Evidence from search asymmetries. *Psychological Review, 95*, 15–48.

Van Zandt, T. (1988). Testing serial and parallel processing hypothesis in visual whole report experiments. Master's thesis, Purdue University, West Lafayette, IN.

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 and H. Pashler (Eds.), *Stevens' Handbook of Experimental Psychology, Vol. 4: Methodology in Experimental Psychology* (3rd ed.). NY: John Wiley & Sons.

Vorberg, D. (1977). On the equivalence of parallel and serial models of information processing. Paper presented to the Tenth Annual Mathematical Psychological Meeting, Los Angeles, CA.

Welford, A. T. (1980). *Reaction Times.* New York: Academic Press.

Wenger, M. J., & Townsend, J. T. (2001). Faces as gestalt stimuli: Process characteristics. In M. J. Wenger and J. T. Townsend (Eds.), *Computational, Geometric, and Process Perspectives of Facial Cognition* (pp. 229–284). Mahwah, NJ: Erlbaum.

Wenger, M. J., & Townsend, J. T. (2006). On the costs and benefits of faces and words: Process characteristics of feature search in highly meaningful stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 755-779.

Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research, 34*, 1187–1195.