9

Assessment of Mental Architecture in Clinical/Cognitive Research

James T. Townsend and Mario Fifić
Indiana University

Richard W. J. Neufeld
University of Western Ontario

The nature of mental structures and perceptual and cognitive processing modes has long been of concern in clinical psychology. However, in recent years, there has been a movement toward more rigorous descriptions and even predictions within a cognitive setting (e.g., Granholm, Asarnow, & Marder, 1996a, 1996b). In this chapter, we consider the strategic issue of mental architecture. *Mental architecture* refers to the organization of a set of mental processes. Two special cases of great importance are parallel processing, which means the simultaneous processing of items, and serial processing, which means the sequential and nonoverlapping processing of items. For instance, whether certain syndromes cause or are associated with a change from parallel to serial processing has often been a question of interest to clinical scientists (e.g., Knight, Manoach, Elliott, & Hershenson, 2000; Magaro, 1983). Mental architecture is one of a set of critical issues contained in our general theoretical approach (e.g., Townsend, 1974; Townsend & Ashby, 1983; Townsend & Wenger, 2004a). *Processing* refers to some perceptual or cognitive operation such as search, comparison, evaluation, or the like. More specific examples are given later.

As an apt example that relates to mental architecture, the dichotomy of automatic versus controlled processing is virtually omnipresent throughout cognitive science (e.g., see review by Shiffrin, 1988) and has been widely employed in clinical science (e.g., Carter, Robertson, Chaderjian, Celaya, & Nordahl, 1992; Hartlage, Alloy, Vázquez, & Dykman, 1993). Roughly, automatic processing is assumed to be effortless, requiring little or no attention, whereas controlled processing is assumed to be effortful with high attentional demands. In addition, all would agree that automatic processing is parallel, but whether controlled processing must be serial, rather than an inefficient form of parallel processing, would in recent times be more controversial. Now as Shiffrin (1988) intimates, it would be impossible to render the notion rigorous at the level of generality that is commensurate with its ubiquity. Nonetheless, in delimited settings it can often be interpreted in a rigorous and even mathematical fashion. In fact, we show in a later section that automaticity should be characterized, in each experimental milieu, in terms of the other critical issues as well as the parallel versus serial distinction. Although we treat the theme of mental architecture in a relatively general way, we subsequently briefly indicate the relationship of our developments to the notion of automaticity. We stress that the major dependent variable with which we labor in this exposition is that of response times.

The implementation of quantitative signatures of mental architecture in clinical studies is much in the spirit of *integrative psychological science*, a movement whose most forceful and articulate proponent has been Richard McFall. His own work, in collaboration with students and coworkers, has exemplified the merits of such synthesis (e.g., Treat, McFall, Viken, & Kruschke, 2001; Treat et al., 2002). Dick has indicted an excessive reliance in clinical cognitive science on assemblies of off-the-shelf measures, or tasks contrived essentially according to clinical hunch, in lieu of choice cognitive-science developments, especially formal versions. He has forcefully taken the discipline to task for what has often amounted to reinvention of the clinical-science wheel, at best, or discharging its scientific mandate with compromised measurement methods, at worst.

Throughout, Dick has put the welfare of the ultimate consumer of clinical science's offerings, specifically clients with problems in living, first and foremost. Practice based on the best psychological science has to offer, and impelled to prove itself in the court of outcome research (witness the prominence of evidence-based practice), has found its most formidable advocate in Dick McFall.

IMPORTANCE OF DISCERNING STATUS OF MENTAL ARCHITECTURE IN RELATION TO PSYCHOPATHOLOGY

Evaluative reviews of the literature on applications of cognitive psychology's information-processing models, most prominent in schizophrenia research, have concluded that these models provide a valid framework for interpreting performance deviations (e.g., Neufeld & Broga, 1981). It follows that the structure of processing systems deemed to bear on cognitive tasks is tantamount to a faculty that is spared with the advent of disorder. Such conclusions, however, have been based on verbal conjecture, rather than mathematically derived diagnostics of processing-system design. The importance of architectural aspects of processing in clinical cognitive science beckons the use of contemporary quantitative signatures, whose paradigms in principle can be appropriated in clinical studies (e.g., Neufeld & McCarty, 1994; Vollick, 1994).

Integrity of mental architecture is of obvious interest in its own right. It is important to know if psychopathology impinges on the usual operation of processing structure, including its apparent adaptation to selected variations in task composition (Townsend & Fific, 2004).

As intimated already (and further elaborated later in this chapter), cognitive architecture is but one component of the automatic-controlled processing construct, ubiquitously invoked in clinical studies. Any one or some combination of this construct's components may effect changes in observed performance. It therefore becomes important to ascertain cognitive architecture's contribution to performance deviations through methods isolating the design of the processing system. If evidently unaltered, for example, other sources can be scrutinized with greater confidence in assumed architectural intactness. Alternatives include overall processing capacity (Neufeld, Townsend, & Jetté, in press) and its parametric constituents (Neufeld, Carter, Vollick, Boksman, Levy, & Jetté, in press).

Evidence bearing on cognitive architecture furthermore is important to complementing analyses of cognitive performance, when a certain architecture (e.g., parallel, serial, or hybrid) is purported to prevail among normal and symptomatic participants alike (e.g., Carter & Neufeld, 1999). Moreover, Bayesian-based approaches to mediating group-level findings to individual participants have deemed selected structures of processing systems as common to the studied groups (Neufeld, Vollick, Carter, Boksman, & Jetté, 2002; Neufeld, 2005).

Finally, in research on functional neurocircuitry (e.g., via functional magnetic resonance imaging [fMRI]), it seems imperative to anchor imputed cognitive functions, notably the temporal arrangement of constituent operations, in mathematically illumined behavioral terms. Apart from being armed with freestanding cognitive-behavioral signatures of mental architecture, we become vulnerable to the circularity inherent in inferring the functions that are at work from the investigated neurocircuitry.

When it comes to treatment interventions aimed at improving information processing, rigorous profiles of clients' strengths and weaknesses of cognitive faculties seem indispensable. The efficiency of biological and psychological interventions in principle can be improved by targeting and monitoring such profiles' disorder-affected elements (e.g., Broga & Neufeld, 1981). Psychological interventions ideally can exploit spared elements, such as the parallel, serial, or other structural aspects of mentation (Townsend & Wenger, 2004a, 2004b). Moreover, proposed cognitive-science entrenched computational methods of assessing individuals' functioning over the course of treatment, and plotting treatment groups' trajectories of response to pharmacological agents, invoke specific parametric-model architectures that fall into the classes articulated here (Neufeld, in press-a).

The preceding are but a sampling of reasons that should motivate delving into quantitative developments for ascertaining cognitive architecture. The exposition that follows is devoted to the most prominent division, parallel versus serial transaction of task elements.

BASIC PROCESSING CHARACTERISTICS AND AVOIDING PITFALLS

Certain fundamental characteristics of human information processing have been known for some time (Townsend, 1974, 1990a; Townsend & Wenger, 2004a). These characteristics, although logically distinct, can interact in ways that can dupe or confound unwary researchers. We briefly outline the major concepts here and delve into more detail subsequently.

In addition to (a) serial versus parallel processing, it is necessary to consider (b) the decision or stopping rule, (c) the question of independence versus dependence of item or channel processing times, and (d) capacity, or how efficiently processes function, especially as the workload is increased.

The major reason for the potential of going astray mentioned in the first paragraph, is that different combinations of values of the listed characteristics can mimic one another. For instance, one of the first outcomes of mathematical research on parallel and serial processing was that limited-capacity parallel processing (i.e., each parallel channel is slowed as more items are being processed in other channels) could so perfectly mimic standard serial processing in the popular experimental designs that the two forms were mathematically identical and thus could not be distinguished in those designs (Townsend, 1969, 1971). Even today, one finds confounding between capacity (efficiency of processing; more on this soon) and architecture (e.g., parallel vs. serial processing; again, more detail later).

It is of the utmost importance to observe that the mathematical identity of certain parallel and serial models does *not* imply that the underlying physical mechanisms, whether neural, electronic, or mechanical, are equivalent! Rather, they simply look and act like one another (in fact, like identical twins raised in the same environment!) in certain experimental settings and under certain assumptions.

To avoid the sloughs of methodological despond threatening the psychological scientist, it is necessary to consider all of the characteristics together and in a rigorous framework. As clinical science moves inexorably into the realms of hard science, it would seem desirable that it not repeat the same mistakes that cognitive psychology has already encountered and begun to surmount.

We must also pay heed to the fact that because mental functions are probabilistic, not deterministic, even at the neural level, it is required that theory and theory-driven methodology be couched in stochastic language. Deterministic models can sometimes yield helpful intuition but they must be engaged with great caution, because sometimes their predictions are at odds with the true stochastic interpretations.

The next section builds up, or reminds the reader of, some needed quantitative tools. Certain of the material may seem overly simple to some readers, but we prefer to be as inclusive as possible. Any of it may be skipped at the reader's discretion.

Some Required Quantitative Tools

The first notion required is that of what the well-known probabilist Emanuel Parzen refers to as a *probability law* (Parzen, 1960). The *probability law* is a general term used to designate any of a number of formulas that define how the underlying random aspects will appear. Many investigators employ the alternate term *distribution* to refer to the probability law. We use *distribution* and *probability law* interchangeably. The term *distribution* can also be employed in a more detailed form, as shown later.

Perhaps the most common formula or designation of a probability law is the frequency function, that is, the idea from elementary statistics that counts up the times or relatively frequency that an event occurs (e.g., proportion of students testing at such-and-such an IQ, etc.). In the ideal or theoretical case, these frequency functions may be continuous, such as the normal curve or the exponential distribution. Probabilists and modelers call the ideal frequency functions *probability densities*, although the concept has nothing to do with the usual physical concept of density. We write a probability density or frequency function (hereafter density) f(t) where t is, of course, time because we are focusing on response times, and naturally t is greater than (or equal to) 0 and less than infinity.

Another useful designator of a probability law is the cumulative distribution function (note the special use of *distribution* here), which is the sum (in a discrete probability law) or integral (in a continuous probability law) from the lower limit (usually 0 in reaction time models) to an arbitrary value of the independent variable (usually time = t here). Thus, if we wish to know the probability that the response time was less than or equal to some t (rather than being exactly t), we calculate $F(t) = \int f(t')\,dt'$ integrating (summing in a continuous way) from 0 to the value of interest, t. In response-time research, the so-called survivor function (from actuarial theory) is of value, indicating the likelihood that the response time or processing time is not yet finished. It is $S(t) = 1 - F(t) = \int f(t')\,dt'$, this time integrating from t to infinity.

There are an infinite number of probability laws and therefore densities and quite a few useful ones, such as the normal and exponential. Everyone is familiar with the normal. A figure of the exponential distribution is shown in Figure 9.1. Let *exp* represent the exponential number 2.7182... (like π, exp goes on forever, without repeating). Then the formula for the exponential density is $f(t) = a\exp(-at)$, where, as usual, two symbols being

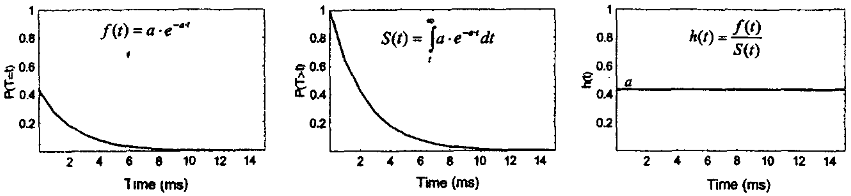

FIGURE 9.1 The exponential probability density function f(t) (left), its corresponding survivor function S(t) (middle), and its hazard function h(t) (right). Note that the hazard function of the exponential density has a constant value.

placed next to one another indicates multiplication. The variable a is the rate of processing in a model of response times. In the case of the exponential distribution, $S(t) = \exp(-at)$, and $F(t) = 1 - \exp(-at)$.

The mean of a distribution, also called the expectation, is just $E(T) = \int f(t')t'dt'$ this time integrating over all possible values from $t = 0$ to infinity. (Note that it is a convention to use capital T instead of lower case here, to indicate that it stands for any possible value, being what is known in the trade as a random variable.)

Finally, we need a finer grain statistic known as the hazard function. The concept, like that of survivor function, comes from actuarial statistics, where it gives the probability that, say, a person will die in the next short time, given that the person has survived until the present moment. It is written as the ratio of the density over the survivor function, which does, indeed, condition on the event not yet having occurred. In response-time theory, of course, it refers to, say, an item finishing in the next instant, given it is not yet completed.

Its formula is $h(t) = f(t)/S(t)$. The hazard function for the exponential distribution is the elementally and unique $h(t) = a$, a constant. The fact that it is a constant indicates that the instantaneous conditional rate of completion neither increases nor decreases. Figure 9.1 shows the various formulae associated with the exponential probability law. The mean of the exponential distribution is simply $1/a$. We next outline the basic processing characteristics involved in psychological systems.

Architecture: The Serial Versus Parallel Issue

Serial processing means processing things one at a time or sequentially, with no overlap among the successive processing times. *Processing* might mean search for a target among a set of distractors in memory or in a display,

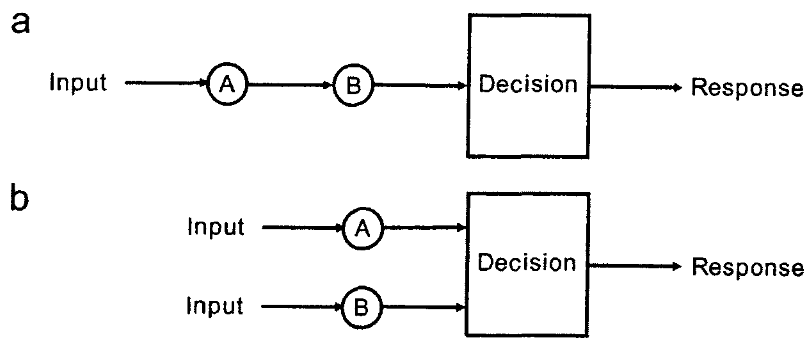

FIGURE 9.2 (a) A serial system, and (b) a parallel system. The input is a source of information for the system, for example a face or a nonface stimulus. "A" and "B" denote two channels of processing, two processes, or two units. For example, "A" and "B" could be face-feature detectors (responding to the presence of eyes and lips). In a serial system both channels process the input information in a nonoverlapping manner, whereas in a parallel system the channels operate simultaneously. After all channels finish processing (for example, the recognition of a face feature) the decision is generated. In other words, upon the positive recognition of all face features the response "I see a face" is generated. Otherwise the response "This is not a face" is generated.

solving facets of a problem, deciding among a set of objects, and so on. Parallel processing means processing all things simultaneously, although it is allowed that they may finish at different times.

Although the term *architecture* might seem to imply rigid structure, we may employ it to refer to such, or to more flexible arrangements. Thus, it might be asserted that certain neural systems are, at least by adulthood, fairly wired in and that they act in parallel (or in some cases, in serial). However, a person might scan the newspaper for, say, two terms, one at a time, that, is serially or, by dint of will, might try to scan for them in parallel. Although parallel versus serial processing is in some sense the most elemental pair of architectures, much more complexity can be imagined and, indeed, investigated theoretically and empirically (e.g., Schweickert, 1978; Schweickert & Townsend, 1989). Figure 9.2 illustrates the flow diagrams associated with serial and parallel processing.

If we are dealing with only one or two channels or items, we shall often just refer to these as a or b, but if we must consider the general case of arbitrary n items or channels, we list them as 1, 2,..., $n - 1$, n. So if $n = 2$, and a and b are stochastically independent (see later material for more on this issue), then the density of the sum of the two serial times is the

so-called convolution of the separate densities. We can't go into the details here (but see e.g., Townsend & Ashby, 1983), but simply note that this new density is designated as $f_a(t) * f_b(t)$, where the asterisk denotes convolution and a and b are processed serially. The mean or expectation of the sum $E[T_a + T_b] = E[T_a] + E[T_b]$, that is, the old result we were taught in statistics that the mean of the sum is the sum of the means: The overall completion time for serial processes is the sum of all the individual means. The standard serial model requires that $f_a(t) = f_b(t)$, which in turn implies that $E[T_a] = E[T_b] = E[T]$, and $E[T_a + T_b] = 2\ E[T]$

In more general settings, one might need to allow for a or b to take different amounts of time, depending on which is processed first. For simplicity, we do not consider that situation here, but even so, it may matter, depending on the stopping rule (see just below), which order is taken. Hence, we can assume that with probability p, a is done first and with probability $1 - p$, b is done first. Figure 9.2 shows the simplest case where, say, item a is always processed first.

In parallel processing, assuming again stochastic independence across the items or channels, the overall completion time for both items has to be the last, or maximum finishing time for either item. Thus, the density that measures the last finishing time is $f_{max}(t) = f_a(t)F_b(t) + f_b(t)F_a(t)$. This formula has an easy interpretation that either a finishes last at time t and b is already done by then, or b finishes last at time t and a is already done by then. In this case, the mean is not so easy to write from first principles. Nonetheless, we can use a trick to do it. It is a very nice fact that the mean of a positive variable T is the integral of the survivor function: $E[T] = \int S(t')\,dt'$, integrating from 0 to infinity. The survivor function in the present situation is $S(t) = 1 - F_a(t)F_b(t)$ and the mean can be calculated from there using the already given integral.

Stopping or Decision Rule: When Does Processing Cease?

No predictions can be made about processing times until the model designer has a rule for when processing stops. In some high-accuracy situations, such as search tasks, it is usually possible to define a set of events, any one of which will allow the processor to stop without error. In search for a set of targets then, the detection of any one of them can serve as a signal to cease processing. A special case ensues when exactly one sought-for target is present. In any task where a subset of the display or memory items is sufficient to stop without error, and the system processor is capable of stopping (not all may be), the processor is said to be capable of *self-termination* (like many terms of specialized argot, this one could perhaps be more descriptive). Because many earlier (e.g., Sternberg, 1966) investigations studied exhaustive versus single-target search, self-termination was often employed to refer to the latter, although it can also have generic meaning and convey, say, *first-termination* when the completion of any of the present items suffices to stop processing. The latter case is often called an *OR design* because completion of any of a set of presented items is sufficient to stop processing and ensure a correct response (e.g., Egeth, 1966; Townsend & Nozawa, 1995).

If all items or channels must be processed to ensure a correct response, then exhaustive processing is entailed. For instance, on no-target (i.e., nothing present but distractors or noise) trials, every item must be examined to guarantee no targets are present. In an experiment where, say, all n items in the search set must be a certain kind of target, called an *AND design*, exhaustive processing is forced on the observer (e.g., Sternberg, 1966; Townsend & Nozawa, 1995). Nevertheless, as intimated earlier, some systems may by their very design have to process everything in the search set, so the question is of interest even when, in principle, self-termination is a possibility.

Hence, in summary, there are three cases of especial interest: (a) minimum time, OR, or first-terminating processing, where the first item to complete stops processing; (b) single-target self-termination, where there is one target among $n - 1$ other items and processing can cease when it is found; and (c) exhaustive or AND processing, where all items or channels are processed. Figure 9.3 depicts AND (exhaustive) and OR (first-terminating) processing in a serial system, whereas Figure 9.4 does the same for a parallel system.

Suppose again there are just two items or channels to process, *a* and *b*, and serial processing is being deployed. Assume that a is processed first. Then the minimum time processing density is simply $f_{min}(t) = f_a(t)$, that is, naturally just the density of a itself. Assume now there is a single target present in channel a and one distractor is in channel b, and self-terminating (ST) serial processing is in force. Then the predicted density is $f_{st}(t) = pf_a(t) + (1 - p)f_b(t) * f_a(t)$. That is, if *a* happens to be checked first, which occurs with probability *p*, then the processing stops. On the other hand, if *b* is processed first and a distractor is found (as it must be), then *a* has to be processed also so the second term is the convolution of the *a* and *b* densities. In the event that both items must be processed (or an

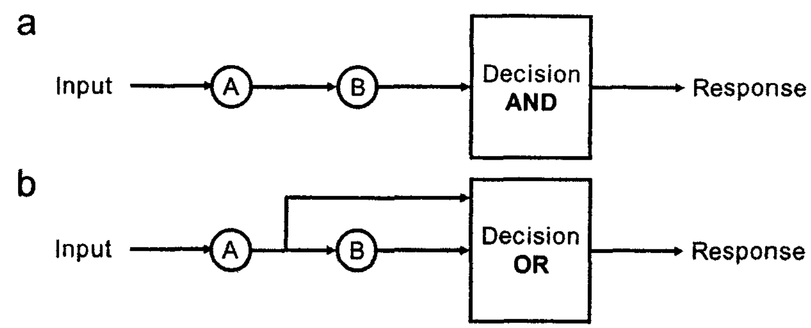

FIGURE 9.3 Schematics of stopping rules in a serial system. (a) A diagram of the standard serial system in the case of "AND" (exhaustive) processing. (b) The stopping rule in the serial system is depicted as an additional arrow which goes from the output of "A" directly to the decision box, allowing for the possibility of bypassing process "B." When the evidence accumulated by process "A" is enough to make a decision then the processing can terminate, and additional processing of "B" is unnecessary.

inflexible serial processor cannot do otherwise), then the prediction is just that given earlier: $f_{max}(t) = f_a(t) * f_b(t)$.

When processing is independent parallel, the minimum time rule delivers a horse race to the finish, with the winning channel determining the processing time (Fig. 9.4a). The density is just $f_{min}(t) = f_a(t)S_b(t) + f_b(t)S_a(t)$. This formula possesses the nice interpretation that a can finish at time *t*, but *b* is not yet done (indicated by b's survivor function), or the reverse can happen. If processing is single-target self-terminating with the target in channel *a*, parallel independence predicts that the density is the simple $f_{st}(t) = f_a(t)$! Finally, if processing is exhaustive (maximum time) and independent, then processing is the same as shown in the introduction to parallel processing, $f_{max}(t) = f_a(t)F_b(t) + f_b(t)F_a(t)$ (Fig. 9.4b).

Independence Versus Dependence Of Channel or Item Processing

The next important issue to discuss is that of independence versus dependence of channels, stages, or subsystems (these terms can be used interchangeably although *stages* is sometimes restricted to serial systems and *channels* to parallel systems).

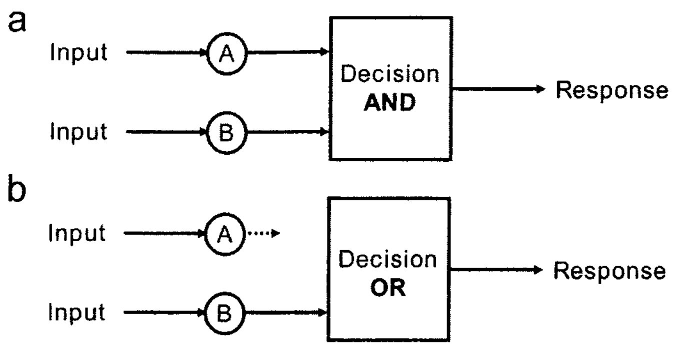

FIGURE 9.4 Schematics of stopping rules in a parallel system. (a) A diagram of the standard parallel system in the case of "AND" (exhaustive) processing. (b) In the "OR" case, the evidence accumulated by process "B" is enough to make a decision and processing can terminate, even though "A" has not yet finished processing (this is indicated by the short arrow).

We have been explicitly assuming independence of the processing times, whether they are serial or parallel. For the present tutorial purposes, somewhat limited space, and without assuming significantly more mathematical background of the reader, we prefer to circumambulate this issue as far as writing out the technical equations goes. Nonetheless, it may be pertinent to give some indication of where it matters.

In serial processing, if the successive items are dependent then what happens on *a*, say, can affect the processing time for *b*. Although it is still true that the overall mean exhaustive time will be the sum of the two means, the second, say *b*, will depend on a's processing time. Hence, if, say, *a* is speeded up, then ordinarily that will affect even the mean time of *b*. Figure 9.5 indicates independence versus dependence in serial systems.

In parallel processing too, the processing times could be dependent. For instance, because they are being processed simultaneously, ongoing inhibition or facilitation (or both!) can take place during a single trial and while processing is ongoing. Townsend and Wenger (2004b) discuss this topic in detail. Figure 9.6 illustrates the concepts of independent versus dependent processing in parallel systems.

It is interesting to note that the above prediction of independent parallel processing in ST situations will no longer strictly hold. However, it will still be true even if processing is dependent that the predicted ST density will be the average or expected value (i.e., known in probability jargon as the marginal) of the density in the channel where the sought-for target is

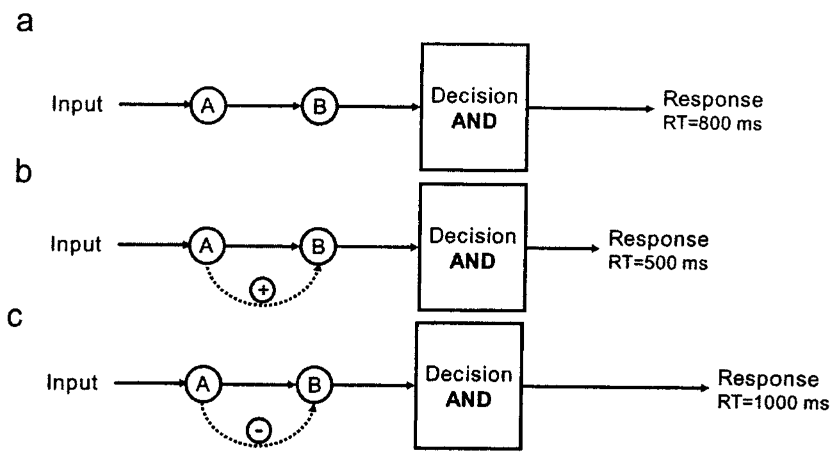

FIGURE 9.5 Dependency between "A" and "B" in a serial system: (a) the standard serial system, and (b) a positively dependent serial system: Duration of "B" depends positively on duration of "A," that is faster processing in "A" will produce facilitation or faster processing in "B" and vice versa. For example, in a face-recognition task faster recognition of the first face feature could give some "confidence" to a second process to speed up processing of a second feature. (c) A negatively dependent serial system: The processing time of "B" is inversely related to a processing time of "A." Faster processing of "A" produces slow processing of "B"; that is, "A" inhibits "B," and vice versa. Overall, a positively dependent system with facilitation exhibits the fastest reaction time (500 msec), whereas a negative dependent system with inhibition exhibits the slowest reaction time (1,000 msec).

located, $E[T_a]$. Only in the nonindependent situation, this expectation has to be taken over all the potential influences from the surrounding channels. The speed-ups or slow-downs shown in Figures 9.5 and 9.6 can be interpreted in terms of the notion of capacity, which we discuss next.

Capacity: Various Speeds on a Speed Continuum

Capacity refers generally to the speed of processing in response-time tasks. We first provide an informal sketch of the major concepts and then turn to a more rigorous exposition. For greater mathematical detail and in-depth discussion see Townsend and Ashby (1978), Townsend and Nozawa (1995), and Townsend and Wenger (2004b). Wenger and Townsend (2000) offer an explicit tutorial and instructions on how to carry out a capacity analysis.

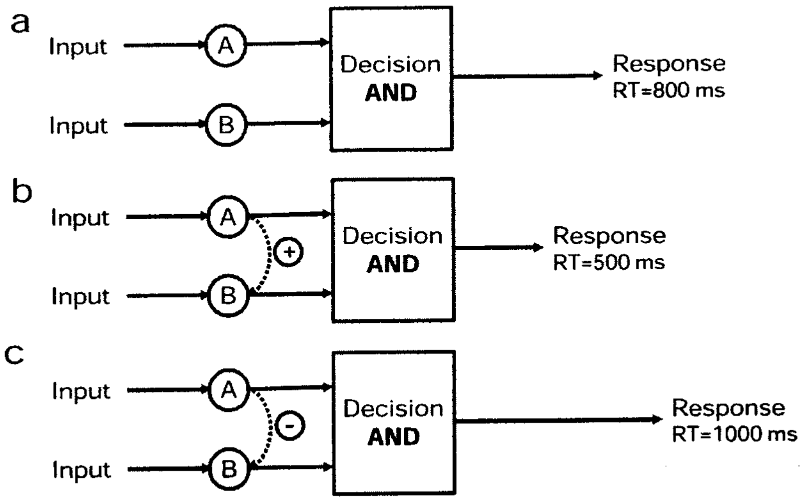

FIGURE 9.6 Dependency between "A" and "B" in a parallel system. (a) The standard parallel system, and (b) a positively dependent parallel system: The positive sign arrow from "A" to "B" indicates positive facilitation. That is, faster processing of one channel speeds up processing in the other channel (as depicted in the figure), and vice versa. (c) A negatively dependent parallel system: The processing time of "A" is inversely related to the processing time of "B." Faster processing of "A" will produce longer processing of "B"; that is, "A" inhibits "B" (as depicted in the figure), and vice versa. Overall, a positively dependent system with the facilitation exhibits the fastest reaction time (500 msec), while a negatively dependent system with the inhibition exhibits the slowest reaction time (1,000 msec).

Informally, the notion of *unlimited capacity* refers to the situation when the finishing time of a subsystem (item, channel, etc.) is identical to that of a standard parallel system (described in more detail later); that is, the finishing times of the distinct subsystems are parallel, probabilistically independent, and the finishing times of each do not depend on how many others are engaged (e.g., in a search task the finishing time of one item is invariant over the total number of items being searched). *Limited capacity* refers to the situation when item or channel finishing times are less than what would be expected in a standard parallel system. *Supercapacity* indicates that individual channels are processing at a rate even faster than standard parallel processing. Figure 9.7 illustrates the general intuitions accorded these concepts, again in an informal manner.

We pause to observe that, although the stopping rule obviously affects overall processing times, as indicated in Figure 9.8 for both serial and

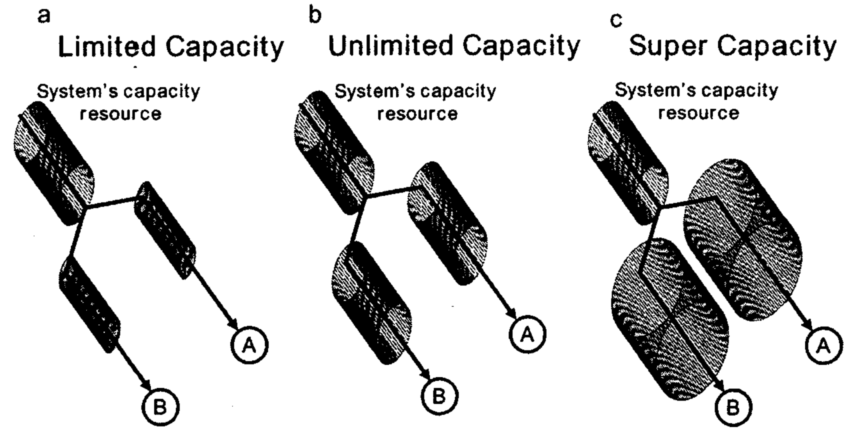

FIGURE 9.7 Graphical intuition of a system's behavior under different capacity bounds: limited capacity, unlimited capacity, and supercapacity. The total system's capacity resource remains the same across all conditions. (a) In the limited-capacity case, the total capacity is split between two channels. (b) In the case of the unlimited capacity, each channel receives the total capacity. (c) In the supercapacity case, the capacity devoted to each channel exceeds the total system capacity. Note that an increase in channel capacity produces faster processing for that channel.

parallel systems, we assess capacity (i.e., efficiency of processing speed) in comparison with standard parallel processing with specification of a particular stopping rule. Thus, although the minimum time (first-terminating or OR processing) decreases as a function of the number of items undergoing processing (because all items are targets), the system is merely unlimited, not super, because the actual predictions are from a standard parallel model (i.e., unlimited capacity with independent channels). But observe that each of the serial predictions would be measured as limited capacity because for each stopping rule, they are slower than the predictions from standard parallel processing.

Figure 9.8 indicates mean response times as a function of workload. *Workload* refers to the quantity of labor required in a task. Most often, workload is given by the number of items that must be operated on in some fashion. For instance, *workload* could refer to the number of items in a visual display that must be compared with a target or memory item. Although Figure 9.8 indicates speed of processing through the mean response times, there are various ways of measuring this speed. The mean = $E[T]$ is a rather coarse level of capacity measurement. A stronger gauge is found in the cumulative distribution function $F(t)$, and the hazard

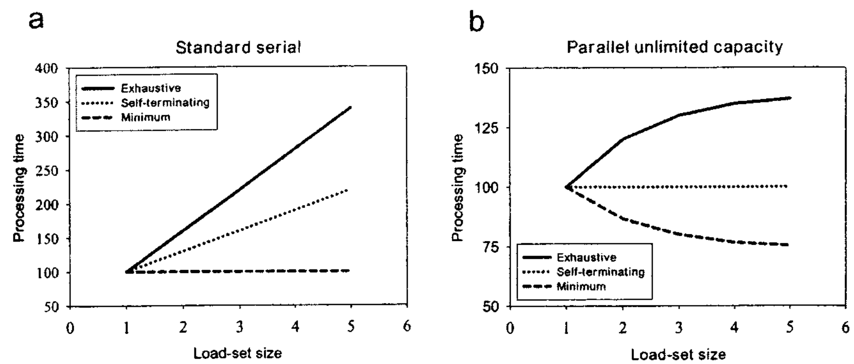

FIGURE 9.8. Expected processing time as a function of load-set size for different stopping rules (exhaustive, self-terminating, and minimum) for (a) the standard serial model, and (b) the parallel unlimited capacity processing model. The load-set size is defined as number of processing units of information: usually number of memorized items or items in the visual filed that have to be searched.

function (to be discussed momentarily) is an even more powerful and fine-grained measure. This kind of ordering is a special case of a hierarchy on the strengths of a vital set of statistics (Townsend, 1990b; Townsend & Ashby, 1978).

The ordering establishes a hierarchy of power because, say, if $F_a(t) > F_b(t)$ then the mean of a is less than the mean of *b*. However, the reverse implication does *not* hold (the means being ordered does not imply an order of their cumulative distribution functions). Similarly if $h_a(t) > h_b(t)$ then $F_a(t) > F_b(t)$, but not vice versa, and so on. Obviously, if the cumulative distribution functions are ordered then so are the survivor functions. That is, $F_a(t) > F_b(t)$ implies $S_a(t) < S_b(t)$.

There is a useful measure that is at the same strength level as *F* or *S*. This measure is defined as $-\log_e[S(t)]$, that is, minus one times the natural logarithm of the survivor function. It turns out that this is actually the integral of the hazard function $h(t')$ from 0 to *t* (e.g., Wenger & Townsend, 2000; illustrative uses of this and related measures, below, in clinical science are presented in Neufeld, Townsend, & Jetté, in press). We thus write the integrated hazard function as $H(t) = -\log_e[S(t)]$. Although it is of the same level of strength as $S(t)$, it has some very helpful properties not directly shared by $S(t)$.

Now we are in a position to compare two or more experimental situations by comparing their statistics. For example we might compare Condition 1

to Condition 2 as in $S_1(t)$ versus $S_2(t)$, or say, $H_1(t)$ versus $H_2(t)$—or,more easily, consider $H_1(t)/H_2(t)$. If this ratio is greater than 1 for all reasonable values of *t*, then we know that the capacity (efficiency, speed) of Condition 1 is greater than that in Condition 2 and in a quite strong sense.

As a special case of great import, assume now that we change the number of channels or items that must be processed, in the present context, say, n goes from 1 to 2. Suppose we wish to measure the effect of this increase in workload in a situation where an efficient system can stop in the first-terminating (minimum) time.

We require a measuring instrument, in a sense, because there is no elementary ruler we can use for arbitrary capacity measurement. Our measuring instrument is that of the set of predictions by unlimited-capacity independent parallel processing. *Unlimited capacity* means here that each parallel channel processes its input (item, etc.) just as fast when there are other surrounding channels working (i.e., with greater *n*) as when it is the only channel being forced to process information.

Now it has been demonstrated that when processing is of this form, then the sum of the integrated hazard functions for each item presented alone is precisely the value, for all times *t*, of the integrated hazard function when both items are presented together (Townsend & Nozawa, 1995). That is, $H_a(t) + H_b(t) = H_{ab}(t)$. This intriguing fact suggests the formulation of a new capacity measure, which the latter authors called the *capacity coefficient* $C(t)$ and set it equal to $C(t) = H_{ab}(t)/[H_a(t) + H_b(t)]$, that is, the ratio of the double item condition over the sum of the single item conditions.

If this ratio is identical to 1 for all *t*, then the processing is identical to that of an unlimited capacity independent parallel model. If $C(t)$ is less than 1 for some value of *t*, then we call processing *limited*. For instance, either serial processing of the ordinary kind or a fixed-capacity parallel model that spreads the capacity equally across a and b predicts $C(t) = ½$ for all times $t > 0$. If $C(t) > 1$ at a time (or any, or maybe all times *t*) *t*, then we call the system *supercapacity* for those times. A tutorial on capacity and how to assess it in experimental data is offered in Wenger and Townsend (2000), with clinical-science applications being illustrated in Neufeld, Townsend, and Jetté (in press). In a recent extension of these notions, we have shown that if configural parallel processing is interpreted as positively interactive parallel channels (thus being dependent or positively correlated rather than independent), then configural processing can produce striking supercapacity (Townsend & Wenger, 2004b).

Up to this point, in addition to providing motivation and relevance to clinical science, we have reviewed some required probability tools and, more important, introduced a set of critical dimensions from our information processing theory. These included the focus of this chapter: parallel versus serial processing. It is possible to construct a huge variety of processing systems simply by combining different values from each dimension (e.g., moderately limited capacity parallel processing with negatively dependent channels and with an exhaustive processing rule imposed). Nonetheless, certain particular systems have gained almost archetypal status in cognitive science. In fact, this is so much so that many times investigators seem to operate as if they are the *only* available or possible systems in nature. Of course, this is far from true, but these prototypical models bear special consideration on our part. We next take up these prominent models. After that,we return to a more in-depth discussion of automatic versus controlled processing. This section is followed by presentation of a relatively recent and powerful experimental approach that permits assessment of the basic dimensions of information processing, using response times.

PROMINENT ARCHITECTURES

The first of the major architectures comprises the standard serial class of model. This class has been quantitatively well understood at least since Sternberg's (1966) initial papers. The standard parallel class of models seems to be less well comprehended at large, although a number of psychological notions, for instance, automaticity, can be captured by this type of processing. The third architecture, coactive parallel processing, is a relatively recent contender—as we show later, these models permit performance that is superior even to standard parallel processing. These models make distinct predictions even at the level of mean response times, as a function of workload, as indicated in this section. However, they are still open to problems with model mimicking at this level. A later section, Experimental Testing of Parallel and Serial Architectures, shows how to effectively circumvent the model-mimicking dilemma.

It is worth pointing out that none of these models' characteristic predictions rely on any special kind of probability density—that is, the predictions are a quantitative form of qualitative, and are thus so-called distribution-free and, of course, are parameter-free as well. The *latter* stipulation simply means that the predictions do not depend on any particular

mean or variance, say of the underlying distributions, although naturally the actual means or variances would depend on the parameters.

Standard Serial Models

This type of model is what most people mean when they only say serial unadorned. Thus, it is the model advocated by S. Sternberg in many of his papers (e.g., Sternberg, 1966, 1969, 1975). To reach it in the case that $n = 2$, simply let $f_a(t) = f_b(t) = f(t)$—that is, the probability densities are the same across items or positions and even n. The latter means that $f(t)$ defines the length of time taken on an item or channel no matter how big or little the entire set of operating items or channels is. Furthermore, it is assumed in the standard serial model that each successive processing time is independent of all others. So, if a is second, say, its time does not depend on how long the preceding item (e.g., b) took to complete its processing.

Note, however, that we still allow in general that different paths through the items might be followed. We also do not confine the stopping rule to a single variety. Now, S. Sternberg's preferred model did assume that exhaustive processing was used even in target-present trials. But because this seems like a secondary issue we allow the standard model to follow other, sometimes more optimal, rules of cessation. Because all the n densities are now the same we can simply write the nth order convolution for exhaustive processing in symbolic form as $f_{max}(t) = f^{*(n)}(t)$. The exhaustive mean processing time is then $E_{max}[T_1 + T_2 + \ldots + T_n] = n\,E[T]$.

Next consider the situation where exactly one target is present among $n - 1$ distractors and the system is self-terminating. Again, it is assumed that the target is placed with probability $1/n$ in any of the n locations. Then it follows that $f_{st}(t) = (1/n) \sum f^{*(i)}$ where the summation goes from $i = 1$ to $i = n$. The mean processing time in this case is the well-known $E_{st}[T] = (n + 1)E[T]/2$. This formula can be interpreted that on average, it takes the searcher approximately one-half of the set of items to find the target and cease processing. Finally, when processing stops as soon as the first item is finished, then we have the result $f_{min}(t) = f(t)$ and the elemental $E_{min}[T] = E[T]$.

Standard Parallel Models

The standard parallel model assumes independence again among the processing items, but this time in a simultaneous sense. At this point, we nake the decision of whether to force all the channels to process :he same (stochastic) speed. Why not? After all, we did just that standard serial model. However, the standard serial model could uce so-called position effects. Position effects are produced when ocations of a target produce different mean processing times and, :e associated with distinct densities (e.g., Sternberg, 1966; Van Townsend, 1993). Standard serial models can do this by letting processing paths through the items occur with different proba- 'or instance, if $n = 3$, the path $\langle a, b, c\rangle$ might be taken with ty 1/3, the path $\langle c, a, b\rangle$ with probability 1/6, the path $\langle a, c, b\rangle$ ›ability ½, and all the other paths never occur. This distribution would make shorter times on average of a over b and c and the 'ould tend to be shorter than those of b.

nly way an independent parallel model can generate position f the distinct channels or items have different densities, as in $f_a(t)$ ence, this provision is usually allowed in standard parallel models. event, because we always assumed independence in the above t of parallel models, the formulas for $n = 2$ stay the same. , for simplicity, take the special case where the densities are all Then $E[MAX(T_1, T_2, \ldots, T_n)] = \int [1 - F^n(t')]\, dt'$ with the inte- ǥ taken from 0 to infinity. It is straightforward to show that in the curve of mean processing times (and therefore response always increasing but with a concave-down shape. In this espe- ple case, the single self-terminating target case, among $n - 1$ dis- s just $E[T]$, the time required for any single item to complete. , the time required for the minimum or first-terminating time, ıe time of the winning horse, is $E[MIN(T_1, T_2, \ldots, T_n) = \int [1 -$ and again the integral is from 0 to infinity. Here, it can be demon- at the mean times in this kind of model (and this is true even if e distributions are identical), the curve of mean times is always g, concave up. Egeth (1966) employed this characteristic fre- › argue for parallel processing because it is an unnatural prediction nodels. The proofs of the theorems on concavity of reaction time a function of load are provided by Townsend and Ashby (1983).

e Parallel Models

n the late 1980s, J. Miller (1982, 1986) began to produce data ed to indicate that processing could be even better, more capacious than even ordinary (or we would now say, standard parallel

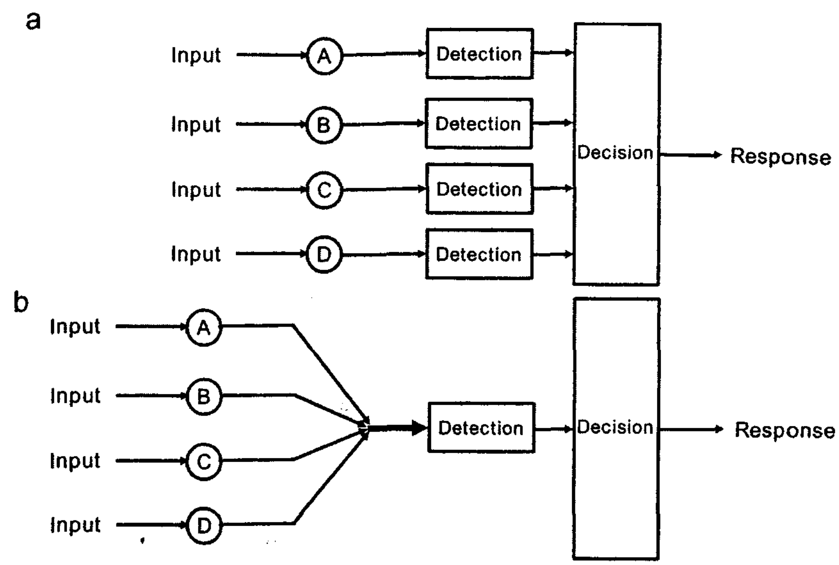

FIGURE 9.9 Schematics of (a) a parallel independent system, and (b) a coactive multiple channels processing systems. A coactive model assumes that an input from separate parallel channels is consolidated into a resultant common processor before a decision is made.

action). His main line of argument used a clever probability inequality that ordinary parallel processing would have to satisfy but that extraordinary or, as we might say today, supercapacity operations could violate.

Before long, a number of investigators, including Miller, commenced to develop actual nonstandard parallel models that indeed would violate the inequality (e.g., Diederich, 1992, 1995; Diederich & Colonius, 1991; Schwarz, 1994; Townsend & Nozawa, 1995; Townsend & Wenger, 2004b). All these models possess the property that activity in the separate channels was summated or pooled into a final common channel before a detection decision was made. In standard and in fact, any parallel model where separate detection decisions are made in their individual channels, this pooling does not occur. Figure 9.9 exhibits a comparison between ordinary parallel processing where separate decisions (detections, etc.) are made on the distinct channels as opposed to coactive processing where the activations on the several channels are combined, for instance, summed arithmetically.

Subsequently, a general theory of capacity was formulated that permitted the measurement of processing efficiency for all times during a trial (Townsend & Nozawa, 1995). Employing standard parallel processing as a cornerstone, the theory defined unlimited capacity as efficiency identical to that of standard parallel processing in which case the measure is $C(t) = 1$. It defined limited capacity as efficiency slower than standard parallel processing. For instance, standard serial processing produces a measure of capacity of $C(t) = .5$. And finally, the theory defined supercapacity as processing with greater efficiency than standard parallel models could produce, that is, $C(t) > 1$. It was then proven that a very broad class of coactive parallel models, which included all the special ones constructed before that, would imply $C(t) > 1$ and would inevitably violate Miller's inequality.

Because our focus here must be on architecture, we have not been able to expend much space on capacity, but the reader is referred to Wenger and Townsend (2000) for a tutorial on capacity measurement, Townsend and Ashby (1978) for the early work on this concept, and Townsend and Wenger (2004b) for the latest theoretical results on capacity in interactive (i.e., not independent) parallel systems.

Note that the capacity construct has been prominent in clinical cognitive science. Initial extensions of mathematical treatments of the construct to this arena of study are reported in Neufeld, Vollick, and Highgate (1993). Subsequent developments are enumerated in Neufeld (in press-a) and Neufeld, Carter, Vollick, Boksman, Levy, and Jetté (in press), and inaugural implementations in this field of the integrated hazard function H(t) and the capacity coefficient appear in Neufeld, Townsend, and Jetté (in press).

We have provided a succinct overview of three highly important kinds of processing models: standard serial, standard parallel, and coactive parallel. The standard serial model is least efficient because it entails sequential processing of each item with average processing times on each always being the same. Standard parallel processing implies quite efficient processing because each item can be processed at the same rate, regardless of how many others are being operated on, and all are processed simultaneously. Coactive parallel processing can be exceedingly fast by virtue of all of the item channels pouring their activation into a single final conduit. The next section revisits the very popular topic of automatic processing.

AUTOMATICITY: INTERPRETATION VIA ARCHITECTURE AND CAPACITY

The notion of automaticity has rarely, if ever, been given a rigorous mathematical definition. One primary correlate has been superior efficiency of processing as expressed in response times.

In fact, often, the criterion investigators use to ascribe processing as automatic, is simply that a mean response-time function of workload be flat. Because, as we have seen, response speed depends on all the major characteristics of a system, the investigator must be wary when drawing conclusions about automaticity even from this simple standard. Thus, even a standard serial model predicts a flat response-time function for a first-terminating stopping rule. Furthermore, a standard parallel model with its independent, unlimited capacity, will predict a flat response-time curve when the system is searching for a single target among $n - 1$ distractors. We further witnessed that with first-termination, such a model will actually predict a decrease in mean response times.

It is not clear to the authors whether the class of standard parallel models are sufficiently free of capacity limitations to qualify as automatic processing in the minds of most investigators of this topic. It would seem that a coactive model, with channels left undegraded as workload n increases, should merit that assessment. This question is primarily one of convention, but it would be propitious if agreement on a rigorous set of criteria could be reached.

We also have to take account of the differing experimental paradigms where the phenomenon is concluded to exist. The modern instantiation of this concept stems from the Schneider and Shiffrin (1977) studies on visual and memory search mentioned in the introduction. We haven't the space to consider the wide array of experimental conditions they ran and in particular must ignore their results on accuracy in favor of the response-time dependent variable under very high accuracy. Their basic findings, replicated scores of times, were that both the single target present (i.e., potentially single-target self-terminating) and target absent (forced exhaustive processing) response times were flat or almost flat. Standard parallel processing models easily predict the first result but not the second. If mean response times for exhaustive processing are flat, then processing is highly supercapacity. For instance, Townsend and Ashby (1983) show that a model whose channels are unlimited capacity at the start of a trial, but that is capable of reallocating capacity from completed channels to uncompleted ones, predicts flat exhaustive mean response-time curves. However, either processing is also exhaustive on the target present trials, or is less super in capacity on the latter—an unlimited capacity parallel model with reallocation would predict decreasing mean response-time functions.

A similar but not identical topic arises when one speaks of gestalt figures or holistic processing. A unified Gestalt or holistic percept might be more or less wired in, or at least installed early in life, as in face perception, or learned for some purpose later on. One expression for the welding of several initially separate parts into a whole is *unitization*, a term employed by Czerwinski, Lightfoot, and Shiffrin (1992) in their study of this kind of learning. Another example stems from Goldstone's (2000) experiments, where observers learned to weld together a set of originally meaningless squiggles into a perceptually holistic object. All of the squiggle features in the designated object had to be perceived in order to be correct, forcing exhaustive processing. Employing concepts from cognitive stochastic process theory (e.g., Townsend & Ashby, 1983), he showed that the various parts of the holistic object were perceived in a supercapacity (i.e., better than standard parallel processing) fashion. Subsequent replicative experiments employing new measures of capacity by Blaha and Townsend (2004a) have confirmed and strengthened Goldstone's conclusions. Blaha and Townsend (2004b) have further developed a neuralistic model based on a dynamic system instantiation of Hebbian concepts, which produces the massive supercapacity found in these investigations.

Another experimental example of a kind of automatization may be that of the well-known and documented pop-out effect. This effect occurs whenever a target that is sufficiently distinct from all the distractors is used, as, for instance, when a colored object is placed among a set of gray distractors or a green object is placed among a set of red distractors (e.g., Treisman & Gormican, 1988; see also Van Zandt & Townsend, 1993). Here the emphasis is obviously on the target present case. Response times on these trials are flat across set size *n*. Hence, standard parallel processing can account for these results without having to posit super capacity.

In summary, as noted before, the literature and phenomena of automatization are vast. Nonetheless, it appears that at least in certain publicized cases, and in particular where a small set of targets is identified with the same speed across increasing workload n (i.e., increasing the number of present distractors), standard parallel processing is a sufficient explanation for automatization. However, when processing has to be exhaustive and yet the response-time curves are flat across n, the system has to be exceedingly supercapacity. Such cases appear in the Schneider and Shiffrin (1977) target absent data as well as the Goldstone (2000) and later the Townsend and Blaha (2004a, 2004b) investigation. Positively interactive parallel processing or coactive parallel processing can readily produce such findings (Townsend & Wenger, 2004b).

It is apparent that a simple and universal isomorphism between the concept of automaticity and rigorous information-processing dimensions is a chimera. It is especially vital to understand that there may be more than one way to *skin the cat*, so to speak—that is, to avoid confusing sufficiency of a model with its necessity. For instance, Kahneman (1973) promoted the idea that systems might in some situations be capable of drawing on extra resources when workload is increased. Such extra resources could transform standard parallel processing into supercapacity (and apparently automatic) parallel processing.

In any event, in each individual research area, our taxonomy can be applied to endow this exceedingly efficient type of processing with precise meaning and experimental implications.

EXPERIMENTAL TESTING OF PARALLEL VERSUS SERIAL ARCHITECTURES

There now exist several experimental assays of mental architecture. Most of these circumvent the major impediment of the ability of limited capacity parallel models to mimic the behavior of standard serial models. However, before embarking on the primary target of this section, we observe that serial models that can mimic unlimited or supercapacity parallel models are typically quite unintuitive, so that evidence of such behavior can be taken as falsifying serial architectures (e.g., Townsend, 1971, 1974). For more complete reviews of, and references to, the available panoply of serial–parallel assessment techniques, the reader is pointed to Townsend and Wenger (2004a), Townsend (1990a), and Townsend and Ashby (1983).

One major set of strategies has sprung from S. Sternberg's (1969) additive factors method. This method and its descendents, rests on the assumption of *selective influence*. Selective influence assumes that specified experimental variables separably affect distinct processing systems. In the additive factors method, the subsystems are often called *stages*. The stages consume some random amount of time and it is postulated that these processing times do not overlap, although it could be that switching times between stages could add some additional time.

Suppose we are concerned with just two subsystems (stages for serial processing). Let us refer to these as a and b as usual and name the experimental factors that selectively affect them A and B. Then the next step is to perform a factorial experiment with factors $A_i \times B_j$, where i and j

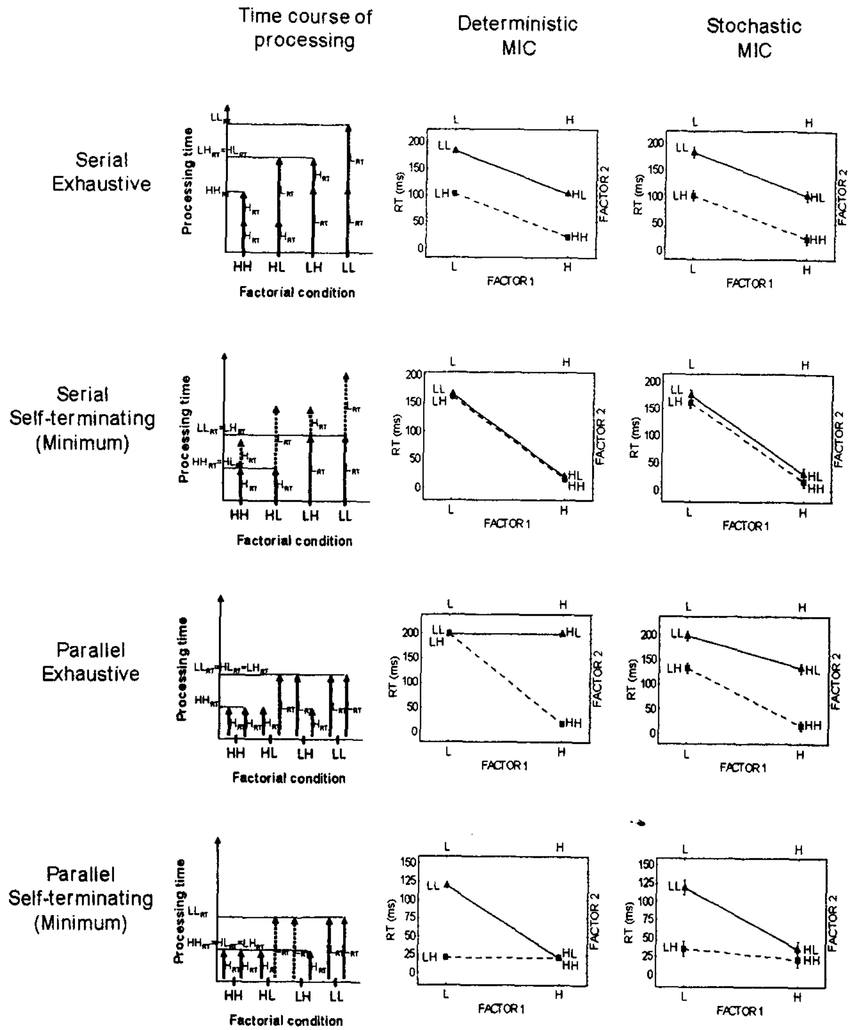

FIGURE 9.10 The time course of processing of two items (left column), the corresponding deterministic (middle column), and stochastic (right column) mean interaction contrast (MIC), across different architectures and stopping rules (rows). The time course of processing depicts the change in total processing time for different factorial conditions (HH, HL, LH, LL) for different architectures. Each upright bold arrow in the graph corresponds to total processing of one unit (in the left column), which could be at the H (high) or L (low) level. A dotted upright arrow indicates a process that did not complete because the processing terminated on a completion of the previous process. The deterministic MIC, in the middle column, represents the duration or the sum of process times (as indicated on the y-axis in the first column). Note that we are not able to directly observe the deterministic MIC in experiments because in a real system processing components will add some variability or noise. The stochastic MIC is an observable measure and is obtained when some variability or noise is added to the overall processing. Error bars around each mean condition represent standard error statistic (added here arbitrarily for the sake of presentation). Also observe also that each architecture combined with a different stopping rule exhibits a different MIC value.

indexes run over the various levels of each factor (e.g., i or j = 1, 2). Because of selective influence and the successive nonoverlapping times from the stages, it is then predicted that the two (or more) experimental factors will exhibit additive effects in the mean response times garnered in the factorial experiment.

Let the random times for the two stages be named $T_a(A_i)$ and $T_b(B_j)$, respectively, and set the overall mean (=expected) exhaustive processing time under condition $A_i \times B_j$ as $E[T_{ij}]$. Then this overall mean exhaustive processing time for condition $A_i \times B_j$ in this serial model is $E[T_{ij}] = E[T_a(A_i) + T_b(B_j)]$ and by virtue of the aforementioned elementary statistical fact that the mean of the sum equals the sum of the means we arrive at $E[T_a(A_i) + T_b(B_j)] = E[T_a(A_i)] + E[T_b(B_j)]$.

By adopting the convention that $E[T_a(A_i)] = t_a(A_i)$, it is straightforward to compute the mean interaction contrast (MIC) as MIC = $E[T_{11}] - E[T_{12}] - \{E[T_{21}] - E[T_{22}]\} = E[T_{11}] - E[T_{12}] - E[T_{21}] + E[T_{22}] = t_a(A_1) + t_b(B_1) - [t_a(A_2) + t_b(B_1)] - [t_a(A_1) + t_b(B_2)] + [t_a(A_2) + t_b(B_2)] = 0$. This little operation demonstrates the additivity, and therefore the zero MIC, of the serial model under selective influence.

Figure 9.10 shows mean reaction time predictions for each mental architecture combined with different stopping rules. For each model time, course of activation is depicted in the first column. Second and third columns of Figure 9.10 show both additive and stochastic MIC predictions for each model, while the stochastic MIC is empirically observed only. For the aforementioned serial exhaustive processing, it is evident that indeed $E[T_{ll}] - E[T_{lh}] = E[T_{hl}] - E[T_{hh}]$, eventuating in the MIC of zero.

Interestingly, first- and single-target terminating stopping rules also result in additive (implying MIC = 0) response-time factors. However, we show later that a more penetrating statistic is able to distinguish the stopping rules.

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.

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. Stopping rule matters here, too. 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 (i.e., first-terminating parallel processing) will evidence a positive contrast (see Townsend & Nozawa, 1995).

The case of single-target termination has not been much discussed or applied, but it is easy to demonstrate that, intriguingly enough, this case implies additive factors. Thus, parallel predictions are shown in Figure 9.10. Note that if only a single target condition were employed, the serial and parallel predictions are the same. The general set of experimental strategies that include nonserial architecture and also statistics other than the mean (discussed next) has been called *systems factorial technology*.

Systems Technology Response-Time Factorial Distributions

As 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 entire cumulative probability distribution (i.e., the integral of the density from 0 to t) function on response times is more powerful than the means alone.

We observe that statistics at the distributional level deliver much deeper and more conclusive information about processing architecture and stopping rules than was possible with mean response times. For instance, a test case presents itself in the question as to whether coactive parallel processing can be distinguished, using factorial methods, from ordinary parallel processing. It turns out that at the level of mean response times (RTs) and within an OR design, coactive parallel processing 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 MIC 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, & Dzhafarov, 2000). Because the original derivations were in terms of the survivor

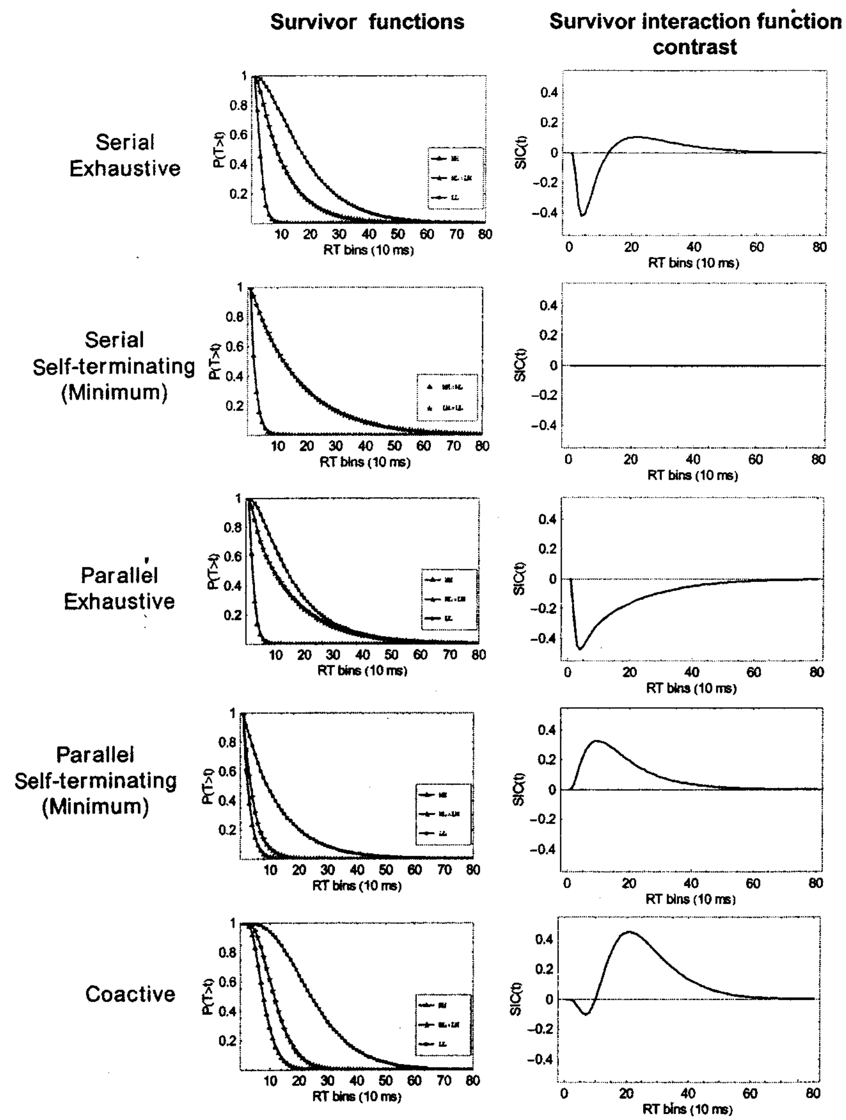

FIGURE 9.11 An ordering of joint survivor functions for different factorial conditions (HH, HL, LH, LL) (left column) and the survivor interaction contrast (SIC) (right column) across different architectures and stopping rules (rows). Note that each SIC function is calculated using SIC(t) = Sll(t) − Slh(t) − Shl(t) + Shh(t). Each joint survivor function on the right-hand side is estimated from data (displayed in the left column). Note that each combination of architecture and stopping rule exhibits a unique SIC function. The shapes of these different SIC functions are independent of the form of the probability density function.

function (Townsend & Nozawa, 1995), we employ that one here, with the interaction contrast for the survivor function (SIC):

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

Figure 9.11 displays the predictions for the various models. Observe that ordinary 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 consistently positive, whereas the contrast for the coactive model possesses a small negative blip at the earliest times before going positive. Because MIC must be positive in coactivation, the positive portion of the SIC always has to exceed the negative portion. Calculation of the SIC function in reaction time experiments could be a laborious job when using some standard statistical packages. In the appendix, we provide guidelines for calculating SIC function. Two scripts that calculate and display the survivor interaction contrast function are available for download on the Psychology Press Web site (http://www.psypress.com/brainscans-etc), written for the Mathematica and Matlab environments. Details of implementations are displayed as the comment sections within each script.

The advantages associated with the use of both the SIC and the MIC go 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, because in the OR case SIC = 0 always, but in the AND case there is a large negative portion of the 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 (2000) to general feedforward architectures, which contain parallel and serial subsystems, and from advances in methods of estimating entire RT distributions (see in particular, Van Zandt, 2000, 2002).

CONCLUSION

In this brief space, we have come all the way from underscoring the importance of discerning mental architecture in clinical cognitive science, through enumeration of the chief issues encountered in engaging this challenge, to mathematical-theory spawned technology for resolution.

Developments in contemporary mathematical cognitive science represent compelling supplements, and even alternatives, to currently proffered batteries of measures aimed at mapping cognitive functioning among clinical samples (e.g., Heinrichs, 2005), or monitoring response to treatment (Nuechterlein, Barch, Gold, Goldberg, Green & Heaton, 2004; cf. McFall & Townsend, 1998). Application in clinical science is not without its problems (Neufeld, 2001). In-depth analyses of clinical-setting exigencies, and tacks to overcoming them, whose exposition is beyond the present scope and space, are available in auxiliary sources (Carter, Neufeld, & Benn 1998; Neufeld, in press-b). We anticipate that advances in quantitative cognitive science will continue to make inroads on the clinical scene, eventuating in significant improvements in cognitive assessment and intervention, and that tutorials of this nature will serve to accelerate the process.

APPENDIX: A GUIDE TO CONSTRUCTING AND USING THE SURVIVOR INTERACTION CONTRAST FUNCTION

1. For each condition, remove the RTs that correspond to errors, equipment failures, anticipatory reactions, and lapses of observer attention.
2. Determine the reaction time bin size. Usually we use 10 msec bin size. The range and the size of the bins vary according to the nature of the task.
3. Count the number of observations in each bin. This step generates a frequency distribution function.
4. Divide each counted frequency (bin) by the total number of observations. This produces relative frequency function (i.e., an empirical probability density function).
5. Calculate the empirical cumulative distribution function (CDF) F(t) for each stimulus condition (HH, HL, LH, LL) by accumulating the empirical probabilities from the lowest to the highest valued bin. This will produce four vectors of data, each describing empirical cumulative distribution function for particular condition ($F_{ll}(t)$, $F_{lh}(t)$, $F_{hl}(t)$, $F_{hh}(t)$).
6. Calculate an empirical survivor function for each condition by subtracting the value of F(t) for each condition from 1; that is, calculate $1 - F(t)$. Do this for four data sets that correspond to each factorial condition. This step will produce four vectors of data,

each describing empirical survivor distribution function for each condition ($S_{ll}(t)$, $S_{lh}(t)$, $S_{hl}(t)$, $S_{hh}(t)$).

7. Calculate the survivor interaction function (SIC) by taking the double difference between survivor data vectors, $\text{SIC}(t) = S_{ll}(t) - S_{lh}(t) - S_{hl}(t) + S_{hh}(t)$, and form the final SIC data vector.

ACKNOWLEDGMENTS

Sources of support for this research included a National Institute of Mental Health grant awarded to Dr. James Townsend and Dr. Mario Fific (NIMH-R01 MH57717-04A1), a Social Sciences and Humanities Research Council of Canada operating grant awarded to Richard W. J. Neufeld, Workplace Safety and Insurance Board, and Canadian Institutes of Health Research operating grants (Richard W. J. Neufeld, co-investigator), and a Canadian Institutes of Health Research New Emerging Teams grant (Richard W. J. Neufeld, co-investigator).

REFERENCES

Blaha, L. M., & Townsend, J. T. (2004a, July). *From nonsense to gestalt: The influence of configural learning on processing capacity.* Paper presented at Society for Mathematical Psychology annual meeting, University of Michigan, Ann Arbor.

Blaha, L. M., & Townsend, J. T. (2004b, November). *A dynamic hebbian model of perceptual learning.* Poster presented at the Psychonomic Society annual meeting, Minneapolis, MN.

Broga, M. I., & Neufeld, R. W. J. (1981). Multivariate cognitive performance levels and response styles among paranoid and nonparanoid schizophrenics. *Journal of Abnormal Psychology, 90,* 495–509.

Carter, C. S., Robertson, L. C., Chaderjian, M. R., Celaya, L. J., & Nordahl, T. E. (1992). Attentional asymmetry in schizophrenia: Controlled and automatic processes. *Biological Psychiatry, 31,* 909–918.

Carter, J. R., & Neufeld, R. W. J. (1999). Cognitive processing of multidimensional stimuli in schizophrenia: Formal modeling of judgment speed and content. *Journal of Abnormal Psychology, 108,* 633–654.

Carter, J. R., Neufeld, R. W. J., & Benn, K. D. (1998). Application of process models in assessment psychology: Potential assets and challenges. *Psychological Assessment, 10,* 379–395.

Czerwinski, M., Lightfoot, N., & Shiffrin, R. M. (1992). Automatization and training in visual search. *American Journal of Psychology. Special Views and varieties of automaticity, 105*(2), 271–315.

Diederich, A. (1992). Probability inequalities for testing separate activation models of divided attention. *Perception & Psychophysics, 52*(6), 714–716.

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology, 39*(2), 197–215.

Diederich, A., & Colonius, H. (1991). A further test of the superposition model for the redundant-signals effect in bimodal detection. *Perception & Psychophysics, 50*(1), 83–86.

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

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 86–112.

Granholm, E., Asarnow, R. F., & Marter, S. R. (1996a). Dual-task performance operating characteristics, resource limitations, and automatic processing in schizophrenia. *Neuropsychology, 10,* 11–21.

Granholm, E., Asarnow, R. F., & Marder, S. R. (1996b). Display visual angle and attentional scanpaths on the span of apprehension task in schizophrenia. *Journal of Abnormal Psychology, 105,* 17–24.

Hartlage, S., Alloy, L. B., Vázquez, C., & Dykman, B. (1993). Automatic and effortful processing in depression. *Psychological Bulletin, 113,* 247–278.

Heinrichs, R. W. (2005). The primacy of cognition in schizophrenia. *American Psychologist, 60,* 229–242.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Knight, R. A., Manoach, D. S., Elliott, D. S., & Hershenson, M. (2000). Perceptual organization in schizophrenia: The processing of symmetrical configurations. *Journal of Abnormal Psychology, 109,* 575–587.

Magaro, P. A. (1983). Psychosis and schizophrenia. In W. D. Spaulding & J. K. Cole (Eds.), *Theories of schizophrenia and psychosis: Nebraska Symposium on Motivation* (Vol. 31, pp. 157–229). Lincoln: University of Nebraska Press.

McFall, R. M., & Townsend, J. T. (1998). Foundations of psychological assessment: Implications for cognitive assessment in clinical science. *Psychological Assessment, 10*(4), 316–330.

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

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics, 40*(5), 331–343.

Neufeld, R. W. J. (2001, April). *Formal models in explanation and measurement of cognitive psychopathology.* Paper presented at Mount Sinai Workshop on Cognition in Schizophrenia: International Congress on Schizophrenia Research, Whistler, British Columbia, Canada.

Neufeld, R. W. J. (in press-a). Composition and uses of formal clinical cognitive science. In W. Spaulding & J. Poland (Eds.), *Modeling complex systems: Motivation, cognition and social processes: Nebraska Symposium on Motivation* (pp.). Lincoln: University of Nebraska Press.

Neufeld, R.W.J. (in press-b). Introduction. In R. W. J. Neufeld (Ed.), *Advances in clinical cognitive sciences: Formal modeling and assessment of processes and symptoms.* Washington, DC: American Psychological Association Publications.

Neufeld, R. W. J., & Broga, M. I. (1981). Evaluation of information sequential aspects of schizophrenic performance: II. Research strategies and methodological issues. *Journal of Nervous and Mental Disease, 169,* 569–579.

Neufeld, R. W. J., Carter, J. R., Vollick, D., Boksman, K., Levy, L., & Jetté, J. (in press). A mathematical process-representation of stimulus encoding subserving memory search and other collateral processes, among paranoid and nonparanoid schizophrenic patients. In R. W. J. Neufeld (Ed.), *Advances in clinical cognitive science: Formal modeling and assessment of processes and symptoms*. Washington, DC: American Psychological Association.

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*(2), 193–226.

Neufeld, R. W. J., Townsend, J. T., & Jetté, J. (in press). Quantitative response time technology for measuring cognitive-processing capacity in clinical studies. In R. W. J. Neufeld (Ed.), *Advances in clinical cognitive science: Formal modeling and assessment of processes and symptoms*. Washington, DC: American Psychological Association.

Neufeld, R. W. J., Vollick, D., Carter, J. R., Boksman, K., & Jetté, J. (2002). Application of stochastic modeling to the assessment of group and individual differences in cognitive functioning. *Psychological Assessment, 14*(3), 279–298.

Neufeld, R. W. J., Vollick, D., & Highgate, S. (1993). Stochastic modelling of stimulus encoding and memory search in paranoid schizophrenia: Clinical and theoretical implications. In R. L. Cromwell & R. C. Snyder (Eds.), *Schizophrenia: Origins, processes, treatment, and outcome: The Second Kansas Series in Clinical Psychology* (pp. 176–196). Oxford, England: Oxford University Press.

Nuechterlein, K. H., Barch, D. M., Gold, J. M., Goldberg, T. E., Green, M. F., & Heaton, R. K. (2004). Identification of separable cognitive factors in schizophrenia. *Schizophrenia Research, 72*, 29–39.

Parzen, E. (1960). *Modern probability theory and its applications*: Oxford, England: Wiley.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1–66.

Schwarz, W. (1994). Diffusion, superposition, and the redundant-targets effect. *Journal of Mathematical Psychology, 38*(4), 504–520.

Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology, 18*(2), 105–139.

Schweickert, R. (1983). Latency network theory: Scheduling of processes in sentence verification and the Stroop effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9(3), 353–383.

Schweickert, R., Giorgini, M., & Dzhafarov, E. (2000). Selective influence and response time cumulative distribution functions in serial-parallel task networks. *Journal of Mathematical Psychology, 44*(4), 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*(3), 328–347.

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson & R. J. Herrnstein (Eds.), *Stevens' handbook of experimental psychology: Vol. 1. Perception and motivation: Vol. 2. Learning and cognition* (2nd ed., pp. 739–811). Oxford, England: Wiley.

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. In W.G. Koster (Ed.), *Attention and performance* (Vol. 2, pp. 276–315). Amsterdam: North Holland.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27*(1), 1–32.

Townsend, J. T. (1969). *Mock parallel and serial models and experimental detection of these*. Paper presented at the Purdue Centennial Symposium on Information Processing, Purdue University.

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

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: Lawrence Erlbaum Associates.

Townsend, J. T. (1990a). Serial versus parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1*(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*(3), 551–567.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. III, pp. 200–239): Hillsdale, NJ: Lawrence Erlbaum Associates.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge, England: Cambridge University Press.

Townsend, J. T., & Fific, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception & Psychophysics*, 66(6), 953–962.

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(4), 321–359.

Townsend, J. T., & Wenger, M. J. (2004a). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review, 11*(3), 391–418.

Townsend, J. T., & Wenger, M. J. (2004b). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 111*(4), 1003–1035.

Treat, T. A., McFall, R. M., Viken, R. J., & Kruschke, J. K. (2001). Using cognitive science methods to assess the role of social information processing in sexually coercive behavior. *Psychological Assessment, 13*(4), 549–565.

Treat, T. A., McFall, R. M., Viken, R. J., Nosofsky, R. M., MacKay, D. B., & Kruschke, J. K. (2002). Assessing clinically relevant perceptual organization with multidimensional scaling techniques. *Psychological Assessment, 14*(3), 239–252.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1), 15–48.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7(3), 424–465.

Van Zandt, T. (2002). Analysis of response time distributions. In J. W. H. Pashler (Ed.), *Stevens' handbook of experimental psychology: Vol. 4. Methodology in experimental psychology* (3rd ed., pp. 461–516): New York: Wiley.

Van Zandt, T., & Townsend, J. T. (1993). Self-terminating versus exhaustive processes in rapid visual and memory search: An evaluative review. *Perception & Psychophysics*, *53*(5), 563–580.

Vollick, D. N. (1994). *Stochastic models of encoding-latency means and variances in paranoid schizophrenia*. Unpublished doctoral dissertation, University of Western Ontario, London, Ontario, Canada.

Wenger, M. J., & Townsend, J. T. (2000). Basic response time tools for studying general processing capacity in attention, perception, and cognition. *Journal of General Psychology*, *127*(1), 67–99.

OGICAL

CAL

NCE

ichard M. McFall

by

otzin • Timothy B. Baker

S

gy Press
Group
ndon