Psychological Review
2004, Vol. 111, No. 4, 1003–1035

0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.4.1003

A Theory of Interactive Parallel Processing: New Capacity Measures and Predictions for a Response Time Inequality Series

James T. Townsend
Indiana University Bloomington and Hanse Wissenschaftskolleg

Michael J. Wenger
Pennsylvania State University

The authors present a theory of stochastic interactive parallel processing with special emphasis on channel interactions and their relation to system capacity. The approach is based both on linear systems theory augmented with stochastic elements and decisional operators and on a metatheory of parallel channels' dependencies that incorporates standard independent and coactive parallel models as special cases. The metatheory is applied to OR and AND experimental paradigms, and the authors establish new theorems relating response time performance in these designs to earlier and novel issues. One notable outcome is the remarkable processing efficiency associated with linear parallel-channel systems that include mutually positive interactions. The results may offer insight into perceptual and cognitive configural–holistic processing systems.

When a person views a work of art in all its complexity, it seems as though all dimensions—color, form, arrangement, perspective, and sharpness of edges—are acting in league. But are they? The antithetical notions of independence and dependence have long played a role in the philosophy and science of human perception and cognition. Whether the focus is on the internal representations that support psychological experience or on the channels or systems that work with those representations, eventually the implications of assumptions about independence or dependence need to be considered.[1]

It is natural to contrast the general notion of independence with ideas such as a holism or a gestalt. However, there have been very few attempts to develop explicit, systematic, quantitative definitions, taxonomies, and explanations of holistic cognitive processes. The work presented here is an attempt to provide these definitions, taxonomies, and explanations, by way of (a) an emphasis on the general characteristics of information processing and (b) a concern with real-time interactions in specific (parallel) information-processing architectures.

It is the case that certain interactive architectures have been shown to be capable of simulating specific gestaltlike phenomena (e.g., Biederman & Kalocsai, 1998; Cottrell, Dailey, Padgett, & Adolphs, 2001; Grossberg, 1991b; Mordkoff & Yantis, 1991; Rumelhart & McClelland, 1981, 1982). Although such studies provide sufficiency arguments, it is not clear what highly simple (in a sense to be defined precisely) and interactive systems can do when compared, within a common framework, with noninteractive systems.

Nevertheless, the great preponderance of quantitative–computational accounts of processing multidimensional stimuli are based on the assumption of independence among channels or items (e.g., Diederich, 1995; Logan, 1988; Raab, 1962; Ratcliff, 1978; Ratcliff & McKoon, 1997; Townsend, 1981; Townsend & Ashby, 1983; Van Zandt, Colonius, & Proctor, 2000). On the basis of a growing body of quantitative theory and methodology, this article develops a new theory of interchannel dependence capable of expressing hypotheses about both positive and negative dependencies as well as independence. Explorations of these hypotheses (presented in detail later in the article) reveal that channel dependencies, positive or negative, hold striking implications for speed of processing, represented formally by notions of capacity (Townsend & Ashby, 1978, 1983). Our theory focuses on response times (RTs) and provides a battery of valuable RT inequalities and a taxonomy of graded capacity and dependence cases for both independent and interactive processing (an excellent overview of modern RT methodology is offered by Van Zandt, 2002).[2]

The theory is developed within the context of two major logical decision structures—OR and AND—corresponding to two types

James T. Townsend, Department of Psychology, Indiana University Bloomington, and Department of Cognitive Neuroscience, Hanse Wissenschaftskolleg, Delmenhorst, Germany; Michael J. Wenger, Department of Psychology, Pennsylvania State University.

This work was supported in part by National Institute of Mental Health Grant 5R01MH57717-02. Portions of this work were completed while James T. Townsend was a fellow at the Hanse Wissenschaftskolleg, and the support of the institute is sincerely appreciated.

Correspondence concerning this article should be addressed to James T. Townsend, Department of Psychology, Indiana University Bloomington, Bloomington, IN 47405. E-mail: jtownsen@indiana.edu

[1] The proper definition of representation engenders controversy to this day (e.g., Van Gelder, 1998). These definitional skirmishes, though important for a field's epistemology, need not waylay us here.

[2] This effort is part of a recent proposal (see, e.g., Wenger & Townsend, 2001) that the stochastic cognitive process theory we have developed over the past several decades provides a natural basis for the exploration and expression of a number of foundational issues (e.g., Townsend & Ashby, 1983; Townsend & Nozawa, 1995). These include capacity and independence, the topics focused on here with regard to correlating channel activity for distinct dimensions or parts of a figure, as well as architecture (e.g., the parallel vs. serial issue) and stopping rule (e.g., when can a search for a target cease), which play a supportive role in the developments. Processing times, with RTs and accuracy as the main dependent variables, have played a major role in the approach. The approach can be thought of as metatheo-

of response instructions for tasks involving multiple elements or dimensions. In an OR design with two targets in a two-item display, the very first target to be completed may stop the ongoing processing. In an AND design, processing of all items in a display must be completed. Both types of decision rule yield valuable information about perceptual and mental processing. The OR designs are important in contrasting coactive processing with independent parallel models (e.g., Egeth & Dagenbach, 1991; Miller, 1982; Mordkoff & Yantis, 1991). The AND designs may be less common in the RT literature but apply when it is necessary for the observer to identify an entire object (i.e., requiring the presence of exactly the proper features or dimensional values).

We use two major quantitative approaches in developing the general theory. The first is based on probability theory, and using this approach, we are able to provide proofs of extremely general theorems. In this approach, the probability distribution can but need not specify a state space that indicates how perceptual information or evidence is accrued. Models in this class can simply place a probability distribution on processing times themselves. This class is entirely general and contains an infinite number of member models.

The second approach uses stochastic dynamic systems theory and simulation methods to assess the properties of different parallel systems. Models developed with this approach do specify a state space indicating how information is accumulating. The dynamic systems are used to explore the properties of various parallel systems. Such models are used less often in cognitive psychology than are models that simply posit a probability distribution on the processing time of each channel (although see Busemeyer & Townsend, 1993; Heath, 2000; Townsend & Ashby, 1983; Usher & McClelland, 2001).[3]

Both the analytic and the simulation approaches play strategic roles in theory development. The analytic approach provides the logical analytic structure and an instrument—a chain of inequalities—for assaying, in an ordinal and stochastic fashion, the speed of processing across OR and AND designs. When an inequality predicted by certain conditions fails, the inequality chain tells in a distribution- and parameter-free way the seriousness and direction (e.g., less or more speed than predicted) of the failure. The analytic approach also allows the development of the concept of capacity and a quantity—the integrated hazard function—to represent stochastic speed at the level of the individual channel (feature, item, etc.). Finally, appropriate functions of time that measure capacity in AND and OR situations assess whether efficiency at the individual channel level has increased or decreased when the number of occupied channels increases and compare them with standard parallel processing speed; all of this is done in a distribution-free fashion. Simulation methods augment the analytic results, and we emphasize that neither of the main approaches could answer the fundamental questions by themselves.

Although the analytic approach is more general, the simulations provide a more intuitive introduction to the major themes under investigation. We therefore begin with a description of the general dynamic tools and use them to instantiate a specific parallel, separate decisions architecture. Next, the class of unlimited-capacity, independent, parallel models are introduced within AND and OR designs. The notion of coactivation as a special type of parallel processing that might underlie configural processing is also introduced. Presentation of central probabilistic expressions and highly important RT inequalities, and then the probabilistic interpretation of coactivation, is treated within these sections.

The next sections[4] offer major theoretical results, many of them new, relating our theory of capacity to the set of RT inequalities for parallel models, in which a decision threshold is included in each channel, as well as for coactive models, which assume that the parallel channels pool their outputs into a single final channel. We also assess whether channel interactions can produce violations of the critical inequalities, for both separate decisions and coactive models. It is concluded that both positive interactions and coactivation can deliver massive improvements in capacity and may be good candidates for modeling the perceptual and cognitive processing of holistic and configural objects. Likewise, negative interactions can lead to extremely limited capacity, even below that associated with standard serial processing.

Theoretical Foundations: The Dynamics of Parallel Processing

The *state space* of a perceptual or cognitive subsystem represents the accumulation of perceptual or cognitive activation of information from the environment or another subsystem. This representation is related to representations of real-time systems in biology, physics, chemistry, and economics as well as a number of important current approaches in psychology. A state space records the state of processing at any given time after the initiation of processing. In contrast, the nonstate approach of stochastic process theory places a probability distribution directly on the finishing times, rather than necessitating the medium of a psychological state space.[5]

One example of state-space models is the class of accumulator models (e.g., Dzhafarov & Bückenholt, 1995; Smith & Vickers, 1989), which generally assume separate decisions across processing channels when more than one channel is operating. Such

retical, as it strives to develop global mathematical properties of psychological models and issues that attend to a broad variety of individual models. These properties can then be used to develop experimental methodologies that can overcome or bypass dilemmas arising from the ability of models to mimic one another in less discriminative experimental designs. Of course, ultimately the metatheory should facilitate a synthesis of representational and process epistemology (O'Toole et al., 2001; R. D. Thomas, 1996; Wenger & Townsend, 2001).

[3] Some well-known models possess a state space, as we show below, but the class of dynamic models to which we refer possesses a rich set of structures and functions based on the state space, such as initial conditions, specification in terms of differential equations, and so on.

[4] We know that readers will come to this article with varying needs and background. Although the article is not designed to be modular, it can be read selectively. For example, although there is, of necessity, fairly rigorous mathematical developments in the body of the article, we have tried to offer the major results in such a way that readers who wish to skip the technical developments can still come to an appreciation of the major results.

[5] Marley (1992) has shown how to convert probabilities on, say, processing times into equivalent statements concerning states. However, the critical facet of the state-space approach is interlocking time and state within the same system in a combined description.

Table 1
Some Important Notation and Variables Used in the Text

Notation and/or variable	Explanation
$x_i(t)$, $i = 1, 2$	Level of activation (state) of each of the two channels
$u_i(t)$, $i = 1, 2$	Level (deterministic) of the input to each of the two channels
$\eta_i(t)$, $i = 1, 2$	Gaussian white noise introduced to each of the channels
$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$	Channel rate and cross-talk parameters: The diagonal elements determine each of the channel accumulation rates, whereas the off-diagonal elements determine the nature and magnitude of channel cross talk
γ_i, $i = 1, 2$	Activation thresholds for each of the channels
$P_i(T_i \leq t)$, $i = \mathrm{A, B}$	CDFs for trials involving a single target
$P_{\mathrm{AB}}(T_i \leq t)$, $i = \mathrm{A, B}$	Marginal CDFs for the processing of a single target, obtained from trials involving both targets
$P_{\mathrm{AB}}(T_{\mathrm{A}} \leq t \text{ OR } T_{\mathrm{B}} \leq t)$	Observed CDF for trials involving both targets when the observer uses an OR decision rule
$P_{\mathrm{AB}}(T_{\mathrm{A}} \leq t \text{ AND } T_{\mathrm{B}} \leq t)$	Observed CDF for trials involving both targets when the observer uses an AND decision rule

Note. CDF = cumulative distribution function.

models also permit either discrete or continuous activation states. A special case of accumulator models is the set of counter models (i.e., discrete state with the states being positive integers; e.g., Townsend & Nozawa, 1995), and a special case of these is Poisson counter models (e.g., Audley & Pike, 1965; Diederich & Colonius, 1991; Pike, 1973; Smith & Zandt, 2000; Townsend & Ashby, 1983; Van Zandt et al., 2000). The other major class of state-space models is that of random walk and diffusion processes, the first being based on discrete and the second on continuous state spaces (e.g., Ashby, 1983; Busemeyer & Townsend, 1993; Link & Heath, 1975; Ratcliff, 1978; Ratcliff, Van Zandt, & McKoon, 1999). Several of the major theorems leading up to this article and all of the new theorems presented here do not require a state-space description: They automatically hold for state-space models as well as for those devoid of a state-space characterization (as in Townsend & Nozawa, 1995). Only the results pertinent to coactivation require a state-space specification.

Within the nonstate-space approach, the theorist is free to arbitrarily select the form and extent of dependency of the component processing times in a way that is autonomous from the other distributional aspects—in particular, the marginal distributions. But the relationships among such quantities emerge as absolutely fundamental in answering basic questions pertaining to interactions and capacity. That is, the nonstate-space approach is, in a sense, too general for certain purposes! In fact, it turns out that there is a trade-off between what happens to the marginal distribution functions and what happens to the joint distribution functions in OR experiments that vitally affects predictions.

The state-space systems we investigate assume a set of accumulator channels that can interact in various ways. Embedded in the theory of dynamic systems, they provide for notions that are unavailable in most other approaches, even a number of other state-space models. These are mentioned in the General Discussion. The present theory also possesses strong linkages with general recognition theory (Ashby & Townsend, 1986). The original general recognition theory (e.g., Ashby & Townsend, 1986; Kadlec, 1992; Kadlec & Townsend, 1992a; Maddox, 1992; R. D. Thomas, 1995, 1996) studied perceptual independence using a static, multidimensional signal-detection approach. Of course, in such approaches, a state becomes simply a multidimensional observation. The current approach is a natural generalization of that theory and, in fact, can be interpreted as an extension of Ashby's (1989, 2000) stochastic general recognition theory.

A global designation for the class of systems within which we work is that of a set of linear channels or filters (e.g., Luenberger, 1979; Townsend & Ashby, 1983, pp. 401–412), with the addition of stochastic noise (e.g., as in Busemeyer & Townsend, 1993; Heath, 2000; Smith, 2000; Usher & McClelland, 2001) and thresholds that implement specific decisional operators. Assuming predecision linearity facilitates the derivation of the relationships holding between channel interactions and capacity as revealed in stochastic processing speed. In addition, this class of systems exhibits considerable flexibility on a number of dimensions, being in fact a member of the sophisticated type of stochastic accumulator system known as a diffusion process (e.g., Busemeyer & Townsend, 1993; Movellan & McClelland, 2001; Ratcliff, 1978; Ratcliff et al., 1999; Smith, 2000).

Basic Specification

We begin the details of our approach by introducing some notation. Table 1 summarizes the most important notation and variables introduced in the sections that follow. Let $x_i(t)$ be the state (the level of activation) of either of two accumulator channels (i.e., $i = 1, 2$)[6] at some time at or after the start of processing. (Note that although we limit ourselves to the case of two parallel channels, the approach is readily extended to deal with n [arbitrary] parallel processing channels; we include example data supporting this assertion at various points below.) We can describe the manner in which the activations in these two channels change over time by way of what one might think of as the prototypical parallel model: a pair of differential equations. Let

[6] Throughout the article, we discuss simulation results, theoretical relations, and empirical regularities that pertain to the parallel processing of two inputs or elements. As a convention, we use the subscripts "1" and "2" when we are discussing our simulation systems and the subscripts "A" and "B" when we are discussing theoretical and empirical regularities.

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

be the vector of channel states, and let

$$\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

be the vector of inputs to these two channels. Let

$$\mathbf{B}(t) = \begin{bmatrix} b_{11}(t) & b_{12}(t) \\ b_{21}(t) & b_{22}(t) \end{bmatrix}$$

be the matrix of coefficients that determines how the inputs in $\mathbf{u}(t)$ are distributed to the two channels. The coefficients b_{11} and b_{22} weight the value of the inputs specific to each channel (e.g., u_1 as input to x_1), and the coefficients b_{12} and b_{21} weight the value of the inputs distributed "across" channels (e.g., u_1 to x_2). Let

$$\mathbf{A}(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix}$$

be the matrix of coefficients that determines how the activations in each of the two channels interact (i.e., how the evidence that is being accumulated in one channel influences the evidence that is being accumulated in the other). This matrix acts to weight and distribute the channel activations as they are being accumulated in the same way that the matrix $\mathbf{B}(t)$ acts to weight and distribute the inputs. For our purposes, we show that the signs of the coefficients $a_{12}(t)$ and $a_{21}(t)$ allow us to instantiate various forms of interactive parallel processing. Given these components, the most general parallel model can then be described, in terms of these matrixes, as

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} b_{11}(t) & b_{12}(t) \\ b_{21}(t) & b_{22}(t) \end{bmatrix}\begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}. \quad (1)$$

Next, we introduce some constraints that simplify our analyses without compromising our objectives. First, we restrict our consideration to systems in which interaction between the channels, if it is present, is present only during the accumulation of evidence. We do not consider systems that allow for very early interactions (as might be appropriate in models of early visual processing). Consequently, we can assume that the off-diagonal elements of $\mathbf{B}(t)$ are equal to 0 and that the weights for the inputs specific to each of the channels are equal to 1.0 and invariant of time, yielding the identity matrix (i.e., $\mathbf{B} = \mathbf{I}$). This simplification allows us to drop $\mathbf{B}$ from the model. Second, we assume that the parameters that weight and distribute the channel activations during processing are constants, invariant of time. This restricts our consideration to a well-understood class of linear systems known as continuous-time linear systems with constant coefficients (e.g., Luenberger, 1979). The approach is further simplified by assuming a symmetry in the rates with which activation grows within each of the channels and is shared across channels. Specifically, we assume that $a_{11} = a_{22}$ and $a_{12} = a_{21}$. Finally, we constrain values of the entries in $\mathbf{A}$ so that the channels are asymptotically stable: That is, neither of the channels are parameterized in such a way as to allow their activations to run off to infinity (or "explode"; see, e.g., the stability parameter in Busemeyer & Townsend, 1993). This assumption has the same effect as the notion of "leaky integrators" in which some of the activation is bled off as it accumulates (see, e.g., Smith, 1995; Usher & McClelland, 2001). This last constraint is important in constructing systems with positive and negative channel interactions. In consideration of all of these simplifying assumptions and constraints, Equation 1 becomes

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{u}(t) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}.$$

When the input vector $\mathbf{u}$ includes randomness (e.g., Gaussian noise), then it is traditional to write $d\mathbf{x}(t) = [\mathbf{A}\mathbf{x}(t) + \mathbf{u}(t)]dt$, to indicate that $(d/dt)\mathbf{x}(t)$ is not a derivative in the usual sense. We ignore this notational nicety—it appears unlikely to cause harm in the present circumstances.

The condition of stability in this system is met by assuming that $a_{11} < 0$ and that always $|a_{11}| > |a_{12}|$ (recall that $a_{11} = a_{22}$ and $a_{12} = a_{21}$ for our systems). The first assumption means that the internal channel feedback is always stabilizing, and the second means that the sum or difference of a_{11} and a_{12} is always negative, whether $a_{12} > 0$, implying positive interaction, or $a_{12} < 0$, implying negative interaction.

To explicate the notion of stability a bit further, set

$$\mathbf{B} = \mathbf{I}$$

as stated above, and consider the simplified input

$$\mathbf{u}(t) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, t > 0.$$

The solution to the consequent differential equation for either channel (as they are the same, because of the symmetry constraint) can be found to be

$$x_1(t) = \frac{1}{a_{11} + a_{12}}\left(e^{(a_{11}+a_{12})t} - 1\right). \quad (2)$$

The key to maintaining stability in this deterministic expression is in the exponential term. Specifically, for the system to be stable as $t \rightarrow \infty$, then $a_{11} + a_{12} < 0$. One important inference from asymptotic stability is that inputs that are limited in their extent (that is, the absolute magnitude is always less than some bound) will inevitably produce outputs that are limited as well (i.e., never become "unstable" by running off to infinity). Another requirement for stability is a limitation on the amount of positive cross talk, because otherwise the effects of $|a_{12}| > |a_{11}|$ would swamp the stabilizing effects produced by $a_{11} < 0$. There is no such limit on the magnitude of negative cross talk, because there would be no limit to how negative $a_{11} + a_{12}$ can be with equal and constant positive inputs to the two channels.[7]

The approach thus far is deterministic. To make contact with the bodies of literature that we seek to integrate, we must make the formulation stochastic. We do this by adding Gaussian white noise, $\eta_i(t)$, to the inputs, which we treat as step functions, that is,

[7] If unequal inputs are presented to the system, then even if $a_{12} < 0$, implying inhibitory (negative) interaction, stability will still require that $|a_{12}| < |a_{11}|$. The reason is that with unequal inputs to the two channels, the channel interaction produces a kind of "sign reversal," which in turn leads to instability, that is, a running off to infinity.

$$u_i(t) = \begin{cases} 0, & t = 0 \\ u_i, & t > 0 \ (i = 1, 2) \end{cases} .$$

Thus, with the addition of the Gaussian white noise to the input, we have

$$\mathbf{u}(t) = \begin{bmatrix} u_1 + \eta_1(t) \\ u_2 + \eta_2(t) \end{bmatrix}.$$

Within the context of this stochastic representation, stability implies that the mean will always be bounded and will approach an asymptote smoothly; specifically, $x(t)$ approaches $-1/(a_{11} + a_{12})$ as t goes to infinity (under our stability assumptions). The variance statistic requires a complicated expression, but simulations indicate that the variances of the channel activations increase to an upper asymptote in very much the same way (but with a different time course) as do the channel activations themselves. An example from the results from the simulation of two independent channels at two levels of noise variance is presented in Figure 1.[8]

The final differential equation for our simplified, stochastic, interactive parallel system is given by

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{u}(t) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} u_1 + \eta_1(t) \\ u_2 + \eta_2(t) \end{bmatrix},$$

with $a_{22} = a_{11}$ and $a_{21} = a_{12}$. Schematic representations of the general parallel model, the simplified and symmetrized interactive parallel model, and the independent channels model are presented in Figure 2.

Various statistics of interactive systems can, with these representations, be compared with predictions from an independent channels system. The independent channels system can be found from the more general system by setting the off-diagonal elements of **A** to 0. Our instantiation of independent parallel processing is thus

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} u_1 + \eta_1(t) \\ u_2 + \eta_2(t) \end{bmatrix}, \quad (3)$$

with $a_{22} = a_{11}$. Even though the individual noise sources are themselves inherently independent, the channels become dependent because of their cross talk or interaction specified by a_{12} ($= a_{21}$, by our symmetry assumption). Therefore, at any point in time, the channels will possess a positive covariance when $a_{12} > 0$ and a negative covariance when $a_{12} < 0$.

Because a primary concern in the results that follow is with comparison of the outputs and finishing times of one versus two channels in OR and AND experimental situations, we need to specialize our assumptions for situations in which only a single channel is operative (e.g., the observer is presented with only one target). When one begins analyzing real-time state-space systems, this topic, perhaps unapparent outside the context of dynamic systems theory, becomes salient and clearly requires attention. In this case, we set $u_i = 0$ (with i indicating the channel that has no input), and we set the diagonal element of the **B** matrix that weights that input to 0. Thus, we assume that when only one of the two channels is presented with an input, the total evidence accumulated is a function solely of what is occurring in that channel. Absent this restriction, the noise that would be present in Channel 2 would be integrated into the accumulating evidence in Channel 1. An alternative, of course, would be to allow cross talk of the noise from the nontarget channel to intrude. However, our numerical simulations indicate that the results we report in this article do not depend on whether the residual cross talk is present in single-target trials.

Decisional Operators

The models generated to this point—independent parallel processing and parallel processing with positive and negative dependencies—implement the stochastic accumulation of task-specific evidence but are incapable of generating any observable responses. To do this, we need some way of determining when enough activation has occurred to trigger a positive detection (identification, recognition, etc.) response in either channel. The natural solution is to insert a pair of decision thresholds on each channel, which we represent as γ_i, $i = 1, 2$. Furthermore, when both channels are involved in the task, a logic gate (i.e., OR or AND) must be imposed after the thresholds are exceeded in each channel. Figure 3 exhibits the separate decisions system, for both independent and dependent processing.

The systems as specified are thus able to represent the stochastic accumulation of evidence in one or two channels and are able to generate responses in the context of specific task logics (i.e., OR or AND tasks). To acquire RT predictions, we need to generate expressions for $P(T_i \le t)$, where T_i is the random variable for the finishing time of each channel, with $i = 1, 2$. The time for any one channel to complete processing is determined by the time required for that channel to accumulate enough evidence to exceed its positive response threshold. Specifically, the completion time for any one channel will be the earliest value of t at which $x_i(t) \ge \gamma_i$. In terms of the distribution of completion times for this channel, this implies that the maximum of $x(t')$ for $t' < t$ will be larger than γ_i or, equivalently, that the minimum time t' such that $x(t') > \gamma_i$ is less than time t. Letting $x_i^*(t) = \max[x_i(t'), \text{ such that } t' < t]$, we can write $P(T_i \le t) = P(x_i^*(t) > \gamma_i)$ for each individual channel.

Furthermore, the joint probability that Channel 1 finishes before time t_1 and Channel 2 before t_2 is simply $P(T_1 \le t_1, T_2 \le t_2)$, and this is the same as saying that the activation for Channel 1 has exceeded its threshold by time t_1 and similarly for Channel 2 completing by time t_2. Likewise, the joint probability that both

[8] Given the well-known ability of chaotic systems to produce stochastic-like phenomena (e.g., Townsend, 1992), it is a continuing temptation to express system randomness in chaotic terms. However, at present, it is not feasible to carry out this program in either a natural or economic fashion. In fact, we are aware of no information-processing models (e.g., ones that are capable of predicting RT) whose stochasticity is founded on chaotic mechanisms, much less in a cogent and meaningful way. Cogency of a chaotic formulation must be considered critical because of the long and successful treatment of chance elements through conceptions of noise and stochastic processes (e.g., Cooper & McGillem, 1999; Davenport & Root, 1958; Papoulis, 1991).

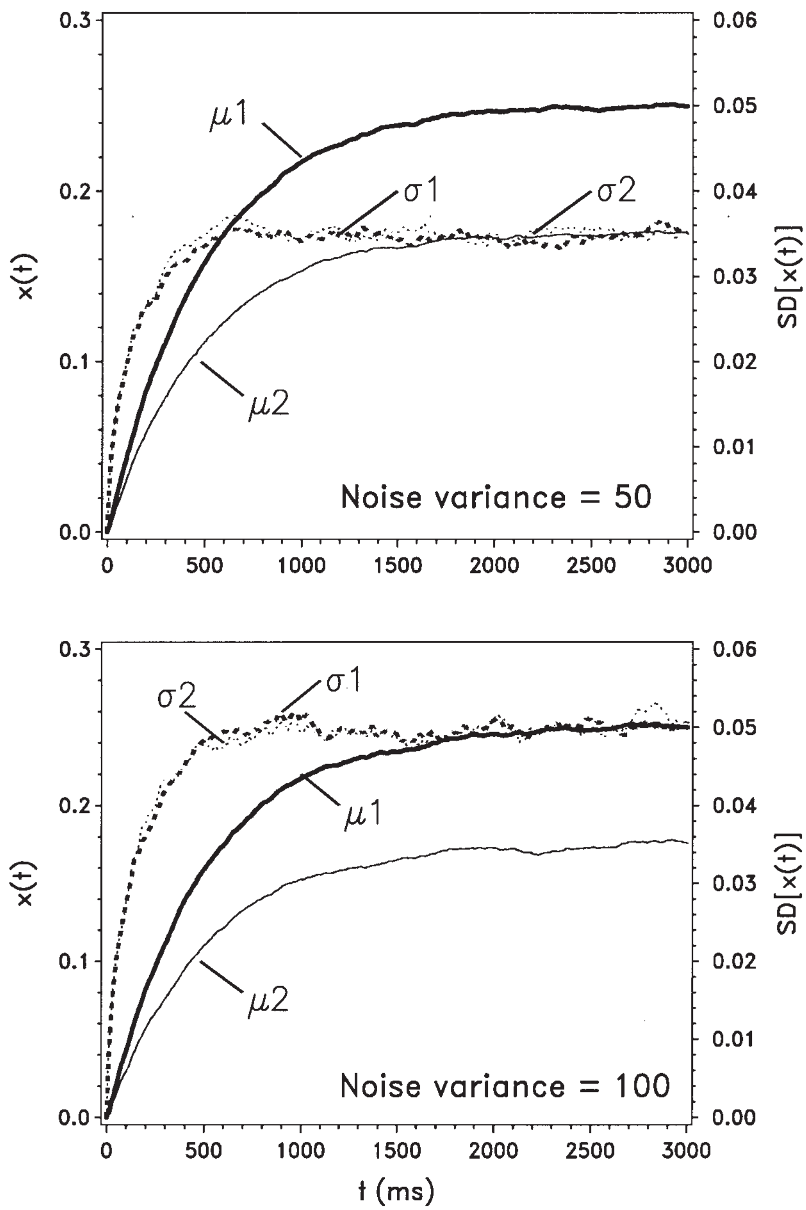

Figure 1. Results of simulating one channel of a two-channel model, assuming independence, two levels for the constant portion of the input, and two levels of variance for the additive noise. Shown are the two mean levels of activation (μ_1 and μ_2) and the standard deviation for each of the two levels of activation (σ_1 and σ_2), for two levels of noise variance (top: 50; bottom: 100), following 1,000 runs of the simulation.

channels are finished by time t is $P(T_1 \leq t, T_2 \leq t)$, just setting t_1 and t_2 equal to t. Such joint probability demands that both events be satisfied and is therefore equivalent to an AND statement. Also, in our AND tasks, we wish to find the probability that both channels finished before or equal to time t. Hence, the latter probability can be found by simply setting $P(T_1 \leq t \text{ AND } T_2 \leq t) = P(T_1 \leq t, T_2 \leq t) = P(x_1^*(t) > \gamma_1 \text{ AND } x_2^*(t) > \gamma_2)$. In the terminology of stochastic processes, the first-passage times of both channels are less than or equal to t. This notation encompasses basically all mathematical processing models in the market place today. A special subclass of these models assumes that activations can only grow (as opposed to growing on the average but possibly decreasing some of the time) across time. In these models, the probability that a channel is finished by time t is identical to the activation being above threshold at time t: $P(T_i \leq t) = P(x_i(t) > \gamma_i)$, but the more complete notation must be used in the general case.[9]

[9] It is worth noting that Dzhafarov (1993) has shown equivalence relations between parallel race models based on random activations with fixed decision thresholds and models based on deterministic activation functions with random decision thresholds.

In the OR situation with redundant targets, if either channel exceeds its decision threshold by time t, the processing activity ceases, and a response can be made. Then the appropriate equation is

$$P(\text{OR RT} \le t) = P(T_1 \le t \text{ OR } T_2 \le t)$$
$$= P(x_1^*(t) > \gamma_1 \text{ OR } x_2^*(t) > \gamma_2). \quad (4)$$

In an AND design, both channels must exceed their thresholds before a response can be made, and hence the expression for RT becomes

$$P(\text{AND RT} \le t) = P(T_1 \le t \text{ AND } T_2 \le t)$$
$$= P(x_1^*(t) > \gamma_1 \text{ AND } x_2^*(t) > \gamma_2). \quad (5)$$

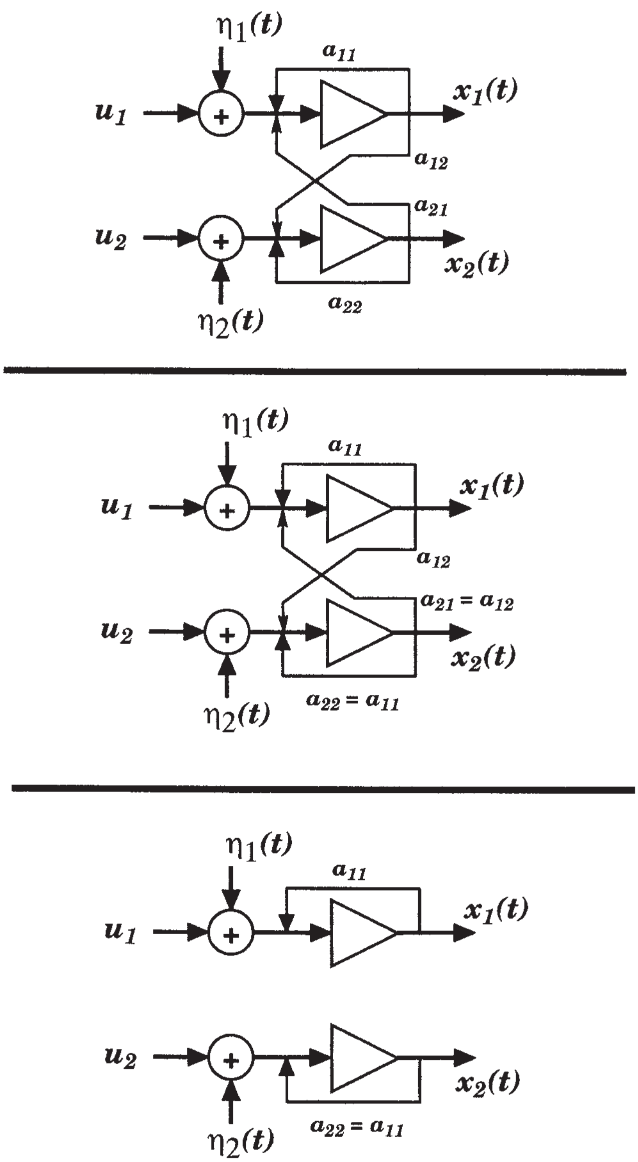

Figure 2. Schematic representations of three versions of the basic parallel-channels models. Top: Unconstrained model, allowing channel interactions. Middle: Symmetrized interactive model. Bottom: Independent channels model. The parameters u_1 and u_2 represent the (deterministic) inputs to the two channels, and $\eta_1(t)$ and $\eta_2(t)$ are the Gaussian white noise inputs.

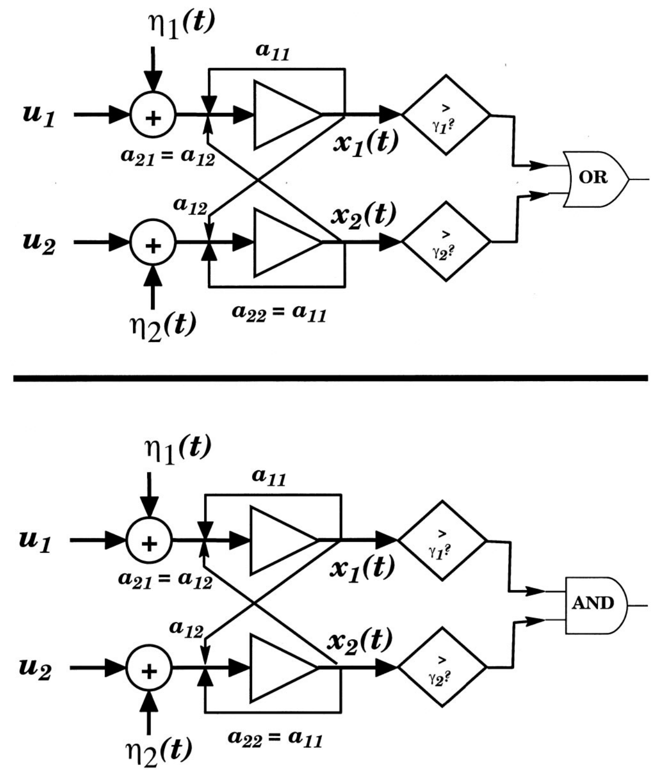

Figure 3. Schematic representation of the independent parallel-channels model, augmented with activation thresholds (γ_i, $i = 1, 2$) and decisional operators (the OR [top] and AND [bottom] gates). Interactive models are created by allowing the cross-talk parameters ($a_{12} = a_{21}$) to be nonzero. The independent channels model is created by setting these parameters to 0. The question marks indicate that the activations are continually being compared with the thresholds.

Thus, it can be seen that positive and negative correlations in activation will be mirrored in RT predictions.

Parallel Processing, Independence, and Capacity

Having described our metatheoretical language and the instantiations of independent and (positively and negatively) dependent systems, we now consider the issues we seek to address with these systems. Our strategy in this section is to discuss the general issues and punctuate that discussion with illustrations of the ability of the models to produce particular empirical regularities. Table 2 summarizes the set of theoretical and empirical regularities and relationships that we consider.

The literature on information processing has been somewhat confused with regard to the issue of parallel (simultaneous) processing. On the one hand, the rule of thumb in myriad studies has been to classify any hypothetical system that exhibits an increase in mean RT as a function of workload as serial (one at a time) and only flat mean RT functions of load as parallel (see review and discussions in Townsend & Ashby, 1983). That this inference is in error was established as early as S. Sternberg's (1966) classic study, in which it was indicated that exhaustive parallel processing would be associated with increasing, concave-down, mean RT functions. Furthermore, it has been known for decades that even

Table 2
Major Empirical and Theoretical Regularities and Relationships to Be Considered

Relationship	Description
$P_{AB}(T_A \le t \text{ OR } T_B \le T) = P_{AB}[\min(T_A, T_B) \le t] \le P_A(T_A \le t) + P_B(T_B \le t)$	Miller inequality (bound; e.g., Miller, 1982)
$P_{AB}(T_A \le t \text{ OR } T_B \le T) = P_A(T_A \le t) + P_B(T_B \le t) - P_A(T_A \le t)P_B(T_B \le t)$	Prediction for first-terminating processing, assuming UCIP
$P_{AB}(T_A \le t \text{ OR } T_B \le t) \ge \max[P_A(T_A \le t), P_B(T_B \le t)]$	Grice inequality (bound; Grice, Canham, & Boroughs, 1984; Grice, Canham, & Gwynne, 1984)
$P_A(T_A \le t) + P_B(T_B \le t) - 1 \le P_{AB}(T_A \le t \text{ AND } T_B \le t) \le \min[P_A(T_A \le t), P_B(T_B \le t)]$	Colonius–Vorberg inequalities (bounds; Colonius & Vorberg, 1994)
$C_o(t) = \dfrac{H_{AB}(t)}{H_A(t) + H_B(t)}$	Capacity coefficient, OR task
$C_a(t) = \dfrac{K_A(t) + K_B(t)}{K_{AB}(t)}$	Capacity coefficient, AND task

Note. UCIP = unlimited capacity, independent, parallel.

linear increasing RT functions can be identically mimicked by limited-capacity parallel models (e.g., Townsend, 1971). Likewise flat RT functions, in experimental circumstances in which they are rare but sometimes found (exhaustive processing), imply very strong and restricted types of parallelism. Flat RT functions of load when one target is embedded within $n - 1$ distractors may also indicate a prototypical type of parallel processing when the stopping rule is self-terminating (for detailed reviews of these issues, see Luce, 1986; Townsend, 1974, 1990a; Townsend & Ashby, 1983).[10] In spite of all of this, one continues to witness theoretical errors in the literature (see related discussion in Townsend & Wenger, 2004b), and all of these errors are occurring in the context of assumptions of independence among the component channels.

With regard to channel dependence, it is well known that parallel channels can be stochastically independent, positively dependent, or negatively dependent in their completion times (e.g., Colonius, 1990; Townsend, 1972, 1974). These three possibilities have intuitive appeal with respect to particular issues in the perception of, and memory for, gestalt or well-configured stimuli such as faces. And the empirical need to consider all of these possibilities comes from observations that characteristics of well-formed perceptual objects, such as faces, can either improve or impede performance, depending on the specifics of the stimuli and task (e.g., Czerwinski, Lightfoot, & Shiffrin, 1992; Goldstone, 1998, 2000; Kuehn & Jolicoeur, 1994; Suzuki & Cavanagh, 1995; Wenger & Ingvalson, 2002, 2003).

It is often difficult to know where to start in dealing with violations of independence. A frequent strategy has been to posit a level of performance expected according to some notion of independence (e.g., C. W. Eriksen & Spencer, 1969; Grice, Canham, & Boroughs, 1984; Grice, Canham, & Gwynne, 1984) and then experiment to see if that level is exceeded in the case of facilitation (e.g., Raab, 1962) or, in fewer cases, to learn if that level is not reached (e.g., Grice, Boroughs, & Canham, 1984). An alternative approach has been to propose detailed process models to explain phenomena that clearly point to nonindependent behavior (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1999; Cottrell et al., 2001; Rumelhart & McClelland, 1981). In addition, as noted, independence is not always amenable to open examination by itself. That is, it is typical that other assumptions—specifically regarding architecture, stopping rule, and capacity—are required to be made. Frequently, these assumptions are not made explicit, and each may well be in error. Capacity issues are most often neglected in this regard.

In the immense literature occupied with such questions, one experimental strategy stands out: one in which comparisons are made across conditions that vary in the number of objects that must be processed, with these objects being often, but not always, of the same type. We concentrate on the prevalent *one versus two* case, although the concepts immediately generalize to a higher number of objects, and (as noted earlier) our focus is confined to RT.[11]

Some additional notation is necessary to develop the arguments. Let A and B be two physical or informational entities, such as target characters in one of two positions, two different anatomical elements of a face, or two sources of cue information for memory. Assume that these two entities can be presented to an observer singly or in combination and that an experimenter wishes to compare RTs when both entities are present—call this T_{AB}—with some combination of the RTs when only one of the two entities is present—call these T_A and T_B. How can the processing times from these three conditions (T_A, T_B, and T_{AB}) be used to determine whether the processing that led to the observable latencies was independent with respect to the two components?

Just as in the case of accuracy (see Ashby & Townsend, 1986), two major distinct subclasses of questions concern across-object

[10] In fact, this prototypical type of parallel processing is based on unlimited capacity (loosely, rates do not change as load is increased), stochastically independent processing time random variables, and self-termination. This model provides the linchpin for capacity measurement in the theory developed below.

[11] For a theory and combined methodology focusing on accuracy, see some of the more recent contributions to the literature associated with general recognition theory (e.g., Ashby & Townsend, 1986; Kadlec, 1999; Kadlec & Hicks, 1998; Kadlec & Townsend, 1992a, 1992b; Perrin, 1992; R. D. Thomas, 1996; see also Movellan & McClelland, 2001, which is pertinent). That line of effort has opened the way for testing various kinds of independence in accuracy contexts but has awaited investigation of behavior in the presence of experimentally stimulated interaction. Kadlec (1999) began work in the latter direction (see also Amazeen, 1999).

invariance and within-object independence. A particular version of within-object independence refers to whether, on a single trial, the processing of two components are stochastically independent of one another. This kind of situation is represented within our systems metatheory as two parallel channels lacking all forms of information exchange (see, e.g., Figures 2 and 3, with the cross-talk parameters set to 0).

This notion of independence is interpreted by way of multiplicativity of the separate probabilities associated with the processing times of the two components. The probability of observing particular values for RTs in an experimental condition is given by the cumulative distribution function (CDF) of the processing times. For example, the CDF for the processing times for Input 1 would be $P(T_1 \leq t)$—the probability that the processing of Element 1 is complete at or before some time *t.* Within our linear systems approach, we can express this CDF in terms of a distribution on activations in the channel responsible for processing this input; specifically,

$$P(T_1 \leq t) = P(x_1^*(t) \geq \gamma_1).$$

Analogous quantities can be defined for the processing of Element 2 and for the joint probability $P(T_1 \leq t, T_2 \leq t)$. Consider, then, from the definition of stochastic independence,

$$P(T_1 \leq t, T_2 \leq t) = P(T_1 \leq t)P(T_2 \leq t). \quad (6)$$

In terms of our linear systems, this definition can be immediately rephrased in terms of accumulated evidence in the two channels. Specifically, if the processing in the two channels is stochastically independent, then

$$P(x_1^*(t) > \gamma_1, x_2^*(t) > \gamma_2) = P(x_1^*(t) > \gamma_1)P(x_2^*(t) > \gamma_2). \quad (7)$$

This within-object type of independence has often been targeted for investigation in experiments in psychophysics and sensory psychology. It is of central importance at many levels of cognition for several reasons. For example, as noted earlier, in the sense of a continuum linking separate and unitary (or configural) processing, the ultimate separateness might entail stochastic independence, whereas a complete unity might entail a maximal degree of positive stochastic dependence. In addition, a number of influential theories of cognition (e.g., Logan, 1988; Massaro, 1998) are based, at some level, on assumptions of independence in processing (see a discussion specific to one particular domain in Wenger, 1999).

The preservation or violation of stochastic independence can be considered as logically independent of a set of foundational issues in processing, including architecture, stopping rule, and capacity (see also Townsend & Wenger, 2004b). The issue of the stopping rule has already been introduced in the guise of the type of logic gate to which the outputs of the processing channels are passed. One of the two major classes is exhaustive processing, in which processing must be completed on all present items or channels before a response can be generated. The other is self-termination, in which it is assumed that whenever sufficient information is processed to accord a response, processing ceases. As noted above, self-termination also refers to the special case in which there is exactly one target among $n - 1$ distractors (e.g., Atkinson, Holmgren, & Juola, 1969; Sternberg, 1966) but has grown to often refer to the general case in which more than one target may be present. The opposite of exhaustive processing happens when the statistic under observation is based on a minimum time or what Colonius and Vorberg (1994) call a "first-terminating stopping rule" (p. 37). The latter case arises when all items in the stimulus display are targets (e.g., Egeth, 1966).

Capacity is perhaps the subtlest issue and, because it turns out to be vital with regard to questions of independence, must receive a bit more attention. This term has long been a theme of importance, and several reviews have been written that survey and discuss theoretical and quantitative approaches to the topic (e.g., Kantowitz, 1985; Schweickert & Boggs, 1984). In addition, many aspects of attention cannot help but intersect the notion of capacity (e.g., Shiffrin, 1976; Shiffrin & Gardner, 1972; Wenger & Gibson, 2004). Within our metatheory of processing systems (cognitive stochastic processing theory, e.g., Townsend & Ashby, 1978, 1983; Townsend & Nozawa, 1995), capacity makes reference to effects on processing efficiency for a single object (component, dimension, etc.) as the number of objects requiring processing (that is, the load) is manipulated.[12] Thus, unlimited capacity is defined by the condition that the marginal probability distribution on individual item processing times is unaffected by the number of other items undergoing processing. At a more microlevel, this implies that the hazard function associated with the marginal probability distribution of an item is invariant over values of $n > 1$. Degradation of performance (implying limited capacity) as n grows can be evidenced in several ways. For example, if the hazard function is smaller for all t for a larger n, then the distribution function is also, and the mean RT for that item increases (Townsend, 1990b).

Once capacity on individual objects is established as a function of n, the question may be asked as to how speed or processing time, as manifested in the mean or other statistics of the RT distribution, may change as n is varied. Assume that the processing time of each item is a random variable with nonzero variance. Then the mean RT for the exhaustive processing of all available objects increases with n in a negatively accelerated fashion, when processing is parallel, stochastically independent across channels, and of unlimited capacity (i.e., the probability distribution on processing times on any single object is invariant regardless of n; Townsend, 1974; Townsend & Ashby, 1978, 1983, chap. 4). Thus, exhaustive processing can produce decreasing performance (increasing RTs), even though processing speed, specified by the probability distribution on individual processing times, at the level of the individual object is constant and therefore "unlimited." The degradation in overall speed for exhaustive parallel processing (even with unlimited capacity and stochastic independence) is caused by what we refer to as *statistical debilitation:* If processing were deterministic (i.e., the processing-time random variables are constants possessing zero variance), mean exhaustive processing time would be flat as a function of n.

In contrast, first-terminating processing can lead to a kind of *statistical facilitation.* Here, with unlimited capacity at the individual item level, independent parallel processing leads to a diminution of the mean processing time as a function of n. In fact, both effects—statistical facilitation with first-terminating, independent, unlimited-capacity, parallel processing and statistical debilitation with exhaustive, independent, unlimited-capacity, parallel process-

[12] See Busey and Townsend (2001) for discussion of other forms of capacity.

ing—order the distribution functions of the exhaustive or first-terminating processing-time random variables, respectively, as the number of channels, *n,* grows (e.g., Townsend & Ashby, 1983; Townsend & Nozawa, 1995).

A reviewer commented that truly unlimited capacity is likely to be rare in human performance. We agree. However, there are circumstances in which systems exhibit unlimited capacity up to some limit (e.g., Fisher, 1984; see also Kahneman, 1973). In addition, many experiments testing for parallelism or independence implicitly assume unlimited capacity (e.g., C. W. Eriksen & Spencer, 1969; Hughes & Townsend, 1998; Shiffrin & Gardner, 1972). The unlimited-capacity model is, in fact, the *modal model* in psychophysics, underlying primary hypotheses. Another major reason for our present quest is the possibility of super-capacity (even better than unlimited capacity) processing with certain stimuli (e.g., gestalt or configural forms) and decision rules (e.g., coactivation, per Miller, 1982, see discussion below).

We should emphasize that both exhaustive and first-terminating parallel systems typically assume stochastic independence: The facilitation or degradation is purely a product of statistical effects. In contrast, in systems endowed with unlimited capacity at the individual object level, positive dependencies among the channels devoted to processing the individual items can produce higher performance at some other levels of analysis (e.g., overall system capacity) than would be expected when processing is stochastically independent (e.g., Townsend, 1974; Townsend & Ashby, 1983). Stochastic dependence may offer a potential mechanism for the superior levels of performance associated with gestalt or holistic perception or cognition. In the extreme case of perfect correlation of all parts, all would finish processing (i.e., reach threshold) at the same time (e.g., Czerwinski et al., 1992; Farah, Wilson, Drain, & Tanaka, 1998; Goldstone, 2000; Kadlec, 1999; Logan, Taylor, & Etherton, 1996; Reicher, 1969; Rumelhart & McClelland, 1981; Wheeler, 1970). Although very little of a systematic nature is known about these issues, the analyses we present in this article provide some intriguing initial results.

An additional reason for considering the possibility of super-capacity processing comes from the notion of unity as it is associated with the idea of a neuron or, more realistically, a set of neurons that acts as a unit to represent (or provide a code for) the object under consideration. In this conception, the system possesses a single "final" channel into which all preceding channels pour their output. A schematic representation of such a system, itself a modification on the system presented in Figure 3, is presented in Figure 4. We refer to this general notion as *coactivation,* a term originated by Miller (1982) to characterize performance that exceeded a certain inequality, to be precisely defined later.[13]

All of these notions can be brought within the scope of experiments comparing performance with one versus two inputs. There are two main subdivisions of the one versus two type of design, although other logical possibilities could be readily defined. We designate one as the OR class of paradigms and the other as the AND class of paradigms, and these correspond to the two types of logical gates that we can place on the outputs of the processing channels (as in Figure 3). Both have substantial histories, which we cannot detail here, with the OR class being emphasized more in recent years (see discussion in Townsend & Nozawa, 1995).

The prototypical prediction of the unlimited-capacity, independent, parallel model with first termination in OR experiments is often called probability summation, and the model itself is referred to as the *race* or *independent race* model. Because we are investigating exhaustive processing as well as first-terminating processing, we require a more general name. In the interest of brevity, and at the risk of memory loss, we use the acronym *UCIP* (unlimited capacity, independent, parallel), with the assumption that the appropriate stopping rule is used in AND and OR experimental paradigms. Although UCIP processing actually refers to a class of models, depending on exact probability distributions, we refer to the UCIP "model" for simplicity.[14]

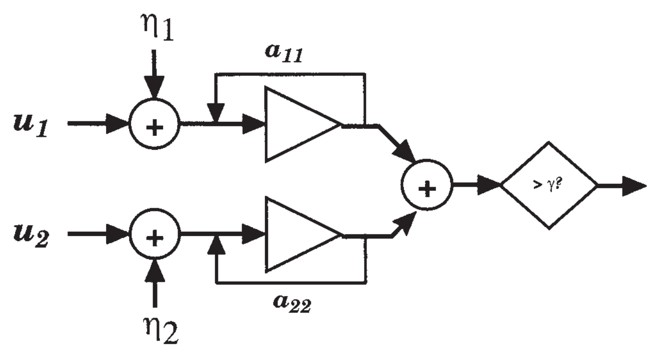

Figure 4. A schematic representation of coactive processing, in which the outputs from each of the parallel processing channels are pooled in a final, single output channel. The activation in this output channel is compared with a single activation criterion (γ). The question mark indicates that the activation is continually being compared with the threshold.

Our specific goals in the remainder of the article are to use our theory to investigate the effects of positive (facilitatory) and negative (inhibitory) channel interactions on the processing of one versus two inputs in AND and OR designs. Given our earlier comments about the possibly tight relationship between channel dependencies and processing capacity, this mission becomes a set of interlocked questions about capacity effects, including those caused by statistical facilitation or debilitation, those caused by channel interactions, and those caused by the actual increase in processing load (e.g., going from one to two targets).

Behavior of Parallel Systems

For both the OR and AND designs, we describe the basic characteristics of the experimental setting, then describe the regularities in RTs that have been documented, and show how systems constructed using our metatheoretical systems language can produce outcomes in parallel systems that are consistent with these regularities. A word is in order regarding errors in relation to RT. The present theory and applications address experiments in which performance is almost

[13] In fact, there are circumstances in which such a possibility does not appear to require unitization (i.e., learning to process an object as a unity rather than as a set of components or features) or a gestalt type of processing. Data exist in which quite different percepts, for instance, signals from the visual and auditory modalities, appear to be processed so fast that something like coactivation is called for (e.g., Diederich & Colonius, 1991).

[14] For an application of UCIP concepts to accuracy in whole-report accuracy experiments, see Busey and Townsend (2001).

perfect (though we consider the robustness of the general effects presented here under variations in accuracy in a later section). The relationship of very low error rates (e.g., less than 10%) to RT has long been of interest to psychologists (e.g., Pachella, 1974). Valuable approaches have appeared that manipulate the time or state of processing and measure the ways in which accuracy and other facets vary as a function of these manipulations (e.g., Dosher, 1979, 1984; Kantowitz, 1978; Pachella, 1974; Wickelgren, 1977). In addition, a number of sophisticated models that represent both RT and accuracy on a within-trial dynamic basis have emerged over the last few decades (e.g., see summaries in Busemeyer & Townsend, 1993; Luce, 1986; Smith, 2000; Townsend & Ashby, 1983).

Few of these approaches have, in our opinion, cast any serious doubt on the rather sizeable knowledge base that continues to accumulate, relating low-error RT to perceptual and cognitive processes. The focus in the present investigation is naturally on the correct “yes” responses in OR or AND designs. We are not concerned here with correct or incorrect “no” responses or with incorrect “yes” responses. For us to produce such responses, it is necessary to consider a variety of modifications to the approach outlined so far. We present results from simulations incorporating one of these modifications in a later section.

We start with the experimental paradigm that has received the bulk of the attention in the last two decades: the OR design. As noted earlier, experiments that implement this paradigm are often referred to as redundant targets experiments (e.g., Bernstein, 1970; Egeth & Dagenbach, 1991; Mordkoff & Egeth, 1993; Mordkoff & Yantis, 1991; Townsend & Nozawa, 1995) because the observer is presented with zero, one, or two targets and is instructed to give a positive response as soon as either or both targets are detected. For this paradigm (and for the AND design to be described later), there have been two main varieties in recent years. The most popular defines a target for each channel, and either channel may be presented with a target or with nothing. Examples from previous work include simple light stimuli that can be present or absent (Townsend & Nozawa, 1995) and the eyes and mouth of a human face that also can be present or absent (Wenger & Townsend, 2001). The second type of OR design defines a positive target (e.g., the letter *X*) and a negative distractor (e.g., the letter *O;* e.g., Egeth & Dagenbach, 1991; Ingvalson & Wenger, in press; Townsend & Wenger, 1999; Wenger & Townsend, 2000b). In both types of designs, the observer is instructed to respond affirmatively if either channel contains a positive target.

The first design has the advantage of being able to assume that when nothing is presented, it neither detracts from nor contributes to the overall RT. This situation allows a better assay of capacity effects in going from one to two targets. The second type of design may be preferable in some cases because on negative trials, there is really something being processed, and thus there may be some symmetry between double-target trials and double-distractor trials that permits better overall model testing (e.g., see Egeth & Dagenbach, 1991; Townsend & Nozawa, 1995).

Stochastic Regularities of OR Processing in Parallel Models

One of the fundamental regularities associated with the OR design was described by Miller (e.g., 1982), who pointed out that the first-terminating RT distribution should be less than (or equal to) the sum of the single target probability distribution functions. Stated in equation form, this assertion, known in the RT literature as *Miller’s inequality,*[15] is

$$P_{AB}(T_A \leq t \text{ OR } T_B \leq t) = P_{AB}[\min(T_A, T_B) \leq t] \leq P_A(T_A \leq t) + P_B(T_B \leq t). \quad (8)$$

It was later noticed (e.g., Ashby & Townsend, 1986; Luce, 1986) that this prediction depends on a critical assumption, one known as context invariance: specifically, the assumption that the time required to process any one of the targets is invariant of what is happening in the other channel. To understand context invariance, let $P_A(T_A \leq t)$ and $P_B(T_B \leq t)$ be the CDFs for processing each of the two targets alone. Then, let $P_{AB}(T_A \leq t)$ and $P_{AB}(T_B \leq t)$ be the marginal CDFs for processing each of the two targets in the context of the other being present. If context invariance holds, then $P_A(T_A \leq t) = P_{AB}(T_A \leq t)$ and $P_B(T_B \leq t) = P_{AB}(T_B \leq t)$. It is of the utmost importance to discern that these equations need not hold and generally will not hold in many experimental situations. If context invariance is in force, it implies unlimited capacity at the level of the marginal distribution (e.g., Colonius & Vorberg, 1994; Townsend & Nozawa, 1997).[16]

It is also critical to observe that context invariance is logically distinct from the notion of stochastic independence. The latter means that

$$P_{AB}(T_A \leq t, T_B \leq t) = P_{AB}(T_A \leq t)P_{AB}(T_B \leq t).$$

However, this definition does not require that $P_{AB}(T_A \leq t) = P_A(T_A \leq t)$, and similarly for the other term. In addition, it can be the case that $P_{AB}(T_A \leq t) = P_A(T_A \leq t)$, implying context invariance, and yet $P_{AB}(T_A \leq t, T_B \leq t) \neq P_{AB}(T_A \leq t)P_{AB}(T_B \leq t)$. That is, context invariance does not imply stochastic independence. Thus, the two conditions are logically independent. Nonetheless, a major part of our study is involved with showing how violation of stochastic independence in real-time systems often leads to a failure of context invariance. In actual experiments, we know no way to ensure that context invariance is satisfied, anymore than we could guarantee stochastic independence: These characteristics will simply depend on the characteristics of the underlying processing system relative to the input and collateral aspects of the experiment.

Consider now why the assumption of context invariance is critical to the prediction in Equation 8. The full prediction for any separate decisions, parallel, first-terminating time model (even with dependencies) can be written as

$$P_{AB}(T_A \leq t \text{ OR } T_B \leq t) = P_{AB}(T_A \leq t) + P_{AB}(T_B \leq t) - P_{AB}(T_A \leq t, T_B < t). \quad (9)$$

Observe that all subscripts, including those on the marginal distributions, indicate the presence of targets in both channels. Hence, a sufficient condition for the inequality in Equation 8 is that $P_{AB}(T_i \leq t) \leq P_i(T_i \leq t)$ for $i = $ A, B. Equality of these terms is equivalent

[15] In probability theory, it is well known as Boole’s inequality (e.g., Parzen, 1960).

[16] Colonius (1990) called this assumption “context independence,” but the numerous uses of the term *independence,* plus the concern of a reviewer, caused us to alter the terminology.

to context invariance. Again, if we expect, at most, limited capacity in going from one to two targets in the channels, this inequality as well as Miller's inequality is still implied, and hence this assumption is theoretically and methodologically harmless.

If instead $P_{AB}(T_i \leq t) > P_i(T_i \leq t)$, the inequality may be violated. But, that is exactly the kind of thing—namely, a supercapacity effect—that we are interested in discovering, along with the kinds of systems that are capable of producing such an effect. When all of the assumptions of the UCIP model hold, the first-terminating prediction becomes

$$P_{AB}(T_A \leq t \text{ OR } T_B \leq t) = P_A(T_A \leq t) + P_B(T_B \leq t) - P_A(T_A \leq t)P_B(T_B \leq t), \quad (10)$$

where, of course, both context invariance and stochastic independence are in force.

Figure 5 plots the results of numerical simulations of the UCIP model (see Figure 3), in terms of the components of Equation 10. The prediction for OR performance, shown as the middle curve, is the result of the difference between the top and bottom curves. Values of the system parameters used to generate the data are presented in Table 3, and a description of the simulation methods is provided in Appendix A.

Before the advent of Miller's inequality, the predictions of the UCIP model presented in Figure 5 (sometimes without explicitly marking the distinction between context invariance and stochastic independence) were the benchmark for assuming independence in many experiments (e.g., Raab, 1962). Again, defining conditions for UCIP processing are parallelism, stochastic independence, and context invariance, but Miller's inequality demands only context invariance and parallelism, not stochastic independence.

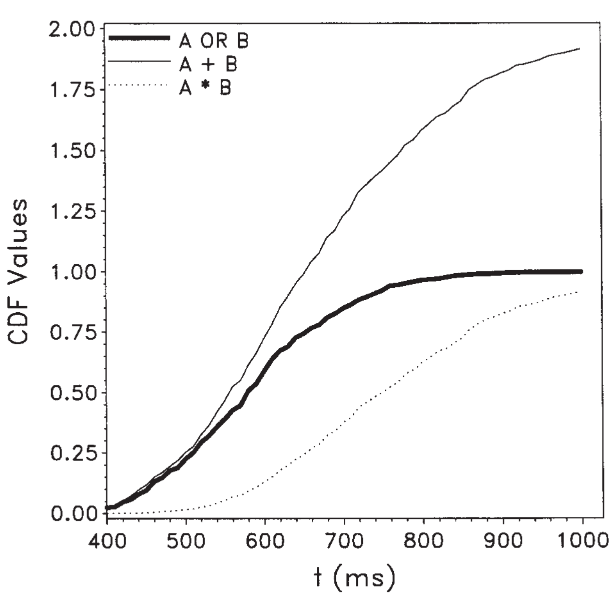

Figure 5. Results of simulating an unlimited-capacity, independent, parallel processing model, operating with an OR response rule. The results are plotted in terms of the components of Equation 10. Note A OR B = $P_{AB}(T_A \leq t \text{ OR } T_B \leq t)$; A + B = $P_A(T_A \leq t) + P_B(T_B \leq t)$; and A ∗ B = $P_A(T_A \leq t)P_B(T_B \leq t)$. CDF = cumulative distribution function.

Table 3
Values of the Parameters Used in the Simulations

Parameter	Value	Explanation
n	10,000	Number of trials per stimulus
u_i, $i = 1, 2$	0.500	Input value when a target is present
a_{ij}, $i = j$	1.000	Channel rate parameters
a_{ij}^{P}, $i \neq j$	0.750	Positive cross talk
a_{ij}^{P*}, $i \neq j$	0.995	Maximal positive cross talk
a_{ij}^{N}, $i \neq j$	−0.750	Negative cross talk
a_{ij}^{N*}, $i \neq j$	−2.500	Maximal negative cross talk
γ_i, $i = 1, 2$	0.250	Channel activation criteria

As is apparent in the form of Equations 9 and 10, Miller's inequality imposes an upper bound on performance in the OR design. There is also a lower bound, which is referred to as the *Grice inequality* (Grice, Canham, & Boroughs, 1984; Grice, Canham, & Gwynne, 1984). Sometimes suggested as an "operational" test of ordinary parallel processing in OR designs, it states that

$$P_{AB}(T_A \leq t \text{ OR } T_B \leq t) \geq \max\,[P_A(T_A \leq t), P_B(T_B \leq t)]. \quad (11)$$

In 1990, Colonius mathematically linked the Miller and Grice inequalities to classical results in probability theory by Fréchet (1951) and others. The basic idea of Fréchet was to learn what the limits of performance are when change is caused by a positive or negative dependence, when the marginal distribution functions are held fixed (i.e., context invariance). Combining the upper and lower bounds represented (respectively) by the Miller and Grice bounds, we get

$$\max\,[P_A(T_A \leq t), P_B(T_B \leq t)] \leq P_{AB}(T_A \leq t \text{ OR } T_B \leq t) = P_{AB}[\min(T_A, T_B) \leq t] \leq P_A(T_A \leq t) + P_B(T_B \leq t). \quad (12)$$

Both ends of this inequality must hold for all times t. However, for the Miller part of the inequality, the interest is in small t because it has to hold (i.e., the right-hand side goes toward 2, whereas the middle term is constrained to be less than or equal to 1) for large t. Figure 6 demonstrates that a model instantiated with our metatheoretical systems language, according to the processing assumptions of the UCIP model, is consistent with these upper and lower bounds. Notice that the OR performance is stipulated by Equation 8.

At this point in our development, it is pertinent to bring up an intriguing dilemma that is of fundamental importance in the overall theoretical puzzle. A natural intuition, and certainly one on which our investigation is based, is that performance improves in the presence of a positive interaction. Paradoxically, Colonius (1990) showed that the upper Miller bound was reached when the correlation was maximally negative, and vice versa in the case of the lower Grice bound. What is going on? The solution to the mystery turns out to depend on the vital concept of context invariance and the consideration of state-based systems as compared with probability distributions that are free to independently manipulate the interactions and the marginal distributions. We see how it all plays out after developing more of the requisite theory.

Stochastic Regularities of AND Processing

Next, consider the predictions for the AND design. If stochastic independence but not necessarily context invariance[17] holds, then

$$P_{AB}(T_A \le t \text{ AND } T_B \le t) = P[\max(T_A, T_B) \le t]$$
$$= P_{AB}(T_A \le t)P_{AB}(T_B \le t),$$

and if both hold, implying the UCIP model, then

$$P_{AB}(T_A \le t \text{ AND } T_B \le t) = P_A(T_A \le t)P_B(T_B \le t).$$

Obviously, it is also possible in principle to have $P_{AB}(T_i \le t) = P_i(T_i \le t)$, i = A, B, even though stochastic independence does not hold. Most studies purporting to test independence in AND designs have, just as in the case of OR designs, overlooked the distinction between channel independence and context invariance. Just as in the OR design, performance in the double-target case may be better than, less than, or equal to that predicted by UCIP processing. In any case, Colonius and Vorberg (1994), extending Colonius's (1990) earlier results, presented inequalities that are analogous to those for the OR situation. For $n = 2$, these are

$$P_A(T_A \le t) + P_B(T_B \le t) - 1 \le P_{AB}(T_A \le t \text{ AND } T_B \le t)$$
$$\le \min[P_A(T_A \le t), P_B(T_B \le t)]. \quad (13)$$

Observe that, just as in the OR case, there is a lower bound for extreme slow performance and an upper bound for very fast performance, relative in both cases to performance in the single-target conditions. Figure 7 shows the results of instantiating the assumptions of the UCIP model, with the channels' outputs subject to an AND gate. As can be seen in the figure, the simulated system produces results that are entirely consistent with the regularities captured in Equation 13.

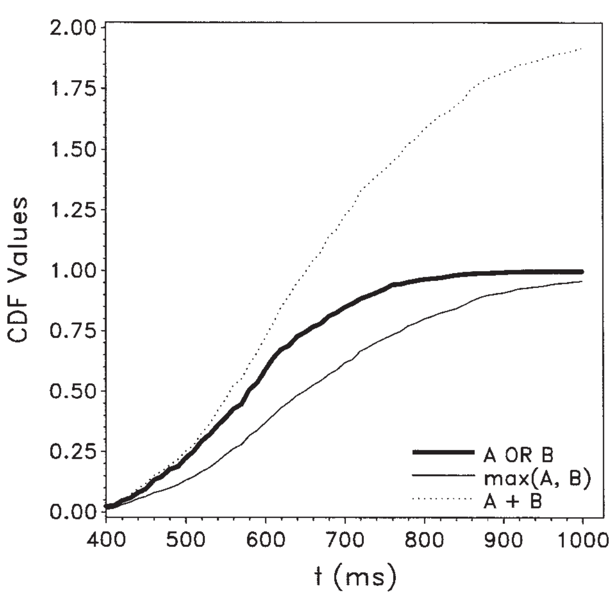

Figure 6. Results of simulating an unlimited-capacity, independent, parallel (UCIP) processing model, operating with an OR response rule. The results are plotted in terms of the components of Equation 12. Note A OR B = $P_{AB}(T_A \le t \text{ OR } T_B \le t)$, which should be interpreted as the prediction of the UCIP processing model; max(A, B) = max $[P_A(T_A \le t), P_B(T_B \le t)]$; and A + B = $P_A(T_A \le t) + P_B(T_B \le t)$. CDF = cumulative distribution function.

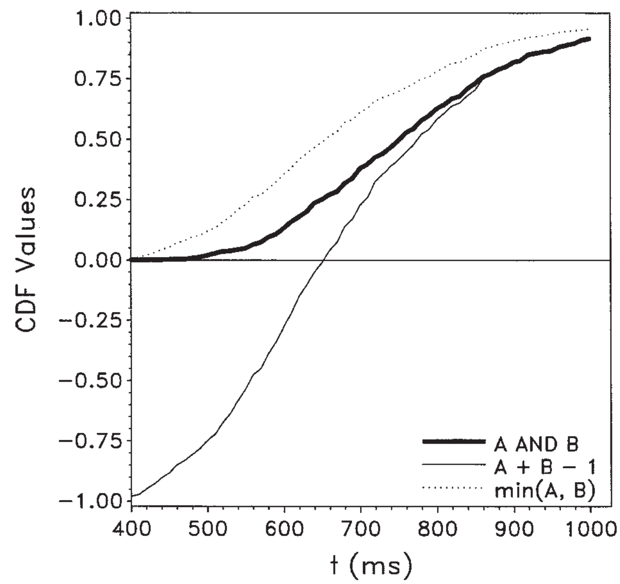

Figure 7. Results of simulating unlimited-capacity, independent, parallel (UCIP) processing with an AND response rule. The results are plotted in terms of the components of Equation 13. Note A AND B = $P_{AB}(T_A \le t, T_B \le t)$, which should be interpreted as the prediction of the UCIP processing model; A + B − 1 = $P_A(T_A \le t) + P_B(T_B \le t) - 1$; and min(A, B) = min $[P_A(T_A \le t), P_B(T_B \le t)]$. CDF = cumulative distribution function.

Aggregating Equations 12 and 13, observing in the meanwhile that $\min[P_A(T_A \le t), P_B(T_B \le t)] \le \max[P_A(T_A \le t), P_B(T_B \le t)]$, we produce an inequality chain that is forced by context invariance in parallel processing. What is more, this chain provides a helpful way to measure capacity in any system in a qualitative, ordinal fashion, through its RTs in OR and AND experiments.

$$P_A(T_A \le t) + P_B(T_B \le t) - 1 \le P_{AB}(T_A \le t \text{ AND } T_B \le t)$$
$$\le \min[P_A(T_A \le t), P_B(T_B \le t)]$$
$$\le \max[P_A(T_A \le t), P_B(T_B \le t)]$$
$$\le P_{AB}(T_A \le t \text{ OR } T_B \le t) \le P_A(T_A \le t) + P_B(T_B \le t).$$

At this point, we reemphasize the strategic role of the chain of inequalities. The chain must be obeyed by ordinary (e.g., not coactive) parallel systems with context invariance. Hence, it serves as a test for that class of models. In particular, the traditional and very important UCIP type of model has to comply with the inequality chain. Nevertheless, the chain also serves a highly valuable function as an ordinal (and therefore model- and distribution-free and nonparametric) kind of measuring scale, which can tell the investigator the coarse degree of any violations and roughly the level of performance. We learn that our finer grained measures of capacity, developed below, provide a

[17] If context invariance does hold, then $P_{AB}(T_A \le t) = P_A(T_A \le t)$, and similarly for the B term.

tight linkage between performance in RT data and the inequality series.[18]

Dynamic Stochastic Representation of Coactivation

Initially, coactivation was defined operationally: Coactivation occurred if Miller's inequality was violated (e.g., Miller, 1982, 1986, 1991). Explicit stochastic models of coactivation were introduced later (e.g., Diederich, 1991, 1995; Diederich & Colonius, 1991; Mordkoff & Egeth, 1993; Mordkoff & Yantis, 1991; Townsend & Nozawa, 1995), with a general theoretical discussion of coactivation and its structural relation to parallel processing offered in Colonius and Townsend (1997). Coactivation models are readily described in terms of a state space, something that can be accomplished by using our metatheoretical language.[19]

Most of the earlier specific coactive models appear to have envisioned parallel channels that pool their activation into a common predecision conduit (e.g., Colonius & Townsend, 1997; Townsend & Nozawa, 1995). That is, the vernacular notion of coactivation involves the summing of the individual channel's activation variables to form the final conduit's activation. The state of this single-output channel is then compared with the decision criterion. Now let us implement this idea using our general metatheoretical approach. For the case of coactivation, the probability that the system has completed processing at an arbitrary time t is equal to the likelihood that the summed activation from the two processing channels exceeds the decision criterion at t, that is,

$$P[T_{(1+2)} \leq t] = P[\max(x_1(t') + x_2(t'),\ t' < t) > \gamma],$$

where the subscript "(1 + 2)" is meant to suggest that the two channels are adding their outputs into a single channel.

In fact, as we suggested earlier, a simple additive connection of the two processing channels transforms our separate decisions, parallel processing model into a coactive model: The two channel outputs are now brought together by summing them in a final common pathway that flows into a single decision criterion comparison mechanism (see Figure 4).

Specific coactivation models have fulfilled theorists' expectations by being able to encompass the high performance levels sometimes found in redundant-targets experiments, in particular the frequent violation of Miller's inequality. Coactivation systems, because of their single output channel, clearly have no facility for making distinct logical decisions. That is, they can separate distinct decision regions along a single activation continuum but have no way to represent AND and OR gates. Hence the performance of a coactive system in AND and OR designs should be quite similar if not identical in the comparisons of single- with double-target yes conditions.

Colonius and Townsend (1997) showed that in terms of the channel activations, coactive models can be considered as a special case of parallel interactive models. Nevertheless, a distinctive feature of coactive models is that a single channel decides whether sufficient information has been acquired for a response to be made. Conventional parallel processing demands separate decisions on the individual channels.

A Capacity Interpretation for Processing Speed

What has been lacking in our discussion so far is an overarching theoretical structure that yields a quantitative and psychologically interesting interpretation of all of the models and inequalities. In particular, there is an absence in the literature of a single theoretical dimension that is absolutely crucial in the violations of the various inequalities. An additional lacuna involves any generic predictions for coactivation models. Both gaps are filled through the concept of capacity.[20]

OR Processing

The notion of capacity, although not always denoted by that term and not always expressed quantitatively, has long been associated with performance differences across conditions involving different numbers of items (e.g., Kahneman, 1973; Kantowitz & Knight, 1976; Schweickert & Boggs, 1984; Townsend, 1972, 1974, 1976; Townsend & Ashby, 1978, 1983; Wenger & Townsend, 2000a).[21] We begin by defining capacity in terms of measures of speed within a stochastic channel.[22] When changes in speed as a function of load are measured, we speak of *load capacity*, and this is the focus of the current discussion. We suggest the term *constant load capacity* when the load (e.g., the number of items to search) is kept constant. One kind of measure of load capacity simply makes reference to a measure of speed, such as mean RT, as n increases. For most stopping rules, it is difficult to observe the individual channel RT distribution and therefore make comparisons with, say, standard parallel processing. Our approach develops an index function (or coefficient) of load capacity that directly compares processing speed for a given stopping rule condition with performance in single-target trials, in such a way that changes in average capacity of individual channels can be assessed. In contrast to load capacity, within-load capacity focuses on the characteristics of speed for a fixed load (e.g., constant n; Busey & Townsend, 2001).

As indicated above, load capacity at the individual channel level can be measured in terms of distinct aspects of the distribution. We have found that a very useful measure of capacity is the integrated hazard function. This precise characterization of capacity leads naturally to development of two coefficients of load capacity:

---

[18] Naturally, the end points of the inequality chain are always satisfied in that $a + b - 1 < a + b$. The important relationships are going on in between.

[19] Miller (1999) suggested a model of coactivation based on the sum of hazard functions of the two separate channels. However, such a sum can always be, and is normally, produced by two independent parallel processes with separate decision thresholds, operating on a first-terminating basis (e.g., Colonius & Townsend, 1997).

[20] Progress is being made with regard to possible neural indicants of process capacity (e.g., Luck & Vogel, 1997). In the future, we look toward the linkage of quantitative concepts such as ours with neural mechanisms.

[21] The original term *capacity* probably derived from its use in information theory (e.g., Shannon, 1948; Shannon & Weaver, 1963), which measured the long-term average of information transmitted by a channel. Our usage takes a different point of departure, although both are concerned with things being accomplished per unit of time.

[22] Our approach to processing speed, and therefore capacity in our terms, has evolved over a number of years, although it has always emphasized parameters, aspects, or functions of the probability distributions on processing times (e.g., Townsend, 1974; Townsend & Ashby, 1978). We have traditionally used capacity in several somewhat different but related ways. For example, the hazard function is a quite precise stochastic measure of

$C_o(t)$, which is specific to OR designs, and $C_a(t)$, which is specific to AND designs (Townsend & Nozawa, 1995, 1997).

The first of these two load capacity coefficients—$C_o(t)$—delivers a benchmark measurement for prototypical but distribution-free UCIP processing in OR experiments (see Townsend & Nozawa, 1995). Use of $C_o(t)$ leads to a pivotal relationship between coactivation, capacity, and violation of Miller's inequality. In this article, the Grice inequality is brought formally into the theory, as expressed in Proposition 1 below.

$C_o(t)$ equals a ratio of logarithm transformations of two quantities: In the numerator is the natural logarithm of the survivor function—the complement of the cumulative distribution function, $S(t) = 1 - F(t) = 1 - P(T \leq t)$—of the double-target RTs, and in the denominator is the sum of the natural logarithms of the two survivor functions—one for each channel—for the single-target RTs. Each of the elements in the numerator and denominator are estimates of the integrals of the hazard function, labeled

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

Specifically, $H(t) = -\ln [S(t)]$, or equivalently, $S(t) = \exp [-H(t)]$. The hazard function itself, denoted by $h(t)$, can be expressed as $h(t) = f(t)/S(t)$, where $f(t)$ is the density function of the attendant distribution. It can be thought of as the function giving the conditional density that processing completes in the next instant, given that it is not finished up until the present (e.g., Luce, 1986; Ross, 1997).

The integrated hazard function is a slightly coarser but probably much more stable measure of capacity than is the more microscopic $h(t)$, where $h(t)$ is analogous to power and $H(t)$ to energy or work done (see Townsend & Ashby, 1978, 1983, pp. 78–79). $H_A(t)$ is then the integrated hazard function when a single target, A, is presented, and similarly for $H_B(t)$. $H_{AB}(t)$ is the comparable function when both targets are present (i.e., a double- or redundant-target trial). At this level of analysis, we can then take advantage of the fact that in the case of independent parallel processing with an OR gate, the integrated hazard function for two inputs will be identical to the sum of the integrated hazards for processing each input separately (see, e.g., Cooper & McGillem, 1999; Townsend & Ashby, 1978, 1983). Pulling all of these identities together for performance in OR designs, we can now define a coefficient of capacity:

$$C_o(t) = \frac{H_{AB}(t)}{H_A(t) + H_B(t)}. \quad (14)$$

speed that bears strong implications for other measures such as the cumulative probability distribution on processing times (Townsend, 1990b). We have also discussed capacity for distributions based on different stopping rules, for instance, first-terminating versus exhaustive processing (e.g., Townsend & Ashby, 1983). That is, if mean RT increased for exhaustive processing, the system was said to be limited capacity at the exhaustive mean RT level. In retrospect, to minimize confusion, we probably should have focused on the individual item or channel level, which we do here, and referred to *performance* or *efficiency* rather than capacity per se when discussing other levels.

This ratio thus involves a direct comparison between performance (expressed in terms of cumulative work completed by time t) in the single-target trials and performance in the double-target trials. If $C_o(t) = 1$ everywhere, the UCIP model is supported because $H_{AB}(t)$ equals the sum of the two separate, single-target integrated hazard functions, which can happen if processing is parallel, unlimited capacity, and independent. For values of t where $C_o(t) < 1$, limited capacity is detected, and when $C_o(t) > 1$, super capacity is the diagnosis.

Townsend and Nozawa (1995) showed that for any times when Miller's inequality is violated, super-capacity processing will be revealed in $C_o(t)$; we reiterate this point in Part 1 of Proposition 1 below. Furthermore, if a system is everywhere super capacity ($C_o(t) > 1$ for all t), then Miller's inequality has to be violated for some t; this result is sharpened in Part 2 of Proposition 1. From these two observations, we see that violation of Miller's inequality demands either slight to moderate super capacity over a long time interval or perhaps more extreme super capacity over shorter intervals. Proposition 1 restates these facts but also includes two new results for the Grice inequality, which is discussed more just below the proposition. The mathematically exact results and proofs are given in Appendix B.

Proposition 1: Miller's and Grice's inequalities.

1. If Miller's inequality is violated at some time t, then the system is super capacity, that is, $C_o(t) > 1$ at that time t.

2. If the system is super capacity for an interval of time near $t = 0$, there is a window of time within that interval where Miller's inequality is violated.

3. If Grice's inequality is violated at some time t, then the system is limited capacity at that time t.

4. Being limited capacity is not itself sufficient to force violation of Grice's inequality. The degree of limited capacity permitted without violation of the Grice inequality is a function of any disparity in speed between Channels A and B.

Note that Part 3 of Proposition 1 is analogous to Part 1, in that any violation of the Grice bound implies limited capacity. However, limited capacity, defined in terms of $C_o(t)$, is insufficient to force a violation of Grice's inequality. To see this clearly, suppose that the speeds of Channels A and B when functioning alone are the same, that is, $H_A(t) = H_B(t)$. Then, capacity can be just a little above fixed capacity (i.e., $H_{AB}(t) = [H_A(t) + H_B(t)]/2$, which is quite drastically limited capacity) without violating Grice's inequality. Fixed capacity with equal distribution means that the overall capacity when both channels are active is equal to the simple (i.e., nonweighted) average of the two alone—that is, the available capacity is, in a sense, being distributed equally among the two channels when both must operate. Observe that a sufficient condition for the Grice bound to be violated is that $H_{AB}(t) \leq \max[H_A(t), H_B(t)]$. Thus, when $H_A(t) = H_B(t)$ in this case, then also obviously $H_{AB}(t) = \max [H_A(t), H_B(t)]$, and $C_o(t)$ will have to be at least lowered to fixed capacity to violate the Grice inequality. Parallel fixed capacity would be identical here to standard serial processing with a first-terminating stopping rule (Townsend &

Ashby, 1983). Furthermore, $P_{AB}(T_A \leq t \text{ OR } T_B \leq t) = \max[P_A(T_A \leq t), P_B(T_B \leq t)]$, indicating that performance is equal to the Grice bound here.

Conversely, when $H_A(t)$ is much larger than $H_B(t)$, then because $H_{AB}(t)$ is bounded above $H_A(t)$, $C_o(t)$ can only be a little lower than 1 (that is, only mildly limited capacity) without leading to violation of the Grice inequality. Thus, if there is great disparity in processing speeds of the two channels, then mild limited capacity can cause a Grice violation. As a special illustrative case, consider the kind of parallel processing in which the hazard functions are proportional to one another. Thus, let $H_B(t) = rH_A(t)$, $r > 0$. Then, it is easy to compute that a Grice violation occurs if $C_o(t) < 1/(1 + r)$.

Capacity predictions for OR processing. Having developed a statistic specific to the notion of capacity, we can now consider the manner in which preservation and violation of channel independence impacts this measure of capacity. We accomplish this by way of models developed with the metatheoretical language we described in the initial portion of this article. In particular, we consider the performance of systems that involve positive and negative dependence in processing and compare both with performance when the processing channels are independent. For each of the systems we consider, the input signal is the same in all cases, and the assumed initial conditions are always 0. As a reminder, Table 3 lists the values of the specific system parameters used to produce the results that follow, and Appendix A describes the simulation methods. In terms of the outputs, we consider (a) the capacity function $C_o(t)$, (b) the Miller and Grice inequalities, and (c) an analysis of the marginal and joint probabilities to better understand how the channel dependencies lead to the overt capacity behavior. The last item provides a solution to the mystery concerning Colonius's (1990) and Fréchet's (1951) results and our intuitions mentioned earlier.

Figure 8 shows $C_o(t)$ for positive (facilitatory) and negative (inhibitory) cross talk. We include in this figure results for two and four channels to illustrate the manner in which our approach "scales up" to larger processing loads. It is apparent that in the case of ordinary parallel processing with channel interactions, $C_o(t)$ varies from a high value of $C_o(t) > 3$, with positive cross talk, to a low value of $C_o(t)$ approximately equal to 0.2, with negative cross talk. Notice that when $a_{ij} = 0$, $i \neq j$, $C_o(t)$ is about equal to 1, as expected. Hence, our fine-grain capacity function, in alliance with Proposition 1, suggests violation of Miller's inequality at the upper end of performance, because capacity is super over a sustained interval of time early in processing.[23] Grice's inequality is predicted to be violated as well in the presence of inhibitory interaction because capacity is less than 0.5, or fixed capacity.[24]

A set of outcomes is also of interest for the simulation of four parallel channels. First, when the channels are independent, the values of $C_o(t)$ are essentially identical to those obtained with two independent channels (see Figure 8, top). Second, when the four channels are positively dependent (see Figure 8, middle) and the magnitude of that dependency (as determined by the cross-talk parameters) is the same as it is in the simulation of two positively dependent channels, the value of $C_o(t)$ increases, as would be expected. Finally, when the four channels are negatively dependent (see Figure 8, bottom), the magnitude of the dependency must be modified so that the magnitude of the sum of the parameters for the cross talk is less than the channel-processing parameter. When

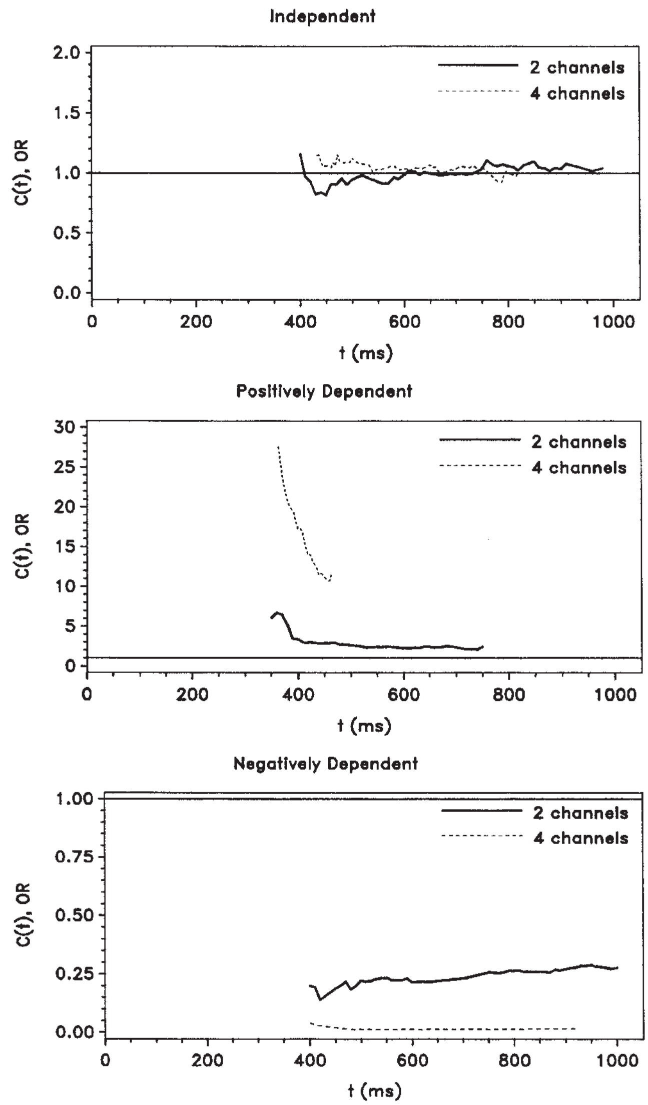

Figure 8. Values of the load capacity coefficient $C_o(t)$ (Equation 14) for independent processing (top) and positive (middle) and negative (bottom) dependent processing, for cases of two and four parallel channels. The reference line at 1 indicates unlimited-capacity processing.

this condition is met, it is still the case that the value of $C_o(t)$ is decreased.

Figure 9 exhibits the behavior of $[P_A(T_A \leq t) + P_B(T_B \leq t)] - P_{AB}(T_A \leq t \text{ OR } T_B \leq t)$ in the top panel and $\max[P_A(T_A \leq t), P_B(T_B \leq t)] - P_{AB}(T_A \leq t \text{ OR } T_B \leq t)$ in the bottom panel. The

[23] The quantitative details of the results cannot say exactly what the magnitude or interval must be to produce a violation, in the case of Propositions 1 and 2. However, these can in principle be calculated for actual single-target data or when one assumes specific processing parameters.

[24] Fixed capacity denotes a boundary between moderately limited capacity and extremely limited capacity (e.g., Townsend & Ashby, 1983; Townsend & Nozawa, 1995).

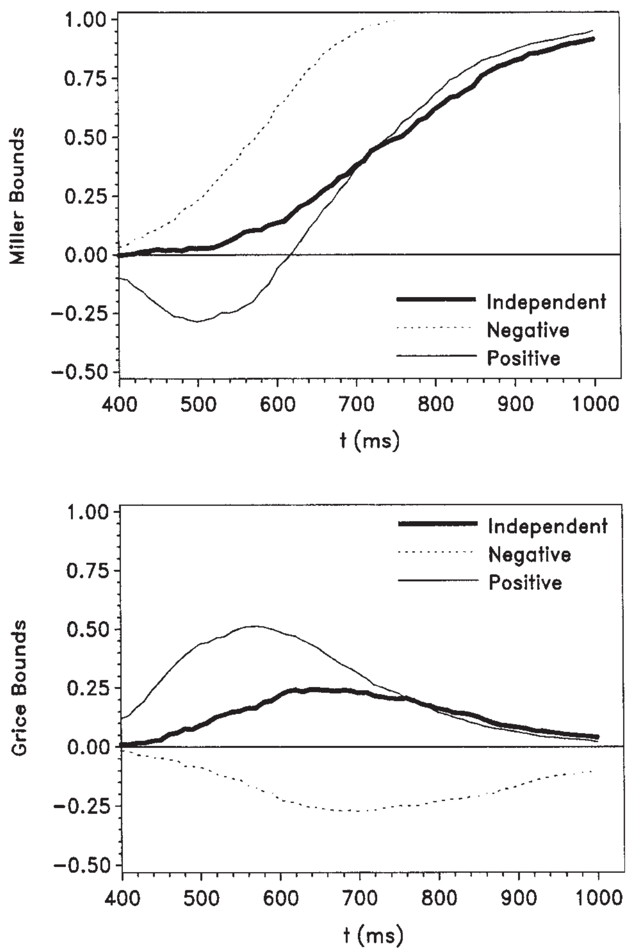

Figure 9. Comparisons against the Miller (top; e.g., Miller, 1982) and Grice (bottom; Grice, Canham, & Boroughs, 1984; Grice, Canham, & Gwynne, 1984) bounds for the two-channel model, under conditions of independent processing and positive and negative dependent processing, in an OR task. Negative values indicate violations of the pertinent bounds.

P_{AB} terms are the double-target entries, whereas the P_A and P_B terms stand for the single-target "results" brought over to make predictions from the upper (Miller) and lower (Grice) bounds, respectively, for the double-target conditions. In this figure and related figures that follow, a violation of either inequality is indicated by a negative value. As expected from the capacity coefficient, no violations are seen when the channels are independent and unlimited capacity (i.e., when $a_{ij} = 0$, $i \neq j$), because then $C_o(t) \cong 1$. When a_{ij}, $i \neq j$, is sizably positive, $C_o(t)$ is greater than 1, and large breaches of the Miller bound are witnessed. Similarly, performance drops below the Grice bound when the off-diagonal elements of **A** are sizably negative, indicating heavy inhibition and causing capacity to drop precipitously.

We are now in a position to solve the apparent paradox involved in the fact that whereas Colonius's (1990) and Fréchet's (1951) results predict that the Miller bound will be reached in the presence of high negative dependence, our intuitions and now the actual dynamic systems results predict reaching and superseding the Miller bound when interactions are highly positive. The solution involves the role played by context invariance. Colonius's and Fréchet's results assume context invariance so that the marginal distributions of the single-target distributions remain unchanged. However, it is possible that natural dynamic systems (e.g., highly interconnected pathways in the visual system) do not obey context invariance and override the kind of effects leading to Colonius's and Fréchet's findings.

A particularly germane aspect, then, of the solution is made clear by comparison of the predictions under context invariance as opposed to what happens in our simulations of real-time systems. Inspection of Equations 9 and 10 reveals that in the OR case, when moving from one target to two and assuming positive dependence, the dependent term subtracts from the marginal distributions, which are constant when context invariance is in force. As observed, the Miller bound, under the condition of context invariance and demarcating extreme super capacity in the OR context, is realized when the RT dependencies are maximally negative rather than positive (Colonius, 1990). This happens because the dependence term disappears, rather than being a positive quantity that is subtracted from the marginals. Hence, at this juncture, we are confronted with the aforementioned paradoxical picture of systems that reach the limits of "ordinary" super capacity (the Miller bound) when the channels are negatively rather than positively dependent.

As anticipated, the missing link in the overall picture of channel interactions lies in the question of what happens to the marginal distributions when channels in a realistic activation system interact in a positive or negative fashion. Do they stay more or less constant, as context invariance demands? Or, with positive interactions, do the marginal distribution terms increase enough to outweigh the detrimental effect of the positive dependency? Conversely, with negative interactions, do we indeed get increased speed through the dominance of the interaction term, or do the marginal distributions decrease so much that an overall inhibitory effect is created? These questions can be seen even more clearly in the identity

$$P_{AB}(T_A \leq t \text{ OR } T_B \leq t) = 1 - P_{AB}(T_A > t, T_B > t),$$

the last term being the quantity involved in Miller's inequality. Here, it is obvious that it is possible that

$$P_{AB}(T_A > t, T_B > t) > P_{AB}(T_A > t)P_{AB}(T_B > t),$$

that is, there is a positive dependence for the condition, and yet we still may have

$$P_{AB}(T_A > t, T_B > t) < P_A(T_A > t)P_B(T_B > t);$$

the probability that neither channel is done by time t is less in the AB condition than in the product of the single-target conditions! Hence, processing would in such instances be faster in the AB condition despite the positive channel dependence when both targets are present. As we now witness, this is exactly what transpired in our simulations.

Figure 10 reveals the critical effects for positive (top) and negative (bottom) interactions through several statistics. First, the difference relevant to the joint probability

$$P_{AB}(T_A \leq t, T_B \leq t) - P_A(T_A \leq t)P_B(T_B \leq t) \qquad (15)$$

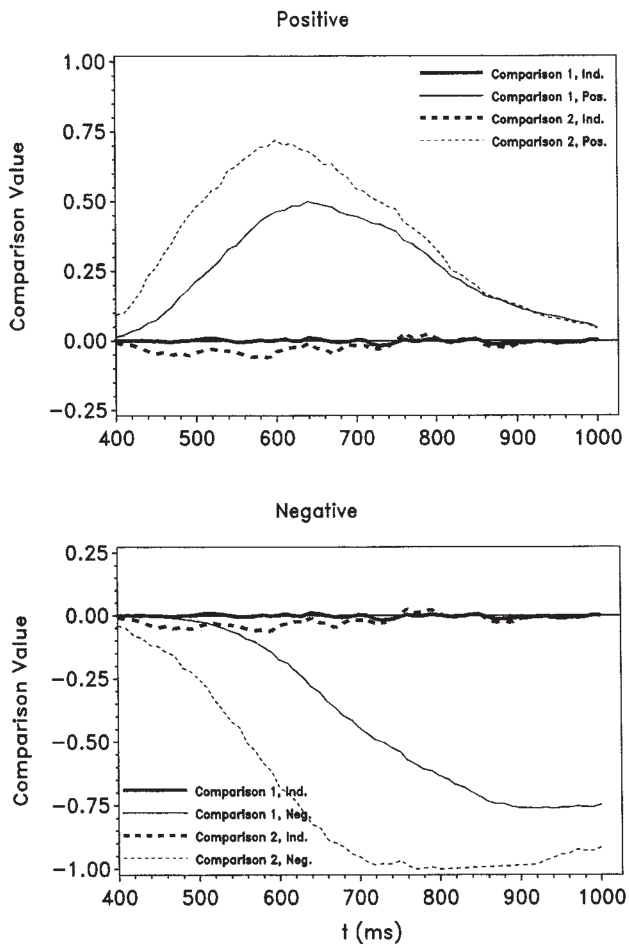

Figure 10. Two comparisons (Equations 15 and 16) involving the marginal distributions, for independent (Ind.) processing and for positive (Pos.) and negative (Neg.) dependent processing. Top: The effects of positive channel interactions contrasted with independent processing. Bottom: The effects of negative channel interactions contrasted with independent processing. See Equations 15 and 16 for details.

is shown and is labeled as "Comparison 1" in the two panels of Figure 10. This comparison assesses the degree of positive and negative dependence as compared with independence in conjunction with context invariance. Recall that the initial term in this comparison is the quantity that is subtracted out in Equations 9 and 10. Thus, the larger the value for Equation 15, the larger the decrement to OR processing (i.e., processing is slowed). The top panel compares independence and positive dependence, whereas the bottom panel compares independence and negative dependence. Notice that in the top panel, the difference is entirely positive, indicating that positive channel dependencies "hurt" OR performance. In contrast, negative dependencies "help" OR performance, as shown in the bottom panel.

Next, we consider the effects of dependence on the marginals in the operative OR equations. The crucial question is whether the positive dependence effect on the joint probability hurts (i.e., lowers the probability) more than potential increasing effects from the marginals help (and, similarly, in the reverse direction for negative dependence). First, observe the difference

$$P_{AB}(T_A \leq t) + P_{AB}(T_B \leq t) - [P_A(T_A \leq t) + P_B(T_B \leq t)], \quad (16)$$

which simplifies to

$$2[P_{AB}(T_A \leq t) - P_A(T_A \leq t)],$$

when the channel characteristics are the same for both channels. We label this as "Comparison 2" in the two panels of Figure 10. What this difference indicates is the summed impact of dependence on the marginal distributions. Given that the marginals are increased, the operative question for positive dependence is whether it can compensate for the hurtful (with regard to speed) effect on the joint probability expressed in the first difference in Figure 10. The reverse issue arises for negative dependence. If this difference is positive at any time *t*, then the contribution of positive interaction more than makes up for the subtractive term through increase of the marginal values. Figure 10 reveals that the facilitatory effect on the marginals in the presence of a positive interaction more than compensates for the inhibitory effect on the joint probability term.

Similarly, it may be observed that the decrease in the marginals when the channels are negatively dependent more than makes up for the decreased subtractive joint probability term. Nevertheless, these benefits (in the case of positive interaction) or detriments (in the case of negative interactions) cannot by themselves assay whether the combined marginal-plus-joint effects are sufficiently strong to force violation of the inequalities: The load capacity function, $C_a(t)$, is required for that.

AND Processing

Even though experimental conditions that demand exhaustive processing have occasionally been used to test parallel independent processing, there has not been the fine-grained structure to measure load capacity in these circumstances. We here devise an analogue to $C_o(t)$, which we call $C_a(t)$, to serve this purpose. First, we need something analogous to the integrated hazard function. We propose a function that corresponds in many ways to $H(t)$. Let $k(t)$ be equal to the density over the distribution function, $k(t) = f(t)/F(t)$. It is analogous to the hazard function $h(t)$ and can be thought of as the conditional probability density that processing completed in just the last instant, given that it completes at or before t.[25]

In addition, it follows that we can write, similarly to the integrated hazard function, $F(t) = \exp\,[K(t)]$, where

[25] The ordinary hazard function has a very long history, and a tremendous amount of information has accumulated concerning its technical properties. A prime intuition for $h(t)$ is that the event has not occurred by time t, and $h(t)$ gives the conditional likelihood that it will happen momentarily. There is a similar logic to $k(t)$. Suppose you are an actuary investigating an accident and you know only that it happened before time t; then $k(t)$ yields the conditional likelihood that in fact, the accident occurred just before that time. Both functions are useful but for slightly different purposes. It appears likely that interesting technical aspects of $k(t)$ analogous to $h(t)$ can be worked out that may aid RT research (cf. E. A. Thomas, 1971). However, whether $k(t)$ will attain the popularity of $h(t)$ remains to be seen.

$$K(t) = \int_0^t k(t')dt'$$

and $K(t) = \ln[F(t)] \le 0$. In analogy to the integrated hazard function definition, $K(t)$ goes to $-\infty$ as t goes to 0, and it has to increase to 0 as t goes to ∞.[26] The AND load capacity coefficient, $C_a(t)$, is then defined as

$$C_a(t) = \frac{K_A(t) + K_B(t)}{K_{AB}(t)}. \tag{17}$$

Note that in this construction, the single-target quantities are in the numerator, unlike the case for $C_o(t)$. This "inversion" results in a measure that has an interpretation that is identical to that of $C_o(t)$. That is, $C_a(t)$ is greater than 1 if performance is superior to UCIP, less than 1 in the case of inferiority to UCIP, and equal to 1 if performance is identical to that of UCIP. Proposition 2 captures the other interesting relationships that are analogous to the ones for OR conditions; Appendix B develops this and further properties relating to $C_a(t)$ more precisely. In the following, we refer to the results of Colonius and Vorberg (1994) as *Colonius–Vorberg bounds* or *inequalities.* Our new results are as follows.

Proposition 2: Colonius and Vorberg (1994).

1. Violation of the Colonius–Vorberg upper bound implies super capacity, that is $C_a(t) > 1$.

2. Being super capacity is insufficient to force violation of the Colonius–Vorberg upper bound. The degree of super capacity permitted without violating the upper bound is a function of the disparity in speed between Channels A and B.

3. Violation of the Colonius–Vorberg lower bound implies limited capacity, that is, $C_a(t) < 1$.

4. If the system is limited capacity for all times exceeding some t, then the Colonius–Vorberg lower bound will be violated for some interval of time that starts at $t' > t$.

Proposition 2 may be interpreted in a straightforward way analogous to that of Proposition 1. Parts 1 and 2 reveal that the upper Colonius–Vorberg bound is quite high, in that its violation implies super capacity but super capacity does not imply its violation. If the speeds of Channels A and B are very different, say $F_B(t)$ is much larger than $F_A(t)$, then if $C_a(t)$ is only a little larger than 1, the upper bound will be exceeded. However, if $F_A(t) = F_B(t)$, then capacity has to be more than double that to push performance past the upper bound. Conversely, Parts 3 and 4 indicate that although violation of the lower bound does imply limited capacity, just having limited capacity for a window of time means that the bound must be violated.

There is an intriguing strong symmetry between Propositions 1 and 2. The symmetries relate (a) Parts 1 and 2 of Proposition 1 to Parts 3 and 4 of Proposition 2 and (b) Parts 3 and 4 of Proposition 1 to Parts 1 and 2 of Proposition 2. In general, violation of Miller's inequality and the lower Colonius–Vorberg bound calls only for mildly super or mildly limited capacity, respectively. However, violation of Grice's lower OR bound or the Colonius–Vorberg upper bound can, depending on channel speed disparity, require very limited or extreme super capacity, respectively. Assume $F_A(t) = F_B(t)$. Then, from Part 2 of Proposition 2, capacity has to be about double what is required for unlimited capacity, that is, double the capacity of the UCIP model, to force violation of the upper Colonius–Vorberg bound. Similarly, capacity has to be less than one-half unlimited capacity to violate the Grice inequality. Naturally, these statements about limited and super capacity must be taken as relative to their respective domains in OR and AND paradigms because the AND bounds are lower than the OR bounds. Moreover, even though both coefficients measure capacity, that they do it on distinct statistics (the minimum vs. maximum processing time) means that $C_o(t)$ being greater than 1, say, does not imply that $C_a(t)$ is greater than 1, and similarly for the other orderings.

Capacity predictions for AND processing. In terms of predictions for capacity, we begin, as we did for the OR task, with the load capacity function $C_a(t)$. Figure 11 displays the results of numerical simulations of the AND task for three situations—independent processing, positive dependency, and negative dependency—for two levels of processing load (two and four channels). In the case of independent processing (i.e., when $a_{ij} = 0$, $i \ne j$), $C_a(t)$ remains very close to 1, exhibiting unlimited capacity. When a_{ij}, $i \ne j$, is positive (for two channels), $C_a(t)$ takes on values considerably greater than 2, indicating the exceptional super capacity sufficient to cause performance to surpass the Colonius–Vorberg upper bound. Finally, when a_{ij}, $i \ne j$, is negative (for two channels), $C_a(t)$ moves toward 0, indicating extremely limited capacity and suggesting that the Colonius–Vorberg lower bound will likely be violated. The basic patterns for the four channel simulations are similar: For the case of independence, $C_a(t)$ hovers around 1; for positive dependence (at the same level of positive cross talk assumed for two channels), capacity is well above 1; and for negative dependence (modulated in magnitude as was done for two channels), capacity is extremely limited.

Performance of AND systems (under conditions of independence and positive and negative dependence) relative to the Colonius–Vorberg bounds is exhibited in Figure 12. As was true for the figures assessing performance relative to the Miller and Grice bounds, negative values indicate violations of the pertinent bounds. It can be seen in this figure that the intuitions developed from $C_a(t)$ are verified: High positive values of a_{ij}, $i \ne j$, produce a violation of the Colonius–Vorberg upper bound, and conversely, large negative values of the cross-talk parameters produce performance below their lower bound.

Analogous questions to those posed for OR situations, with regard to the combinations of channel dependencies and marginal context variability, can be asked with regard to AND conditions. We know from the previous analyses that the overall effects are performance enhancement in the case of positive dependencies and decrements in the case of negative dependencies. However, it is interesting to see in more detail how this came about. The distribution $P[\max(T_A, T_B) \le t]$ in the UCIP case is just $P_A(T_A \le t)P_A(T_B \le t)$ versus $P_{AB}(T_A \le t, T_B \le t)$ in the general two-target

[26] In addition, $k(t)$ is assumed to be continuous (this can be weakened), and $K(t)$ will be strictly monotonic increasing.

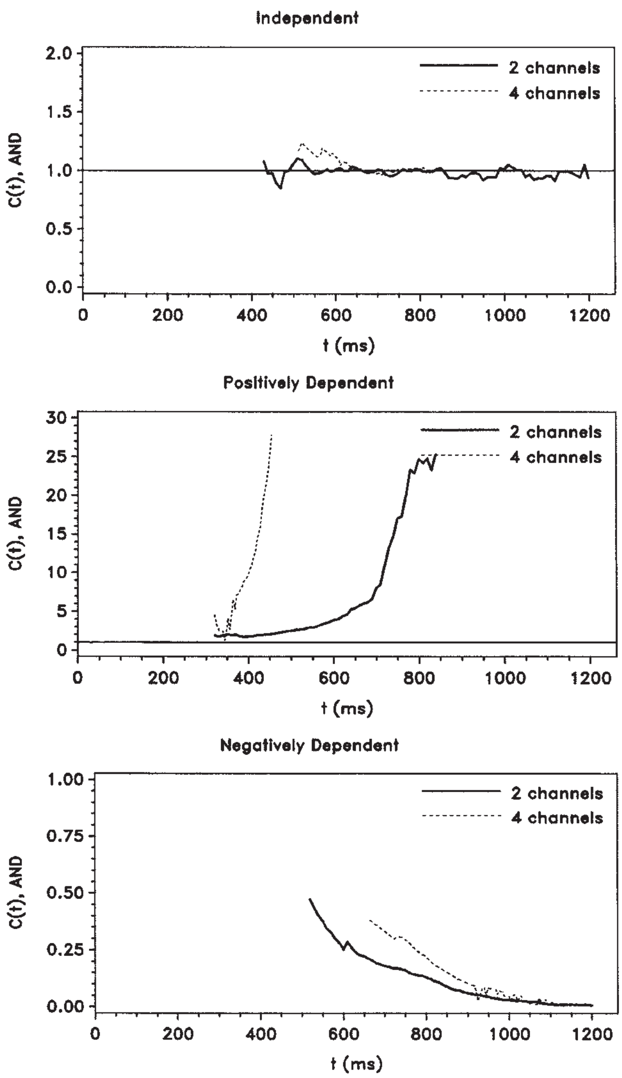

Figure 11. Values of the load capacity coefficient $C_a(t)$ (Equation 17) for independent processing (top) and positive (middle) and negative (bottom) dependent processing, for simulations of two and four channels. The reference line at 1 indicates unlimited-capacity processing.

case, so the dependency term is exactly the pertinent joint distribution function. Nevertheless, a positive dependence alone is not guaranteed to produce superior performance in the AB condition—it depends on the overall capacity effects, just as in the OR case.

Figure 10 can be used to intuit the influence of cross talk on AND performance, even though it was originally constructed for the OR case. It can be seen there that

$$P_{AB}(T_A \leq t, T_B \leq t) \neq P_A(T_A < t)P_B(T_B \leq t)$$

(Comparison 1 in Figure 10), whenever $a_{ij} \neq 0$, $i \neq j$. Because these quantities are the probability of finishing by time t in the AND design, they immediately prove that negative interactions slow down performance and positive interactions speed up performance. However, just as in the OR case, they do not by themselves prove that bounds will be superseded: The more powerful load capacity functions fulfill that role.[27]

Channel Dependencies and the "Wrong" Bounds

As we have discussed in the preceding sections, each of the two experimental designs—OR and AND—have distributional inequalities that provide upper and lower bounds on performance when the systems are consistent with the assumptions of UCIP. However, we have seen that when we introduce violations of independence between the channels, the upper and lower bounds can be violated for both designs. Given that an OR system with negative dependencies can violate the Grice inequality and given that an AND system with positive dependencies can violate the upper Colonius–Vorberg bound, it seems reasonable to inquire about the possibility that the dependencies in these two cases might be sufficient to cause violation of the "wrong" bound. That is, it seems possible that an OR system with great negative dependencies might be able to travel below the lower Colonius–Vorberg bound, even though that bound represents extremely low capacity when $F_A(t) = F_B(t)$. Furthermore, perhaps an AND system with strong positive dependencies might be able to violate the Miller inequality, even though when $F_A(t) = F_B(t)$, the Miller bound is twice UCIP performance with an OR decision gate.

But, we have a specific constraint that we must deal with in parameterizing our models to address this question. Specifically, as we mentioned earlier, we are assuming that all of our models are asymptotically stable. Consequently, the values for our cross-talk parameters, a_{ij}, $i \neq j$, can take on only a restricted range of values relative to each other and to a_{ij}, $i = j$.

With this constraint in mind, we examined the effect of allowing the maximum amount of positive cross talk in an AND system and an extreme amount of negative cross talk in an OR system (see Table 3 for the values of a_{ij}^{P*} and a_{ij}^{N*}). Figure 13 shows that massive negative interactions in OR processing can produce a

[27] A reviewer raised the issue about whether capacity coefficient functions such as $C_o(t)$ and $C_a(t)$ can be found for other logical combinations. For instance, suppose the well-known exclusive-or combination (say, A or B but not both) is required for a "yes" response. Now assume that the design is factorial on A, B (hence, not-A = B, and vice versa), and the response assignment is binary (i.e., the "positive" set requires "yes" and the "negative" set requires "no"); then (A, B) and (B, A) stimuli require "yes" responses, whereas (A, A) and (B, B) require "no" responses. Note that all stimuli in this paradigm demand exhaustive processing, even though the stimulus–response assignment is 2:1! Furthermore, in this design, there would be no stimuli that would yield estimates of single-target distributions. However, if the experimenter is willing to run some blocks of A versus not-A trials and other blocks of B versus not-B trials, then these distributions can be used to compute $C_a(t)$ in the various cases. In fact, it turns out that for all binary factorial stimuli with binary responses, there are only three kinds of stopping rules that apply: (a) first terminating (race, minimum time, OR), (b) self-terminating on a given position but not on both (the latter would lead to first terminating; for instance, in certain arrangements, processing could cease if and only if an A is found in the left position), and (c) last terminating (exhaustive, AND, conjunction). $C_o(t)$ and $C_a(t)$ apply to (a) and (c). As an example for possibility (b), let the integrated hazard function for an $n = 2$ case, for instance, the stimulus AA be $H_{AA}(t)$; then an appropriate capacity function would be $H_{AA}(t)/H_A(t)$, where the denominator is an estimate of the single-target processing term. Therefore, for this class of paradigms, the present methodological apparatus covers all cases for capacity computation. Beyond this class, we do not have the answers.

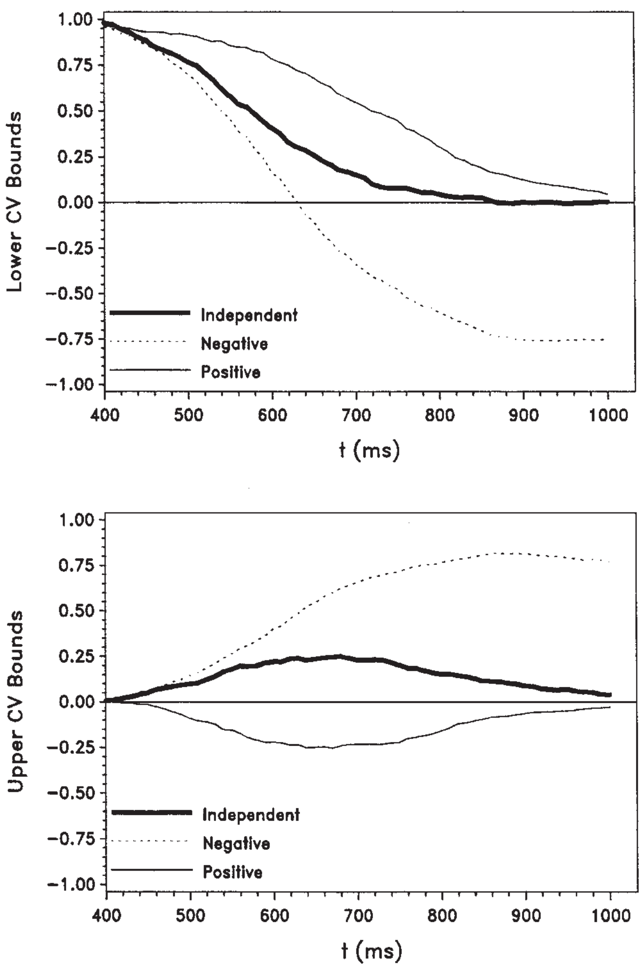

Figure 12. Comparisons of performance for AND processing, for independent processing and for positive and negative dependent processing, relative to the lower (top) and upper (bottom) Colonius–Vorberg (CV) bounds (Colonius & Vorberg, 1994). In both panels, negative values indicate violations of the pertinent bounds.

violation of the lower Colonius–Vorberg AND bound (values less than 0 in the upper panel) and that even more surprisingly, maximal positive interactions in AND processing can produce a violation of the Miller OR bound (values less than 0 in the lower panel).

Coactive Processing and Capacity

Given the intimate connection between violations of the major RT inequalities and the concept of capacity, it is not surprising that coactivation models' abilities to violate Miller's inequality can be interpreted in these terms as well. Townsend and Nozawa (1995) considered a quite general class of counting model representations of the coactivation hypothesis. They referred to this class of models as *channel summation models*, a class that included all of the specific coactivation models (for instance, those based on Poisson counters) fit to data up until that time. Their Propositions 7 and 8 showed that all stochastically independent channel summation models, assuming context invariance, were always super capacity (i.e., for any time t). Furthermore, coactive models with any type of channel dependence (assuming context invariance) inevitably predicted violation of Miller's inequality (and thus implied super capacity at those times when the inequality was violated). It is important to note that because a coactivation system is incapable of making logical decisions (AND, OR, etc.), even in an AND experiment, the AB distribution function will automatically violate Miller's inequality as well as the Colonius–Vorberg upper AND bound. This is a strong prediction of coactivation models. The compression of information into a single channel (and therefore a single dimension) is a powerfully constraining aspect of coactivation. Undoubtedly, other rigorous tests of coactivation will arise out of this constraint.

Thus, we know that coactivation can produce super capacity, even when the two channels do not interact before their summation (e.g., Diederich, 1995; Diederich & Colonius, 1991; Miller, 1982,

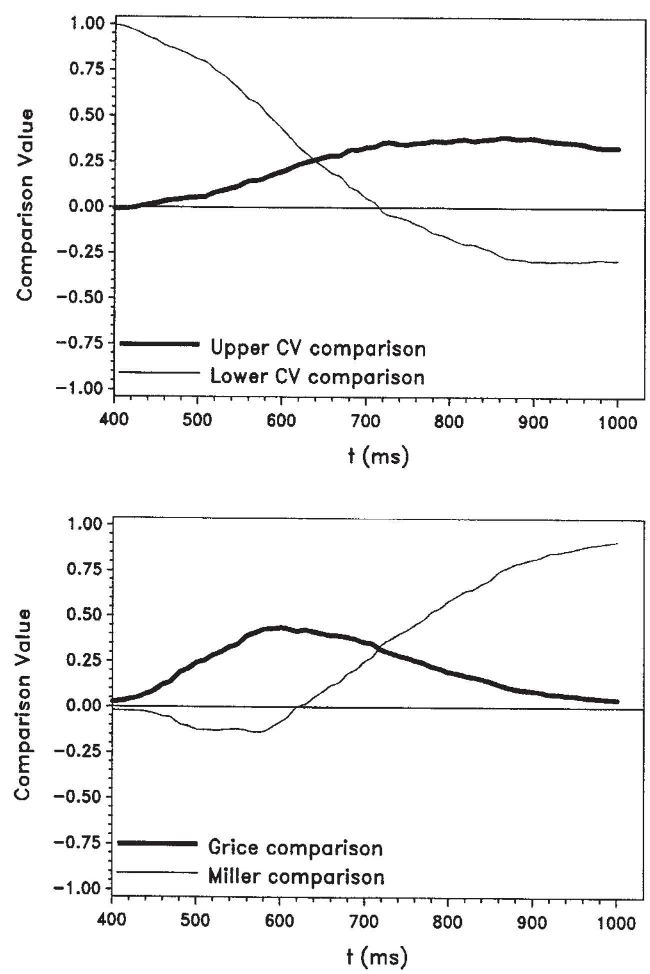

Figure 13. The results of simulating OR processing with massive negative channel interactions and AND processing with maximal positive channel interactions. Top: OR performance compared with the bounds (Colonius & Vorberg, 1994) for AND processing. Bottom: AND performance compared with the AND bounds (Grice comparison: Grice, Canham, & Boroughs, 1984; Grice, Canham, & Gwynne, 1984; Miller comparison: e.g., Miller, 1982) for OR processing. Negative values indicate violations of the pertinent bounds. CV = Colonius–Vorberg.

1986; Mordkoff & Egeth, 1993; Mordkoff & Yantis, 1991; Townsend & Nozawa, 1995, among others). However, what happens in the presence of positive interaction? Can any further increment in capacity be detected? And, what tends to occur in the presence of negative cross talk? Can inhibition wipe out the effects of facilitatory coactivation? To answer these questions, we simulated the channel summation model in Figure 4. As in the simulations of the ordinary parallel models, parameters for channel cross talk were confined to values that kept the deterministic version of the system stable. A critical difference here is that because of the single output channel, results from an AND task will be identical to those from an OR task. Consequently, we can compare one set of data for single- and double-target trials for coactive systems with the bounds for both OR and AND tasks.

Consider first the outcomes associated with analyzing the performance of a coactive system with the capacity coefficients and bounds associated with OR processing (see Figure 14). A result that is immediately apparent in these data is the extreme positive values for $C_o(t)$, extending well beyond the range of values we observed for positive dependencies with normal parallel channels. In addition, these extreme positive values (indicating extreme super-capacity processing) are obtained under conditions of independence in the channels prior to pooling and under conditions of both positive and negative dependence. That is, the presence of negative cross talk prior to pooling does not come close to offsetting the advantage obtained by pooling the channels.

The magnitude of these values for $C_o(t)$ can be understood by making reference to the definition (Equation 14). Any deviations above 1 in the value of $C_o(t)$ will be due to the magnitude of the "separation" between the distribution functions for the double- and single-target trials. As performance in the double-target trials becomes increasingly faster, relative to performance in the single-target trials, the value of $C_o(t)$ will increase. As can be seen in Figure 15, the survivor functions for the double-target trials in all

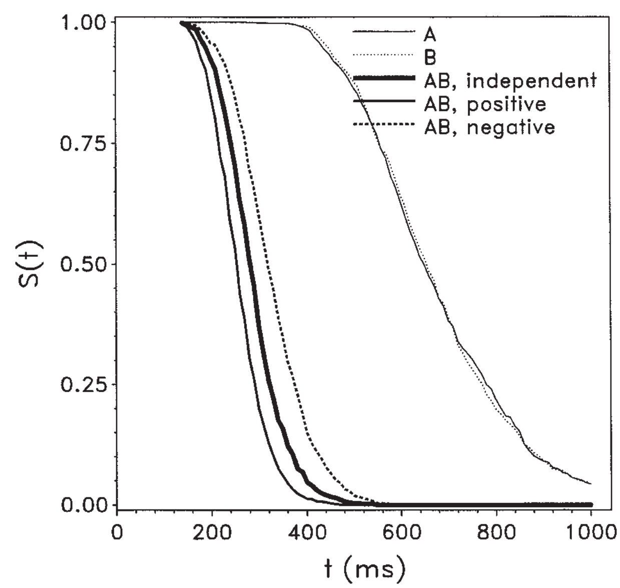

Figure 15. Values of the survivor functions ($S(t)$) for coactive systems under conditions of independence, positive dependence, and negative dependence, while processing single- (A or B) and double-target (AB) trials.

cases—independence, positive dependence, and negative dependence—are greatly separated from those of the single-target trials. In fact, it appears that the nonasymptotic values of the component double- and single-target trials show no overlap.

As we have discussed, we can use the Miller and Grice bounds to corroborate the inferences we have made using $C_o(t)$. Figure 16 presents the results of comparing performance observed with the coactive systems to these two bounds. On the basis of our earlier discussions, we would expect the Miller bound to be violated in all cases and the Grice bound to be violated in none. As can be seen in the figure, this is exactly what we obtained.

Consider next the outcomes associated with analyzing the performance of a coactive system with the tools developed for the AND paradigm: $C_a(t)$ and the Colonius–Vorberg bounds. Figures 17 and 18 illustrate these outcomes when the parallel channels that are summed evidence independence and positive and negative dependence. As we mentioned above, these comparisons involve the same data used in the comparisons against the Miller and Grice bounds. Here, we again see values of load capacity, assessed by $C_a(t)$, that are extreme and positive. And we see these values, indicating extreme super-capacity processing, when the processing channels are independent prior to pooling, when they are positively dependent, and when they are negatively dependent. These data would thus suggest that we should see consistent violations of the upper Colonius–Vorberg bounds but no violations of the lower bounds. Inspection of Figure 18 indicates that this is what we obtained.

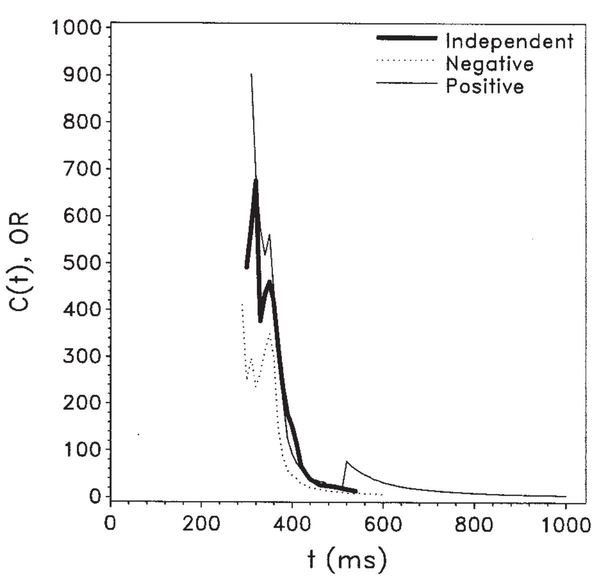

Figure 14. Values of the load capacity coefficient for OR processing, $C_o(t)$, for the channel summation (coactivation) model, for independent processing and positive and negative dependent processing.

Effects of Varying Accuracy

The results presented so far have all been for cases in which response accuracy is nearly perfect (e.g., with error rates less than

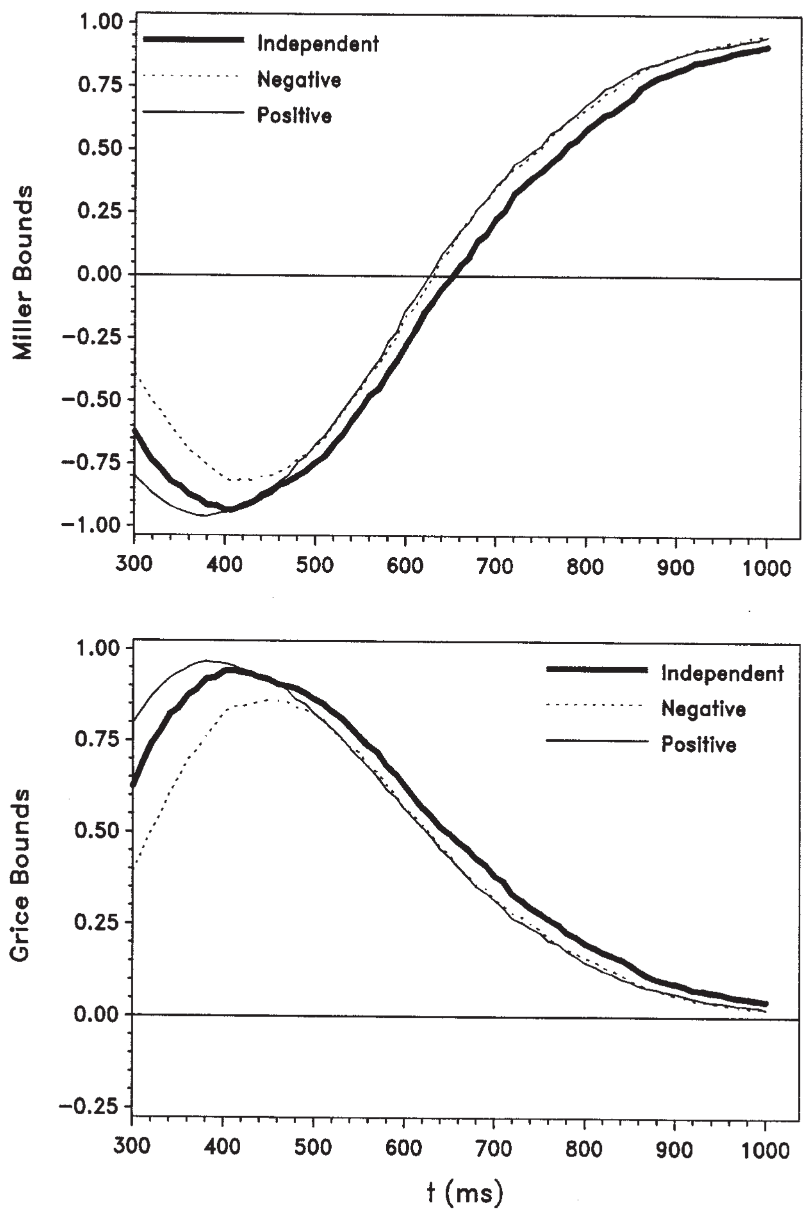

Figure 16. Values for the comparisons against the Miller (top; e.g., Miller, 1982) and Grice (bottom; Grice, Canham, & Boroughs, 1984; Grice, Canham, & Gwynne, 1984) bounds for OR processing, for the channel summation (coactivation) model, for independent processing and positive and negative dependent processing.

5%). What happens in situations in which the error rates are higher than this? The development of the entire theory for response times and errors combined lies far beyond the scope of the present study. However, we can indicate the probable generalizability to such conditions with some limited simulations, which we carry out for separate-decisions models. To simulate moderate- to high-error conditions, we considered one possible mechanism for errors, one that assumes a pair of channels for every dimension or feature presented for processing. One of the channels is responsible for accumulating evidence in favor of one of the two responses, while the other is responsible for accumulating evidence in favor of the alternative response. To simplify the discussion, we denote the correct processing and decision mechanisms as involving the "positive" channel and the incorrect mechanisms as involving the "negative" channel. Note that this is but one of the possible models of error responses in our framework. It allows our systems to be directly comparable with standard diffusion models (e.g., Ratcliff, 1978).

To examine the extent to which our basic predictions hold as accuracy changes, we extended our two-channel model to possess two pairs of channels. In each pair, one channel was dedicated to accumulating positive evidence; this was accomplished by allowing the sign of the constant portion of the input to be positive (and of the same magnitude as used in our earlier simulations). The other channel in each pair was dedicated to accumulating negative

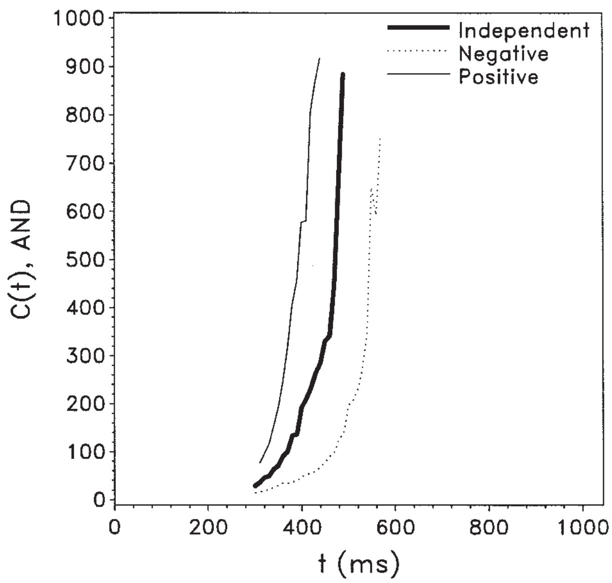

Figure 17. Values of the load capacity coefficient for AND processing, $C_a(t)$, for the channel summation (coactivation) model, in the case of independent processing and positive and negative dependent processing.

evidence; this was accomplished by allowing the sign of the constant portion of the input to be negative. The level of activation in this negative channel was compared against a criterion that was equal in magnitude but different in sign from the criterion for the positive channel. The channel in each pair that reached its response criterion first determined the response (positive or negative). The relationship between the magnitudes of the positive and negative inputs on each type of trial determined the error rate. To the extent that the magnitudes were equal, the error rate approached 50%.

Figure 19 presents the capacity coefficients for a set of simulated OR and AND trials, with error rates being approximately 25% in each case. Only the trials involving correct responses were used in the calculations and figure. As can be seen in this figure, the basic predictions for the capacity coefficients, assuming UCIP processing, are virtually identical to those for channels operating at near-ceiling levels of accuracy. Subsequent research will investigate the situation in the presence of dependence. In addition, work exploring predictions for measures used in identifying processing architecture and stopping rule (reported in part in Townsend & Wenger, 1996) indicates that the basic patterns discussed for parallel independent channels are stable across levels of accuracy until error rates exceed (approximately) 30%. Much effort remains to be expended to provide a complete account of why and how this happens; in the meantime, we are reasonably confident that the patterns presented in the current discussion are those that will be observed with moderate error rates.

General Discussion

In the preceding pages, we developed the formal structures representing parallel processing, assumptions and rules, and correlative associations between the formal and empirical domains. We established two qualitative–ordinal chains of inequalities especially germane to OR and AND perceptual–decisional mechanisms, respectively. We proved and discussed a number of theorems about channel interdependencies and their experimental implications with regard to capacity as measured within our theoretical framework. Simulations within the comparatively simple class of dynamic linear systems with noise and decision thresholds revealed valuable information about how real systems can perform when their channels are positively or negatively correlated. In the remainder of the article, we provide a set of brief synopses of applications of these tools in a small set of empirical projects from our lab and then consider the manner in which our enterprise intersects with other approaches and results in the literature.

First Empirical Applications

A natural application of the theoretical technology we have developed in this article is to questions pertaining to the perception of organized, meaningful forms. Human faces as visual stimuli, for example, have drawn considerable experimental and theoretical

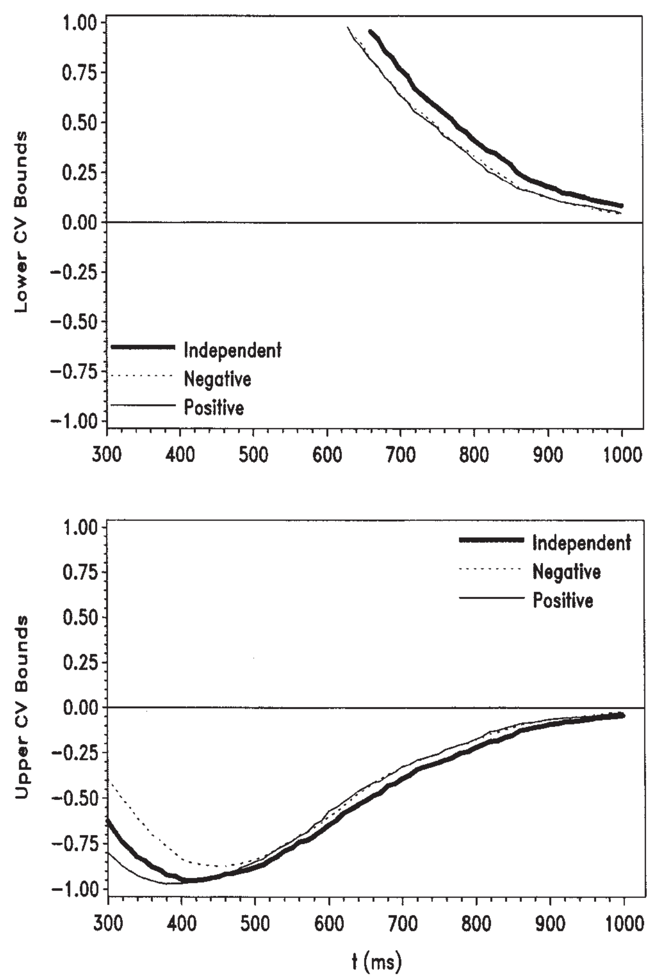

Figure 18. Comparisons against the lower (top) and upper (bottom) Colonius–Vorberg (CV) bounds (Colonius & Vorberg, 1994) for the channel summation (coactivation) model, in the case of independent processing and positive and negative dependent processing. Negative values indicate violation of the pertinent bounds.

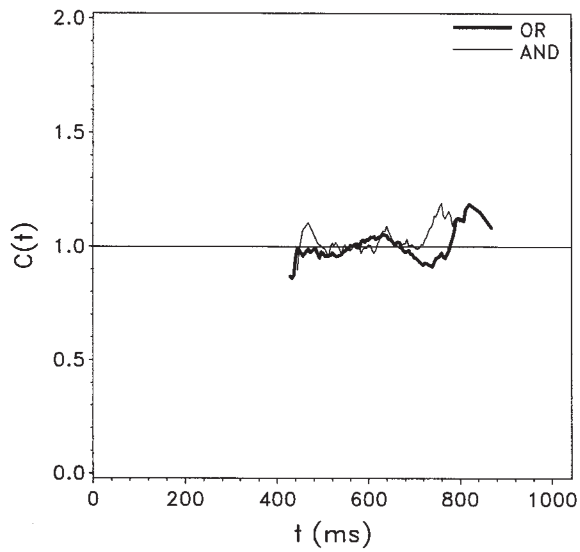

Figure 19. Values of the capacity coefficient for OR and AND trials, with approximately 25% errors. The reference line at 1 indicates unlimited-capacity processing.

attention because of numerous compelling empirical effects (e.g., Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Many of the theories that have been offered to account for these effects have posited specialized representations or processes (or both) for the perceptual and cognitive processing of faces.

We have been pursuing a set of studies examining these ideas, using many of the ideas presented in this article to guide hypotheses, measurement, and inferences. In each of these studies, the ability to quantitatively characterize certain dimensions of processing—centering on capacity—has allowed us to uncover similarities in perceptual processing across forms and to track variations in process capacity as a function of stimulus organization.

On the one hand, we are sensitive to a possible inclination in the literature to hail an empirical result as indicative of true face processing if and only if it differs from one or more other types of object perception: unfortunately, such a strategy can become circular. On the other hand, it may well be that not only the stimuli but also the task demands (instructions, perceptual set, etc.) can set the stage for configural versus nonconfigural kinds of processing (e.g., Bartlett & Searcy, 1993; Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Wenger & Townsend, 2001). Thus, our experimental research program varies such aspects as the set of stimuli, instructions, perceptual set, and response designations to explore when and where empirical indicators of particular process architectures and levels of load capacity might appear. The barest of overall results, regarding usage of some of the tools developed here, are described for two of these studies.

Each of the two studies emphasized the OR design because the theory was first developed in that context. The full methodologies involved aspects of systems factorial technology as applied to parallel versus serial processing (e.g., Townsend, 1984), plus the capacity theory advanced here. Synthesized in a particular fashion, they comprise the double factorial design (Townsend & Nozawa, 1995).

The first of these efforts (reported in detail in Wenger & Townsend, 2001) was an application of the theoretical work reported in Townsend and Nozawa (1995). The OR task involved detection of facial features (eyes, nose, and mouth) in one of four stimulus organizations. In brief, we found evidence for either (a) normal parallel processing, in the context of a first-terminating (i.e., minimum-time) response rule, or (b) coactive processing. From the standpoint of the present theory–methodology, $C_o(t)$ was moderately less than 1 but greater than fixed capacity. Immediately, we know that the Miller inequality should be obeyed, and it was. In addition, the capacity function clearly suggests that capacity was less than expected by UCIP, that is, less than a standard UCIP race model. But the fact that $C_o(t)$ was roughly greater than 0.5 (for the actual exact demands here, see the earlier in-depth discussion) indicates that capacity is greater than, for example, standard serial processing would predict, and it is in turn predicted that the Grice inequality should be satisfied, which the data in fact did.

These basic inferences held across all stimulus organizations, whether the facial features were present upright in their biologically appropriate configuration, inverted, or scrambled (i.e., placed in a biologically inappropriate configuration). These results challenge any simplistic notion that an intact, properly oriented face as a gestalt or holistic perceptual object must inevitably deliver superiority (e.g., as in Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). If that were so, results should have indicated (a) super-capacity coactive processing or (b) super-capacity parallel processing, with positively dependent channels. Note also that the hypothesis that feature processing in valid faces might be exhaustive because of a gestalt, holistic type of perception was falsified.

The second study (Townsend & Wenger, 2004a; Wenger & Townsend, 2000b) was based in part on work exploring the object superiority effect (e.g., Weisstein & Harris, 1974). We constructed a set of stimuli in which the target features—two small down-turned arcs and a single large up-turned arc—were all of the same size and in the same relative location and presented these features in four contexts. In the first, the features were the eyebrows and smile in a schematic face. In the second, these features were the tops of a pair of room dividers and the bottom of a coffee table in a schematic drawing of an interior. In the third context, these features were the tops of the fenders and the bottom of a display banner in a schematic pickup truck at an automotive show. In the fourth context, these features were arbitrary parts of (what was described to observers) as a "word" in an abstract symbol language. To examine the extent to which observers' intent might influence perception, we gave observers one of two sets of orienting instructions. The first asked observers to treat the image as something meaningful (i.e., look for a particular person), whereas the second emphasized only the target features (i.e., look for particular types of curves) and never mentioned the stimulus context. Previous research has shown that reporting names rather than individual features tends to encourage interfeatural dependencies in artificial alphabetic identification accuracy (e.g., Goldstone, 2000; Townsend, Hu, & Evans, 1984). Finally, as it has long been suggested that physically inverting a stimulus, particularly a face, causes the image to be processed in a different and possibly nongestalt manner, we presented each of these forms upright and inverted.

In terms of inferences regarding processing architecture and stopping rule, results of this study were consistent with those from our work with facial features. Specifically, for all stimulus categories, for both types of orienting instructions, and for both upright and inverted presentation, the data supported normal parallel processing in the context of a first-terminating response rule. There was some evidence for super-capacity processing, particularly for upright presentation and for the response instruction that emphasized the meaningful nature of the stimuli. The evidence for super-capacity processing was not restricted to the schematic faces but was obtained for all stimulus categories. The highest load level of capacity was, in fact, obtained for the schematic drawings of the interiors. An interesting aspect of this stimulus class is that it was the only stimulus class in which the target features showed the gestalt characteristics of good continuation and connectedness (see Pizlo, Salach-Golyska, & Rosenfeld, 1997), as the element features were connected to other aspects of the drawing. Conversely, both of these studies found situations in which a "good face" along with holistic instructions led to poorer performance (signified by lower capacity) than did more elemental stimuli (e.g., scrambled features) accompanied with attention to individual features. This kind of "facial inferiority" may be the other side of the coin with configural stimuli and is supported by other accumulating evidence (e.g., Kuehn & Jolicoeur, 1994; Suzuki & Cavanagh, 1995). A final remark relates to the rather incredible magnitude of capacities in our earlier simulations of coactive systems. We so far have not discovered capacities (using the $C(t)$ functions) that are anywhere nearly so extreme in our experiments. This may suggest that milder forms of cross-channel facilitation than coactivation per se are operating in our experiments.

Overall then, these results suggest that the kind of methodology made available by our theory is useful in studying processing mechanisms in many contexts. In particular, within these two studies, it was found that the orienting instruction and the configural quality of the stimuli do have an impact on processing and that the extent to which we may observe differences in characteristics of processing—particularly capacity—may depend on traditional gestalt principles like good continuation.

Relations to Other Theories

Quantitative theories relying on explicit or implicit assumptions of parallelism in information processing are becoming omnipresent in psychology and cognitive science and cannot be completely reviewed here. Nevertheless, it may be helpful in placing the current theoretical work in perspective to sketch in certain approaches that have figured prominently in concerns about stochastic independence over the past few decades.

As noted earlier, the present state-space branch of our theory can be considered an extension of the earlier static general recognition theory of Ashby and Townsend (1986) and as a further generalization of Ashby's (1989, 2000) recent stochastic general recognition theory. As such, it belongs to the general class of accumulator models. Stochastic general recognition theory is based on stochastic linear systems supplemented with decision units, as is the present theory. Ashby (2000) observed that such systems produce a diffusion process, the continuous analogue to discrete-step random walk, and hence are not quite so simple as they may first appear. In addition, our linear stochastic dynamic systems share a number of properties in common with very recent theoretical approaches advanced by Usher and McClelland (2001) and Movellan and McClelland (2001). The present theory emphasizes the details of the parameter structures in stochastic general recognition theory and how they relate to interaction, capacity, and the simplest logical decision gates, OR and AND. The assumption of stability in the underlying linear system is related to the kind of diffusion process under study. Many random walk or diffusion models in psychology are based on or closely related to the Wiener process, named after the mathematician who, along with Einstein, provided a rigorous foundation for Brownian motion. The Wiener process is unstable and permits activation to grow forever. One way to interpret a stable linear system is as a so-called "leaky integrator" that permits some of the accruing activation to leak away. It is a form of the classical law of diminishing returns. The Wiener process provides the basis for Ratcliff's (1978) diffusion model. The random walk model of Link and Heath (1975) can be viewed as a discrete-time version of the Wiener process. However, the Usher and McClelland (2001) leaky integrator model and the Busemeyer and Townsend (1993) decision model are stable, diminishing returns models. Diffusion models along with counter models are major contenders in the contemporary theoretical arena of RT models (e.g., Van Zandt et al., 2000).

With regard to the interpretation of our model as a diffusion model, there are several points worth noting. First, diffusion processes do not automatically offer the systems-based interpretations of process and structure offered by models constructed according to our systems metatheory. It is necessary to implement or find them in, for example, a set of stochastic differential equations that generate the diffusion process. One also requires a system-based interpretation of the equations, for instance, as a set of interrelated system components (e.g., Heath, 2000; Luenberger, 1979).

Linear systems theory provides a rich environment, especially when endowed with noise and logical decision gates, to investigate a myriad of interesting problems in cognitive science. The existence of the state-space concept is perhaps the most critical. Other valuable aspects of using an approach such as the one presented here include the following:

1. Specification of any system requires specification of the form of the input to the system, something often overlooked in the modeling of cognitive processes (see O'Toole, Wenger, & Townsend, 2001). Specification of the form of the input allows one to explore, at varying levels of granularity, effects of variation of the input signal on aspects of system performance across a range of tasks (see, e.g., Heath, 2000; Rouder, 2000).

2. Specification of a system requires explicit consideration of the nature of the initial conditions on any trial. For the situations considered here, it was plausible to assume that the initial condition of all channels in all cases were constant and equal to 0. However, this need not be the case, as might be appropriate in considering hypotheses pertinent to situations in which nonconstant and nonzero initial conditions might be plausible, such as perceptual learning (e.g., Adini, Sagi, & Tsodyks, 2002; Goldstone, 1998).

3. Once a system is specified (in terms of inputs, initial conditions, presence and magnitude of channel interactions, etc.), it is quite easy to introduce decision thresholds on each of the channels and then combine the outputs from these decisional operators in a variety of ways. This ability allowed us to readily model the OR and AND tasks we considered.

4. Dynamic aspects of memory within a system can be represented in the specified systems' impulse response functions (e.g., Townsend & Ashby, 1983, pp. 401–409).

5. Our systems are basically linear, using the simplest kind of nonlinearity to produce decisions: decision thresholds. These sys-

tems present themselves as natural "first steps" in modeling the continuous flow of information in a cognitive system (e.g., Ashby, 1982; McClelland, 1979). Continuous flow systems, in which subprocesses continuously feed their outputs as inputs to subsequent subprocesses (e.g., Liu, 1996; Miller, 1993; Schweickert, 1989; Schweickert & Mounts, 1998; Townsend & Fikes, 1995) can be awesomely complex in general settings, and linearity as we have used it here makes initial exploration tractable. In addition, linearization remains as one of the more tried and tested ways of probing nonlinear systems (e.g., Luenberger, 1979). Naturally, systems that are nonlinear in their basic processing units (that is, in their filter or integrator units) will have an important role to play in explorations of continuous flow systems, as exemplified by the research of Grossberg (e.g., 1978, 1991a) and in important recent texts (e.g., Heath, 2000; Ward, 2002). However, we feel it would be quite myopic to ignore the great progress that has been made in the other sciences through the use of stochastic linear dynamic systems. Notable advances in cognition and perception issuing from the linear systems approach include studies by Anderson (1993); Kohonen (1989); and Loftus, Busey, and Senders (1993). We feel that the present investigation in fact serves as an illustration of this promise and provides the proper level of generality in laying the foundation of dependence–capacity relationships.

6. Not only the dynamics but also the entire system, including the inputs, can be analyzed in terms of wave-form characteristics (as in Fourier analysis; see, e.g., Papoulis, 1991).

7. The stochastic linear dynamic approach shares significant structure with modern approaches to neural and connectionist modeling (e.g., Anderson, 1993; Bar-Yam, 1997; Bishop, 1995; Grossberg, 1978; Kohonen, 1989; Lacouture & Marley, 1995; Mignault & Marley, in press; Ratcliff et al., 1999; Usher & McClelland, 2001). Even with the predecisional linearity, the ability to place various kinds of noise (e.g., possibly non-Gaussian white noise) and interactions at the preprocessing, midprocessing, and postprocessing levels, as well as in a lateral, feed-forward, or feedback mode, enables a wide variety of recurrent and nonrecurrent connections among neural units. In addition, with the availability of sophisticated, massive, and systems-dedicated software, it is possible to investigate rather serious nonlinearities as well. A reviewer expressed doubt that linear systems ". . . can account for sudden changes that are frequently evident in cognitive process, e.g., emergent phenomena such as detecting the woman's face in an ambiguous stimulus." We certainly bear no prejudice against nonlinear systems, but we believe that such questions are basically empirical and need to be answered with direct experimental comparisons of linear versus nonlinear models. In relation to several of the previous items, the present approach is quite compatible with more detailed accounts of information processing, for instance, attention (e.g., Dosher & Sperling, 1998) or memory (e.g., Shiffrin & Steyvers, 1997). As a potential example, we anticipate connections with approaches that investigate sources of noise in perceptual and memorial processing and use such information to learn more about perceptual processing systems (e.g., Dosher & Liu, 1999, 2000).

Conclusion

Interactions among channels, modules, dimensions, features, or items being processed remain one of the least understood yet one of the most critical facets of models of cognitive functioning. Although specific empirical studies and models have indicated the potential power of interactions, in terms of facilitation or inhibition, there is little in the way of general, systematic, quantitative knowledge about the characteristics of such interactions. The present investigation constructs the foundations of a general metatheory of psychological interaction, based on stochastic linear systems theory augmented with logical decision gates, and stochastic processes in general.

The particular types of interaction that we have focused on are produced by channels that are coupled by leading the output of each channel to the input of the other (a cross-feedback arrangement). Within our metatheory, internal feedback is also allowed and the systems are constrained so as to be dynamically stable. The metatheory draws on a growing body of knowledge about capacity in RT, founded on our cognitive stochastic process approach to modeling cognition. This theory emphasizes the analysis of cognitive systems in terms of (a) architecture (e.g., parallel vs. serial vs. coactive), (b) stopping rules (e.g., self-terminating vs. exhaustive), (c) capacity (e.g., limited vs. unlimited vs. super), and (d) dependence relations (with there being many different kinds). Although these issues are logically independent of one another, they can interact in many ways, producing interesting emergent phenomena when they are put together in various combinations in specific model arrangements. Note that we are by no means suggesting that all system interactions act in the same way.

Much of our previous work, as well as the present work, has emphasized capacity as reflecting changes in RT or accuracy as a function of variations in workload, for instance, number of items to be processed (e.g., Townsend, 1974; Townsend & Ashby, 1978; Wenger & Townsend, 2000a). However, recent efforts have begun to delineate capacity issues in situations in which the load is constant: for example, how processing speed or accuracy may change within a trial (e.g., Busey & Townsend, 2001). We suggested the term *load capacity* for the first usage and *constant-load capacity* for the second, in which the meaning might be ambiguous. Note that these two categories are subject to further discriminative terminology as the need arises. The present investigation centered on load capacity, but within-trial effects are evident through the fact that the primary measures are functions of time, rather than single-number statistics, such as means.

The load capacity function, $C_o(t)$, originated in a study of parallel and serial architectures in an OR task, in which completion of either of two subtasks can elicit a correct response (e.g., Townsend & Nozawa, 1995). $C_o(t)$ is a function of the integrated hazard functions from single- and double-target conditions. It was necessary in the present work to invent a new capacity measure, termed $C_a(t)$, which in turn required an analogue of the hazard function more appropriate for AND designs.

The present results form a case in point with regard to the emergent aspects of combinations of positions on the fundamental issues stated above. The observable effects on capacity in terms of speed of processing and as manifested in OR and AND designs are consequences of null (no interaction), negative (inhibitory), or positive (facilitative) couplings. The metatheory articulates the capacity functions and links them with RT inequalities (Colonius & Vorberg, 1994; Grice, Canham, & Boroughs, 1984; Miller, 1982) for OR and AND designs. We have shown that the inequality chain serves as a coarse but highly useful measuring scale against which to assay performance in OR and AND data and to affirm the capacity functions in addition to delineating the circumstances under which super or limited capacity, in their terms, lead to over- or undershooting the bounds appropriate to separate decisions, parallel models in general and UCIP processing in particular. In these senses, violations of the inequality chain (or positions within the chain) emphatically do not imply falsification of the chain itself.

In numerical simulations, it was shown that positive interaction could produce super capacity of a magnitude that forced violations of the upper RT bounds both in OR and AND designs. Such interactions and/or coactivation could be an integral aspect of the formation of gestalts, configural relationships, and the results of unitization in perceptual learning or automatization. Similarly, negative interactions were demonstrated to be capable of evoking serious and extreme limited capacity and violations of lower RT bounds in both types of experimental design. Potential application of negative interactions is to lateral inhibitory circuits (e.g., such as in flanker effects; B. A. Eriksen & Eriksen, 1974; Mordkoff, 1996), opposing perceptual states such as bistable figures, and certain Garner and Stroop effects (e.g., Pansky & Algom, 1999). The dynamic systems large-scale simulations also supported the linkage of the abstract capacity functions to the observable RT inequalities.

One fascinating outcome of our systems simulations was the almost staggering super capacity of the coactive models, relative to positive interactions in stable parallel systems. This extremely super-capacity effect occurred even in the presence of sizeable negative interactions, which caused very limited capacity in the separable decisions models. Such exceptional quantitative differences may aid in the testing of coactive versus interactive separate decisions, parallel models in the future.

Moreover, the powerful capacity effects assayed above, which were caused by positive and negative channel interactions, may help explain the specific strategic strong effects in earlier data as well as some of the previous interactive models by others cited earlier. One of our major current goals is the fleshing out of capacity and architecture characteristics in organized patterns (e.g., faces) versus stimuli consisting of unorganized parts. This research is using the findings of the present study as a strategic aspect of the strategy. Performance efficiency has always been one critical feature associated by psychologists with configural processing. Our results make rigorous exploration of quantified versions of "efficiency" feasible. In addition, there exist other concepts associated with the perception of configural objects, a notable example being that of emergence. A problem has been that even though we can find examples of this idea and claims about nonlinear models' ability to produce emergence are made, there are actually few quantitative examples of much generality. It is likely that such models will need to merge information about the structure of patterns relative to the observer with the kind of simpler state-space results like those worked out here. Other intriguing open research areas include linkages of the present theory with perceptual learning models and experiments and forays exploring potential connections with the growing body of results from cognitive neuroscience.

References

Adini, Y., Sagi, D., & Tsodyks, M. (2002, February 14). Context-enabled learning in the human visual system. *Nature, 415,* 790–793.

Allison, P. D. (1995). *Survival analysis using the SAS system: A practical guide.* Cary, NC: SAS Institute.

Amazeen, E. L. (1999). Perceptual independence of size and weight by dynamic touch. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 102–119.

Anderson, J. A. (1993). The BSB model: A simple nonlinear autoassociative neural network. In M. Hassoun (Ed.), *Associative neural memories* (pp. 77–103). New York: Oxford University Press.

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review, 89,* 599–607.

Ashby, F. G. (1983). A biased random walk model for two choice reaction times. *Journal of Mathematical Psychology, 27,* 277–297.

Ashby, F. G. (1989). Stochastic general recognition theory. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms, and models* (pp. 435–457). Amsterdam: Elsevier.

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology, 44,* 310–329.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1999). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 442–481.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

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

Audley, R. J., & Pike, A. R. (1965). Some alternative stochastic models of choice. *British Journal of Mathematical and Statistical Psychology, 18,* 207–225.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25,* 281–316.

Bar-Yam, Y. (1997). *Dynamics of complex systems.* Reading, MA: Perseus Books.

Bernstein, I. H. (1970). Can we see and hear at the same time? Some recent studies of intersensory facilitation of reaction time. *Acta Psychologica, 33,* 21–35.

Biederman, I., & Kalocsai, P. (1998). Neural and psychophysical analysis of object and face recognition. In H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, & T. Huang (Eds.), *Face recognition: From theory to applications* (Vol. NATO ASI Series F, pp. 3–25). Berlin, Germany: Springer-Verlag.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford, England: Clarendon.

Busemeyer, J. B., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100,* 432–459.

Busey, T. A., & Townsend, J. T. (2001). Independent sampling vs. inter-item dependencies in whole report processing: Contributions of processing architecture and variable attention. *Journal of Mathematical Psychology, 45,* 283–323.

Collett, D. (1994). *Modeling survival data in medical research.* London: Chapman & Hall.

Colonius, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology, 34,* 253–275.

Colonius, H., & Townsend, J. T. (1997). Activation-state representation of models for the redundant-signals-effect. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 245–254). Hillsdale, NJ: Erlbaum.

Colonius, H., & Vorberg, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology, 38,* 35–58.

Cooper, G. R., & McGillem, C. D. (1999). *Probabilistic methods of signal and system analysis.* New York: Oxford University Press.

Cottrell, G. W., Dailey, M. N., Padgett, C., & Adolphs, R. (2001). Is *all* face processing holistic? The view from UCSD. In M. J. Wenger & J. T.

Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 347–396). Mahwah, NJ: Erlbaum.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data.* London: Chapman & Hall.

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

Davenport, W. B., & Root, W. L. (1958). *An introduction to the theory of random signals and noise.* New York: McGraw Hill.

Diederich, A. (1991). *Intersensory facilitation: Race, superposition, and diffusion models for reaction time to multiple stimuli.* Frankfurt, Germany: Peter Lang.

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology, 39,* 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,* 83–86.

Dosher, B. A. (1979). Empirical approaches to information processing: Speed-accuracy tradeoff functions or reaction time—A reply. *Acta Psychologica, 43,* 347–359.

Dosher, B. A. (1984). Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 541–574.

Dosher, B. A., & Liu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research, 39,* 3197–3221.

Dosher, B. A., & Liu, Z. L. (2000). Noise exclusion in spatial attention. *Psychological Science, 11,* 139–146.

Dosher, B. A., & Sperling, G. (1998). A century of human information-processing theory: Vision, attention, and memory. In J. Hochberg (Ed.), *Perception and cognition at century's end. Handbook of perception and cognition* (2nd ed., pp. 199–252). San Diego, CA: Academic Press.

Dzhafarov, E. N. (1993). Grice-representability of response time distribution families. *Psychometrika, 58,* 281–314.

Dzhafarov, E., & Bückenholt, U. (1995). Decomposition of recurrent choices into stochastically independent counts. *Journal of Mathematical Psychology, 39,* 40–56.

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

Egeth, H. E., & Dagenbach, D. (1991). Parallel versus serial processing in visual search: Further evidence from subadditive effects of a visual quality. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 550–559.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16,* 143–149.

Eriksen, C. W., & Spencer, T. (1969). Rate of information processing in visual perception: Some results and some methodological considerations. *Journal of Experimental Psychology, 79*(2, Pt. 2), 1–16.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105,* 482–498.

Fisher, D. L. (1984). Central capacity limits in consistent mapping, visual search tasks: Four channels or more? *Cognitive Psychology, 16,* 449–484.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont donnés [On tables of correlation when the margins are given]. *Annales de l'Universite de Lyon, Section A, Séries 3, 14,* 53–77.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585–612.

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

Grice, G. R., Boroughs, J. M., & Canham, L. (1984). Temporal dynamics of associative interference and facilitation produced by visual context. *Perception & Psychophysics, 36,* 499–507.

Grice, G. R., Canham, L., & Boroughs, J. M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics, 35,* 451–463.

Grice, G. R., Canham, L., & Gwynne, J. W. (1984). Absence of a redundant signals effect in a reaction time task with divided attention. *Perception & Psychophysics, 36,* 565–570.

Grossberg, S. (1978). Competition, decision, and consensus. *Journal of Mathematical Analysis and Applications, 66,* 470–493.

Grossberg, S. (1991a). Nonlinear neural networks: Principles, mechanisms, and architectures. In G. A. Carpenter & S. Grossberg (Eds.), *Pattern recognition by self-organizing neural networks* (pp. 36–109). Cambridge, MA: MIT Press.

Grossberg, S. (1991b). Unitization, automaticity, temporal order, and word recognition. In G. A. Carpenter & S. Grossberg (Eds.), *Pattern recognition by self-organizing neural networks* (pp. 595–614). Cambridge, MA: MIT Press.

Heath, R. A. (2000). *Nonlinear dynamics: Techniques and applications in psychology.* Mahwah, NJ: Erlbaum.

Hughes, H. C., & Townsend, J. T. (1998). Varieties of binocular interaction in human vision. *Psychological Science, 9,* 53–60.

Ingvalson, E. M., & Wenger, M. J. (in press). A strong test of the dual mode hypothesis. *Perception & Psychophysics.*

Kadlec, H. (1992). *Information processing of stimulus modules and dimensions in visual perception: Independence and interactions.* Unpublished doctoral thesis, Purdue University, West Lafayette, IN.

Kadlec, H. (1999, July). *Emergent properties as perceptual interactions in the general recognition theory.* Paper presented at the 32nd Annual Meeting of the Society for Mathematical Psychology, Santa Cruz, CA.

Kadlec, H., & Hicks, C. L. (1998). Invariance of perceptual spaces and perceptual separability of stimulus dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 80–104.

Kadlec, H., & Townsend, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology, 36,* 325–374.

Kadlec, H., & Townsend, J. T. (1992b). Signal detection analysis of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–228). Hillsdale, NJ: Erlbaum.

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

Kantowitz, B. H. (1978). On the accuracy of speed-accuracy tradeoff. *Acta Psychologica, 42,* 79–80.

Kantowitz, B. H. (1985). Channels and stages in human information processing: A limited analysis of theory and methodology. *Journal of Mathematical Psychology, 29,* 135–174.

Kantowitz, B. H., & Knight, J. L. (1976). On experimenter-limited processes. *Psychological Review, 83,* 502–507.

Kohonen, T. (1989). *Self-organization and associative memory* (3rd ed.). Berlin, Germany: Springer-Verlag.

Kuehn, S. M., & Jolicoeur, P. (1994). Impact of the quality of the image, orientation, and similarity of the stimuli on visual search for faces. *Perception, 23,* 95–122.

Lacouture, Y., & Marley, A. A. J. (1995). A mapping model of bow effects in absolute identification. *Journal of Mathematical Psychology, 39,* 383–395.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40,* 77–105.

Liu, Y. (1996). Queuing network modeling of elementary mental processes. *Psychological Review, 103,* 116–136.

Loftus, G. R., Busey, T. A., & Senders, J. W. (1993). Providing a sensory basis for models of visual information acquisition. *Perception & Psychophysics, 54,* 535–554.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1996). Attention in the acquisition and expression of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 620–638.

Luce, R. D. (1986). *Reaction times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Luenberger, D. G. (1979). *Introduction to dynamic systems: Theory, models and applications.* New York: Wiley.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Erlbaum.

Marley, A. A. J. (1992). Developing and characterizing multidimensional Thurstone and Luce models for identification and preference. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 299–334). Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a general principle.* Cambridge, MA: Bradford.

MathWorks. (2001). Matlab 6.1 [Computer software]. Natick, MA: Author.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

Mignault, A., & Marley, A. A. J. (in press). A real-time neuronal model of classical conditioning. *Adaptive Behavior.*

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

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

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 160–169.

Miller, J. O. (1993). A queue-series model for reaction time, with discrete-stage and continuous-flow models as special cases. *Psychological Review, 100,* 702–715.

Miller, J. (1999). Effects of stimulus–response probability on choice reaction time: Evidence from the lateralized readiness potential. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1521–1534.

Mordkoff, J. T. (1996). Selective attention and internal constraints: There is more to the flanker effect than biased contingencies. In A. F. Kramer & M. G. H. Coles (Eds.), *Converging operations in the study of visual selective attention* (pp. 483–502). Washington, DC: American Psychological Association.

Mordkoff, J. T., & Egeth, H. E. (1993). Response time and accuracy revisited: Converging support for the interactive race model. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 981–991.

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 520–538.

Movellan, J. R., & McClelland, J. L. (2001). The Morton–Massaro law of information integration: Implications for models of perception. *Psychological Review, 108,* 113–148.

O'Toole, A. J., Wenger, M. J., & Townsend, J. T. (2001). Quantitative models of perceiving and remembering faces: Precedents and possibilities. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 1–38). Mahwah, NJ: Erlbaum.

Pachella, R. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). Hillsdale, NJ: Erlbaum.

Pansky, A., & Algom, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 39–59.

Papoulis, A. (1991). *Probability, random variables, and stochastic processes.* New York: McGraw-Hill.

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

Perrin, N. A. (1992). Uniting identification, similarity, and preference: General recognition theory. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 123–146). Hillsdale, NJ: Erlbaum.

Pike, A. R. (1973). Response latency models for signal detection. *Psychological Review, 80,* 53–68.

Pizlo, Z., Salach-Golyska, M., & Rosenfeld, A. (1997). Curve detection in a noisy image. *Vision Research, 37,* 1217–1241.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences, 24,* 574–590.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review, 104,* 319–343.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106,* 261–300.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81,* 275–280.

Ross, S. M. (1997). *Introduction to probability models* (6th ed.). San Diego, CA: Academic Press.

Rouder, J. N. (2000). Assessing the roles of change discrimination and luminance integration: Evidence for a hybrid race model of perceptual decision making in luminance discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 359–378.

Rumelhart, D. E., & McClelland, J. L. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88,* 375–407.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

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

Schweickert, R., & Boggs, G. J. (1984). Models of central capacity and concurrency. *Journal of Mathematical Psychology, 28,* 223–281.

Schweickert, R., & Mounts, J. (1998). Additive effects of factors on reaction time and evoked potentials in continuous-flow models. In C. E. Dowling & F. S. Roberts (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement* (pp. 311–327). Mahwah, NJ: Erlbaum.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27,* 379–423.

Shannon, C. E., & Weaver, W. (1963). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Shiffrin, R. M. (1976). Capacity limitations in information processing, attention, and memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Memory processes* (Vol. 4, pp. 25–68). Hillsdale, NJ: Erlbaum.

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

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review, 102,* 567–593.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology, 44,* 408–463.

Smith, P. L., & Vickers, D. (1989). Modeling evidence accumulation with partial loss in expanded judgment. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 797–815.

Smith, P. L., & Zandt, T. V. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology, 53,* 293–315.

Sternberg, S. (1966, August 5). High-speed scanning in human memory. *Science, 153,* 652–654.

Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 901–913.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46*(A), 225–245.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25,* 583–592.

Thomas, E. A. (1971). Sufficient conditions for monotone hazard rate: An application to latency-probability curves. *Journal of Mathematical Psychology, 8,* 303–332.

Thomas, R. D. (1995). Gaussian general recognition theory and perceptual independence. *Psychological Review, 102,* 192–200.

Thomas, R. D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology, 40,* 318–341.

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

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

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133–168). Hillsdale, NJ: Erlbaum.

Townsend, J. T. (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). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should be) distinguished. *Psychological Sciences, 1,* 46–54.

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

Townsend, J. T. (1992). Chaos theory: A brief tutorial and discussion. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (Vol. 1, pp. 65–96). Hillsdale, NJ: Erlbaum.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3, pp. 200–239). Hillsdale, NJ: Erlbaum.

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

Townsend, J. T., & Fikes, T. (1995). *A beginning quantitative taxonomy of cognitive activation systems and application to continuous flow processes* (Tech. Rep. No. 131). Indiana University Bloomington, Cognitive Science Program.

Townsend, J. T., Hu, G. G., & Evans, R. J. (1984). Modeling feature perception in brief displays with evidence for positive interdependencies. *Perception & Psychophysics, 36,* 35–49.

Townsend, J. T., & Nozawa, G. (1995). On the spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39,* 321–359.

Townsend, J. T., & Nozawa, G. (1997). Serial exhaustive models can violate the race model inequality: Implications for architecture and capacity. *Psychological Review, 104,* 595–602.

Townsend, J. T., & Wenger, M. J. (1996). *Evidence monitoring theory: A dynamic extension of general recognition theory and cognitive stochastic processing theory.* Chapel Hill, NC: Society for Mathematical Psychology.

Townsend, J. T., & Wenger, M. J. (1999, November). *Evidence monitoring theory: Foundation and experimental application.* Paper presented at the 40th Annual Meeting of the Psychonomic Society, Los Angeles, CA.

Townsend, J. T., & Wenger, M. J. (2004a). *The costs and benefits of faces and words.* Manuscript in preparation.

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

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review, 108,* 550–592.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences, 21,* 1–14.

Van Zandt, T. (2002). Analysis of response time distributions. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology* (3rd ed., pp. 461–516). San Diego, CA: Academic Press.

Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review, 7,* 208–256.

Ward, L. M. (2002). *Dynamical cognitive science.* Cambridge, MA: MIT Press.

Weisstein, N., & Harris, C. S. (1974, November 22). Visual detection of line segments: An object superiority effect. *Science, 186,* 752–755.

Wenger, M. J. (1999). On the whats and hows of retrieval in the acquisition of a simple skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1137–1160.

Wenger, M. J., & Gibson, B. S. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 708–719.

Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 872–892.

Wenger, M. J., & Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1106–1118.

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

Wenger, M. J., & Townsend, J. T. (2000b, July). *Tracking capacity in object and face perception: New measures and new findings.* Paper presented at the 33rd Annual Meeting of the Society for Mathematical Psychology, Kingston, Ontario, Canada.

Wenger, M. J., & Townsend, J. T. (2001). Faces as gestalt stimuli: Process characteristics. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition* (pp. 229–284). Mahwah, NJ: Erlbaum.

Wheeler, D. D. (1970). Processes in word identification. *Cognitive Psychology, 1,* 59–85.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85.

(*Appendixes follow*)

Appendix A

Simulation Methods

Data for all the simulations reported in the text were generated with the system parameters listed in Table 3. As noted in the text, parameter values were chosen to generate a realistic range of observable RTs. That is, all simulated data possessed minimum and maximum values that correspond to those typically observed in standard laboratory tasks. Although it is possible to generate data values similar to those reported in the text with other combinations of parameter values (particularly for combinations of the parameters specifying values of the constant portion of the input, the channel rate parameters, and the thresholds), the particular values chosen allowed for the range of system configurations to be considered. In addition, we should emphasize that the results we present are obtainable across a range of parameter value combinations.

All models were constructed and simulated with Matlab (MathWorks, 2001). For any given system (e.g., two or four channels, independent, positively or negatively dependent), numeric estimates of channel activations, as given by the solution to the system of differential equations under consideration, were obtained for each value of t (in units of milliseconds) across a 2,000-ms interval. Noise was added to each operative channel at each instant in time. Observable RTs were obtained by, first, determining when each of the channels met or exceeded its response threshold and then, second, feeding those values to the appropriate decision gate (i.e., OR and AND).

Data from each set of replications were analyzed with SAS. The distribution functions reported in the text were estimated with the standard actuarial (life table) method; virtually identical results were obtained (but are not reported) for distributions estimated using Kaplan–Meier (product limit) methods (see discussions in Allison, 1995; Collett, 1994; Cox & Oakes, 1984).

Appendix B

Proofs of Propositions 1 and 2

The proofs of Propositions 1 and 2 are given here, along with some technical remarks. We assume $F_A(t)$ and $F_B(t)$ are absolutely continuous with continuous densities, strictly increasing, and that either $F_A(t) \geq F_B(t)$ or $F_B(t) \geq F_A(t)$ for all $t \geq 0$.

Proposition 1: OR Capacity Relationships

1. If Miller's inequality is violated at time $t = t_0$, then supercapacity processing is entailed; that is, $C_o(t_0) > 1$.

2. If $C_o(t) > 1 + \delta$ for some arbitrary constant $\delta > 0$ for an interval of time near $t = 0$, then Miller's inequality is violated for some window of time within that interval.

3. Being limited capacity is not sufficient to force violations of Grice's inequality. The degree of limited capacity permitted without violation of the Grice inequality is a function of any disparity in speed between Channels A and B.

Proofs of Parts 1 and 2 of Proposition 1 are given in Townsend and Nozawa (1995). To prove Part 3 of Proposition 1, we need a lemma:

Lemma 1: Let $S_i(t) = 1 - F_i(t) = 1 - P_i(T_i \leq t)$, $i = A, B$. Violation of Grice's inequality occurs at arbitrary time t if and only if OR load capacity

$$C_o(t) < \frac{\ln\{\min[S_A(t), S_B(t)]\}}{\ln[S_A(t)S_B(t)]},$$

where the right-hand side is always less than 1.

Proof of Lemma 1. Suppose the Grice inequality is violated, meaning $S_{AB}(t) > \min[S_A(t), S_B(t)]$ (see Townsend & Nozawa, 1995, for a demonstration of equivalence of this expression to the original Grice equality). We can write

$$S_{AB}(t) = [S_A(t)S_B(t)]^{C_o(t)};$$

this expression, conjoined with the violation of the inequality, implies that

$$C_o(t) < \frac{\ln\{\min[S_A(t), S_B(t)]\}}{\ln[S_A(t)S_B(t)]};$$

and it is easy to see that the right-hand side is less than 1, so the necessity follows. Sufficiency is entailed by reversing the argument.

Proof of Part 3 of Proposition 1

Without substantial loss of generality, suppose $S_A(t) = \min[S_A(t), S_B(t)]$ for all $t \geq 0$. Then, Lemma 1 is immediately interpreted as implying a violation if and only if

$$C_o(t) \leq \frac{H_A(t)}{H_A(t) + H_B(t)}.$$

Clearly, the ceiling for $C_o(t)$ is bounded between 1/2 and 1, depending on the magnitudes ($H_A(t) > H_B(t)$) of $H_A(t)$ and $H_B(t)$. $C_o(t) = 1 - \epsilon, \epsilon > 0$, would indicate a capacity close to unlimited. See the text for this discussion.

Proposition 2: AND Capacity Relationships

1. Violation of Colonius–Vorberg upper bound at $t = t_0$ implies super capacity at that time. That is, $C_a(t_0) > 1$.

2. Being super capacity is insufficient to force violation of the Colonius–Vorberg upper bound. The degree of super capacity permitted without violating the upper bound is a function of any disparity of processing speeds between Channels A and B.

3. Violation of the Colonius–Vorberg lower bound at time t_0 implies limited capacity, that is, $C_a(t_0) < 1$.

4. If the system is limited capacity for all times exceeding some time t_0, then the Colonius–Vorberg lower bound will be violated for some interval of time that starts at t_0.

Proof of Parts 1 and 2 of Proposition 2

For proofs of Parts 1 and 2 of Proposition 2, we may use Lemma 2:

Lemma 2. The Colonius–Vorberg upper bound is violated if and only if

$$C_a(t) > \frac{\ln[F_A(t)] + \ln[F_B(t)]}{\ln\{\min[F_A(t), F_B(t)]\}}.$$

Proof of Lemma 2. Assume the Colonius–Vorberg upper bound is violated, that is, $P_{AB}(T_A \leq t, T_B \leq t) = F_{AB}(t) > \min[F_A(t), F_B(t)]$ (see Equation 13). Assume that $F_A(t) = \min[F_A(t), F_B(t)]$ for all $t \geq 0$. Then, $F_{AB}(t) > F_A(t)$. Equation 17 and other equations from the *AND Processing* section form the basis of the following derivation. From Equation 17, we can write

$$F_{AB}(t) = \exp[K_{AB}(t)] = [F_A(t), F_B(t)]^{1/C_a(t)}$$

$$= \exp\left[\frac{K_A(t) + K_B(t)}{C_a(t)}\right] = \{\exp[K_A(t)]\exp[K_B(t)]\}^{1/C_a(t)}.$$

Hence, the violation implies that $F_{AB}(t) > F_A(t) = \min[F_A(t), F_B(t)]$, that is,

$$[F_A(t)F_B(t)]^{1/C_a(t)} > F_A(t),$$

which becomes

$$C_a(t) > \frac{\ln[F_A(t)] + \ln[F_B(t)]}{\ln[F_A(t)]}.$$

Reversing the argument completes the proof.

Because

$$\frac{\ln[F_A(t)] + \ln[F_B(t)]}{\ln[F_A(t)]}$$

represents the appropriate comparison for $C_a(t)$ under the conditions of Part 1 of Proposition 2 and this expression is always at least as great as 1, Lemma 2 leads directly to a proof of Part 1 of Proposition 2. It can also be seen from the result of Lemma 2 that $C_a(t)$ could exceed 1 without necessarily forcing a violation of the upper bound. In fact, it can be observed that when $F_A(t) = F_B(t)$, capacity must be greater than 2, or twice what ordinary unlimited capacity requires. Conversely, as $\ln[F_B(t)]/\ln[F_A(t)]$ becomes very small ($F_A(t)$ is very small relative to $F_B(t)$), $C_a(t)$ could be at most only a little larger than 1 (i.e., mildly super capacity) without violating the upper bound.

Proof of Part 3 of Proposition 2

Suppose the Colonius–Vorberg lower bound is violated at some time $t = t_0$. We readily find that

$$[F_A(t_0)F_B(t_0)]^{1/C_a(t_0)} < F_A(t_0) + F_B(t_0) - 1.$$

Note that for the F values close to 0, violations are impossible, and for both close to 1, they are guaranteed if $C_a(t_0) < 1$. Therefore, it is for F values in the intermediate range, that is, where $0 < F_A(t) + F_B(t) - 1 < 1$, that the question has interest. Concentrating on that range of F values, we find first that

$$C_a(t) < \frac{\ln[F_A(t_0)] + \ln[F_B(t_0)]}{\ln[F_A(t_0) + F_B(t_0) - 1]}.$$

Now, the right-hand side is less than 1 for the constrained F region if and only if

$$[1 - F_A(t_0)][1 - F_B(t_0)] > 0,$$

which it is; so, $C_a(t_0) < 1$, and limited capacity processing is entailed.

Proof of Part 4 of Proposition 2

Assume limited capacity processing in the sense that $C_a(t) < \delta$, $0 < \delta < 1$ for $t > t_0 \geq 0$, and yet suppose the Colonius–Vorberg lower bound is satisfied:

$$F_A(t) + F_B(t) - 1 \leq [F_A(t)F_B(t)]^{1/C_a(t)}.$$

This inequality then implies that

$$\frac{\ln[F_A(t)] + \ln[F_B(t)]}{\ln[F_A(t) + F_B(t) - 1]} \leq C_a(t) < \delta.$$

Now examine the ratio on the left as t gets large. Because of the "niceness" of the CDFs, we can use L'Hospital's rule to examine that function:

$$\frac{\dfrac{f_A(t)}{F_A(t)} + \dfrac{f_B(t)}{F_B(t)}}{\dfrac{f_A(t) + f_B(t)}{F_A(t) + F_B(t) - 1}},$$

which because $F_A(t)$ and $F_B(t) \to 1$ smoothly, approaches 1. Hence, $C_a(t)$ cannot stay less than δ, contradicting the assumption.

Received April 3, 2001
Revision received October 29, 2002
Accepted November 3, 2003 ■