

Basic Response Time Tools for Studying General Processing Capacity in Attention, Perception, and Cognition

MICHAEL J. WENGER
Department of Psychology
University of Notre Dame

JAMES T. TOWNSEND
Department of Psychology
Indiana University

ABSTRACT. *One of the more important constructs in the study of attention, perception, and cognition is that of capacity. The authors reviewed some of the common meanings of this construct and proposed a more precise treatment. They showed how the distribution of response times can be used to derive measures of process capacity and to further illustrate how these measures can be used to address important hypotheses in cognition.*

IMAGINE THAT YOU ARE A CLERK in a 24-hr convenience store. You work the midnight to 8:00 a.m. shift—a dangerous one, because the store has been robbed numerous times during this period. You have been trained to watch each customer very carefully for any signs of threat, such as a narrowing of the eyes, a *sneer* or twitch of the lips, rapid glances around the store, even combinations of these signs. And you have to watch for these signals both in the upright faces of the customers as they pass in front of the counter and in their inverted images in the store's security mirrors.

The authors are indebted to Thomas Palmeri, Vanderbilt University, and Lael Schooler, Pennsylvania State University, for their extensive, constructive comments on a previous version of this article. Special thanks are due to MaryLou Cheal for her patience and attention to detail during the preparation of this article.

Address correspondence to Michael J. Wenger, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556; Michael.J.Wenger.4@nd.edu (e-mail).

The situation just described contains a number of tasks similar to those commonly used to study whole and divided attention (e.g., Lavie, 1995). In fact, various aspects of this imaginary situation correspond to commonly used experimental manipulations, including variations in the number of target locations (e.g., places in the store and in the security mirrors), variations in the number of distractors (e.g., customers who have already been deemed as unthreatening), and variations in load (e.g., looking for both narrowed eyes and a twitch of the lips as compared with narrowed eyes alone; assessing more than one new customer). Consequently, it would be meaningful to ask how much of this environmental information you, as the clerk, can process, as well as how efficiently you are processing it, in the various possible situations.

Some of the most basic questions about the ability of humans to attend to and process environmental information take the form of how much and how efficiently. Certainly, some of the most well known of the earliest investigations (e.g., Sir William Hamilton's examination of the span of apprehension) concerned the first of these, though the basic questions appear to have been posed as early as Aristotle. The import of these questions and the implications of their possible answers have continued to the present day. Indeed, assumptions regarding the answers to these types of questions have formed the bedrock for numerous contemporary theories of attention, perception, and higher levels of cognition. Our goal in the present article is to illustrate how a construct fundamental to attention—capacity—can be informatively assessed by using response times (*RTs*). We hope to demonstrate that the functions of time that we call $H(t)$ (the integrated hazard function) and $C(t)$ (a capacity coefficient) can be used in a broad spectrum of psychological tasks in which *RTs* can be observed. In particular, divided attention (e.g., Bonnel & Hafer, 1998; Nyberg, Nilsson, Olofsson, & Baeckman, 1998), selective cuing (e.g., Cheal & Gregory, 1997; Henderson, 1996; Luck, Hillyard, Mouloua, & Hawkins, 1996; Tellinghuisen, Zimba, & Robin, 1997), and designs that vary number of items (e.g., dimensions, objects; de Haan, Lutz, & Noest, 1998; Lavie, 1995) are obvious targets for the application of these functions as are any investigations that use *RT* to measure or compare efficiency of processing in distinct task conditions.

In presenting these functions, we use the tools of stochastic processes (e.g., Parzen, 1960; Ross, 1997). However, the developments that follow, although they have a solid mathematical foundation, are intended for empirical scientists with little background in mathematical modeling. Indeed, the measures we constructed possess an intuitive appeal and should be usable by any laboratory psychologist. Although our presentation is intended as a tutorial, the techniques we present are rather new: Only now are they beginning to be used in actual experimentation, and there are important theoretical questions that will demand future attention.

We begin by sketching some of the conceptual relationships between a number of constructs in the study of attention, perception, and cognition, and constructs pertinent to notions of capacity. Doing this allows us to restrict our focus

and helps us develop some of the intuitions for the tools we will be presenting. We then move to developing the relationships between these notions and the characteristics of the distribution of *RTs* that can be observed in an experimental task. At that point, we introduce some mathematical terminology and notation and develop the specific measures that form the basis for a pair of example applications. We conclude by giving an algorithm for deriving and applying these measures.

Notions of Capacity in Attention, Perception, and Cognition

Capacity is used in two general ways in contemporary cognitive literature (e.g., Payne & Wenger, 1998). The first way tends toward a static concept, whereas the second way (the one with which we are concerned in the present study) is more dynamic. The first use of the notion of capacity—how much—concerns some aspect of containment ability. Examples of this use would include the “size” of the short-term memory buffer (e.g., Atkinson & Shiffrin, 1968; Shibuya & Bundesen, 1988; Shiffrin, 1975; Sperling, 1963; Theios, 1973; Townsend, 1981) and, perhaps, the ultimate capacity of long-term memory. The size of the visual iconic store (e.g., Sperling, 1960) would also be an example of the use of this notion of capacity.

A related, though not identical, form of this type of capacity—one that is as yet somewhat ambiguous—is the size of the attentional “spotlight.” A related use concerns the “amount” of one or more resources that might be needed in a psychological task (e.g., Kahneman, 1973; Navon, 1984; Navon & Gopher, 1979; Schweickert & Boggs, 1984). Examples of this use would be the familiar, unitary, and multiple pools of cognitive resources available for various tasks (see also Bonnel & Hafter, 1998; Kantowitz & Knight, 1976; Luck et al., 1996; Nyberg et al., 1998).

The second use of the notion of capacity is more directly related to the question of how efficiently information is processed. Here the focus is analogous to modern concerns about “bandwidth,” or how much work can be done (or information processed or transmitted) in some unit of time. Familiar examples that make use of this notion of capacity include the “bottlenecks” and “filters” in the models of early and late attentional selection (e.g., Broadbent, 1958; Moray, 1959; Norman, 1968; Treisman, 1960) and aspects of selective maintenance in models of sensory and short-term memory (Atkinson & Shiffrin, 1968; Shiffrin, 1975; see also Marsh & Hicks, 1998). Similar issues have long been of concern in physical systems, where aspects of capacity have been expressed relative to a system’s robustness. For example, it is possible to consider how long a chain might be able to sustain a certain load, with one possible measure of that chain’s capacity being the likelihood of one of its links breaking instantaneously, or how long the chain could last at some load before failing (e.g., Weibull, 1951).

However, we are not proclaiming that, even within these two categories of

concepts, various interpretations and possible measures of capacity are homogeneous. Indeed, one major problem has been a relative paucity of shared, rigorous definitions and measures of capacity, embedded within a theory or metatheory¹ of human information processing. In the work presented here, we will adhere to the definition and codification originated by the second author (e.g., Townsend, 1972, 1974).

Capacity Within and Across Trials

In refining our conception of capacity, we need to be aware that capacity has two "faces": across trial and within trial. That is, as one varies processing load (e.g., number of items presented for processing, number of dimensions of perceptual or cognitive objects, etc.) and assesses changes in performance, one must be clear about whether the evidence speaks to the capacity available within the execution of a single trial or the capacity available on each of a series of trials. To better understand this distinction and its implications, let us go back to the imaginary task from our introductory remarks.

Imagine now that you, as the store clerk, catch a brief glimpse of the face of a new customer as you are completing a transaction. You hold that information in memory while you complete the current transaction, but then you need to search memory in order to determine whether the new customer poses a threat. One type of parallel model of memory search that might be applied to this situation is of *fixed* capacity (a special form of limited capacity, a distinction we will have more to say about later), and we can talk about what happens to that fixed amount of capacity both across and within trials.

For the sake of this example, assume you have three features or dimensions of the new customer's face in memory (e.g., the eyebrows, the eyes, and the mouth). According to a common parallel model of search (e.g., Townsend, 1972), your overall capacity for processing these three features is a constant, $K(t)$, which is allocated equally to each of the three features. That is, as you begin your search among the three features in memory, each of the features has $K(t)/3$ capacity allocated to it. If at some later point in your shift, you have to repeat this task but this time have to hold and search only two features, then each of those features has $K(t)/2$ capacity allocated to it. Thus, this model predicts variations in performance as a function of changes in load across trials. If processing of each of the elements is also independent, then once a "quantity" of capacity is assigned to a feature, it stays unchanged until the processing of that feature is complete.

¹We should note that our entire treatment is firmly within the tradition of what we refer to as metatheories; that is, general theoretical approaches in which mathematical characterization of major operating principles (e.g., serial or parallel processing) leads to invention of experimental designs that rigorously test among major representatives of those principles.

But a variant of this model is capable of predicting variations in capacity within a single trial. In the aforementioned example of searching among three features of the face held in memory, consider what might happen as the first (or fastest) item to be assessed is completed. The capacity $K(t)/3$ that was allocated to that feature is now freed up and can perhaps be reallocated to the remaining two features. Each of these then has $K(t)/2$ capacity available. Subsequently, as the second feature is completed, its capacity allocation is freed up, and the final feature then has the entire $K(t)$ capacity available for completing its processing. Thus, this model predicts variations in capacity within a trial. Now suppose $K(t) = K$, the rate of exponentially distributed processing. Then this model is the famous parallel mimicker of the standard serial Poisson model (Townsend, 1972), although in general, the observer may be able to distribute K unevenly across the features.

For simplicity, we have, in the preceding paragraphs, ignored some rather important issues. To develop our conception of capacity more precisely, we often will need to make explicit assumptions about other characteristics of the human information-processing system. With physical systems, such as chains, aircraft parts, bridge structures, and so forth, concern with capacity can often disregard some of the other fundamental characteristics of the system. For example, a civil engineer might need to consider only the amount of load the bridge could sustain in a unit of time. Concern for the particulars of the bridge's structure (e.g., its architecture) would be secondary. Still, these other levels of analysis, although logically distinct, are functionally related in important ways, among themselves and to capacity. This does not, however, prevent the engineer from being able to assess the capacity of the bridge, whatever its architecture.

Psychologists face a similar situation when they consider the real-time processing capacity of humans. Capacity is really only one of four logically distinct but critically interrelated characteristics of human information processing (e.g., Townsend & Ashby, 1978, 1983; Townsend & Nozawa, 1995). The other characteristics are the *architecture* of processing (e.g., whether the elements of a stimulus need to be processed one at a time or can be processed simultaneously), the *stopping rule* for processing (e.g., whether all of the stimulus elements need to be processed or whether there is some sufficient minimum), and the preservation or violation of *independence* in the rates of processing aspects of the stimulus. Implicit in our distinction between across-trial and within-trial variations in capacity is the fact that our assumptions about these other characteristics of processing will have fundamental implications for the strength of our inferences regarding capacity. For example, when we returned to our example task, we assumed a parallel processing architecture and an exhaustive stopping rule. These implications do not, however, prevent us from being able to assess the capacity of the human information-processing system, as a function of a variety of stimulus conditions, whatever the underlying architecture, stopping rule, or aspects of independence may be. Considering capacity outside specific assumptions regard-

ing these other aspects of processing does place certain limits on the specificity with which we can characterize the system. But we can still make meaningful and important inferences about the real-time characteristics of processing that we can observe in any given task, and these inferences can be used to place important constraints on developing theories.

Levels of Analysis and a Taxonomy

In particular, it is important to observe that capacity can be assessed at different levels. For example, in our memory search task, we can consider the total amount of capacity available for processing: $K(t)$ in our example. A quite coarse measure of capacity would be the average time to complete a set of n items. We can also talk about the capacity available for processing, for example, the first feature to be completed. Even if this first feature were sufficient for generating a response, then the capacity allocated to this item, $K(t)/3$, would not be the same as the total capacity available for the task, $K(t)$. Similarly, we can talk about the capacity available for processing the second feature and, finally, the third feature. At each of these levels (e.g., the individual feature, an item—such as a face—composed of a set of features, exhaustive processing of all these features, etc.), we can quantify capacity at different levels of granularity, from the reasonably coarse (e.g., mean RT) to quite fine-grained. The measures described later, $H(t)$ and $C(t)$, fall into the latter category.

Once we have specified the level at which we want to examine capacity and have chosen a desired level of granularity of description, we can then consider a taxonomy of capacity. This fundamental characterization (Townsend, 1972, 1974) makes reference to the changes we observe (if any) at the level of granularity chosen, as we vary the amount of work the system has to do. To consider this taxonomy, go back again to our imaginary memory search task, where we had to hold three features of a customer's face in memory and then perform a search among those features.

Imagine that we want to characterize capacity at the level of the entire task (i.e., we will not be concerned with the capacity allocated to any of the features) and that, in terms of granularity, we will use the mean RT (which we will notate as \overline{RT}). Now imagine that we can vary load across trials by adding a fourth feature; thus, we might need to search among the eyebrows, eyes, and mouth, or we might need to consider those three features and the nose (which might be flared). We will use \overline{RT}_3 to indicate our mean RT when three features are being searched and \overline{RT}_4 to indicate the mean RT when four features are involved.

If we consider the relationship between these two means, we see that there are three possible outcomes: $\overline{RT}_3 = \overline{RT}_4$, $\overline{RT}_3 < \overline{RT}_4$, $\overline{RT}_3 > \overline{RT}_4$. These three outcomes correspond to the three "flavors" of capacity in the fundamental taxonomy of capacity. In the first case, if $\overline{RT}_3 = \overline{RT}_4$, then at this level of analysis and granularity, we see that our measure of capacity is unaffected by load. We refer

to this situation as reflecting *unlimited* capacity. The idea is that we can increase the workload, and the ability to perform the search is unaffected. In the second case, if $\overline{RT}_3 < \overline{RT}_4$, as we increase the workload, we slow down. We refer to this situation as one that reflects *limited* capacity, in the sense that increases in workload push at the limits of our ability to perform the search. In the third case, $\overline{RT}_3 > \overline{RT}_4$. For some reason, going from three to four features actually allows us to complete the task faster. We refer to this situation as one that reflects *super* capacity, such as we might expect from a well-configured or “gestalt” stimulus (e.g., Wenger & Townsend, in press).

For this taxonomy to be meaningful, the three categories need to be considered with respect to the level of analysis (e.g., the individual input, the entire task, etc.) and the granularity of the measure of performance (e.g., \overline{RT} , or one of the more fine-grained measures described later). However, once capacity is specified for a certain level—for instance, at the individual-element level—then capacity characteristics are implied at coarser levels.

Developing a More Precise Characterization

Now that we have an appreciation for some of the basic issues associated with characterizing capacity—within trial versus across trial, level of analysis, granularity of measurement, and the fundamental taxonomy—we can begin developing our more precise characterization of capacity. Imagine that we take a large number of observations of performance in our memory search task, across a range of different conditions. For example, we might take hundreds of observations at different levels of load, from faces in different orientations (upright or inverted, as seen in the security mirror), all in order to determine (a) whether capacity in this task is limited, unlimited, or super and (b) whether capacity might be affected by the “naturalness” of initial presentation.

After taking all of these observations, we know the range of the observations (in ms). We arbitrarily decide to divide the total range of task times into 10-ms “bins.” Then, for each observation, we determine into which range (which time bin) that observation falls. After doing this for all observations, we determine first the number of observations in each bin and then the proportion of the total number of observations in each bin. The dark line in panel A of Figure 1 plots this summary of a set of observations for this imaginary task, at one level of load in one orientation.

We now adopt some notation to develop our discussion. Let T denote the duration of the task on any one observation, and let t denote any of the possible time values that could be observed. The dark line in panel A of Figure 1 summarizes the proportion of observations that fall into any one of the time bins (between 500 and 700 ms) and allows us to talk about the likelihood of observing a task duration (T) that would fall into any one of those time bins (t). In other words, this line gives an estimate of $P(T = t)$. In abstract terms, what we have is

probability mass or *probability density function* (pdf), $p(t)$ or $f(t)$.² The dark line in panel A of Figure 1 thus gives an estimate of $f(t)$. We will use $\hat{f}(t)$ to indicate the empirical probabilities and $f(t)$ to denote the abstract, general pdf (for the important notation we will be using, see Appendix). At this point, we have completed our job of describing the distribution of task durations, but we have yet to specify our own measures on capacity. To get there, consider two other ways we can describe this distribution of times.

The first, which may be the most familiar, is one we obtain by starting at the lowest time bin and adding up the proportions in each of the time bins until we get to the highest time bin. Assume that we have a total of n time bins. Then, for any one bin m , where $1 \leq m \leq n$, we can talk about the probability that we observed an RT that fell in one of the bins at or before bin m . If we let t_m indicate the RT range for time bin m , then we have

$$P(T \leq t_m) = \sum_{i=1}^m \hat{f}_i(t). \quad (1)$$

What this quantity is giving us is the probability that the memory search has finished at or before the time bin represented by t_m . For example, if t_m represents the time bin that runs from 600–610 ms, $P(T \leq t_m)$ tells us the probability that the process was done at or before 610 ms. Note that we can check to see if we have constructed this function—and $f(t)$ —correctly, by checking that

$$P(T \leq t_m) = \sum_{i=1}^m \hat{f}_i(t) = 1.$$

In more general terms, for any time $0 \leq t^* \leq \infty$, what Equation 1 is estimating for us is

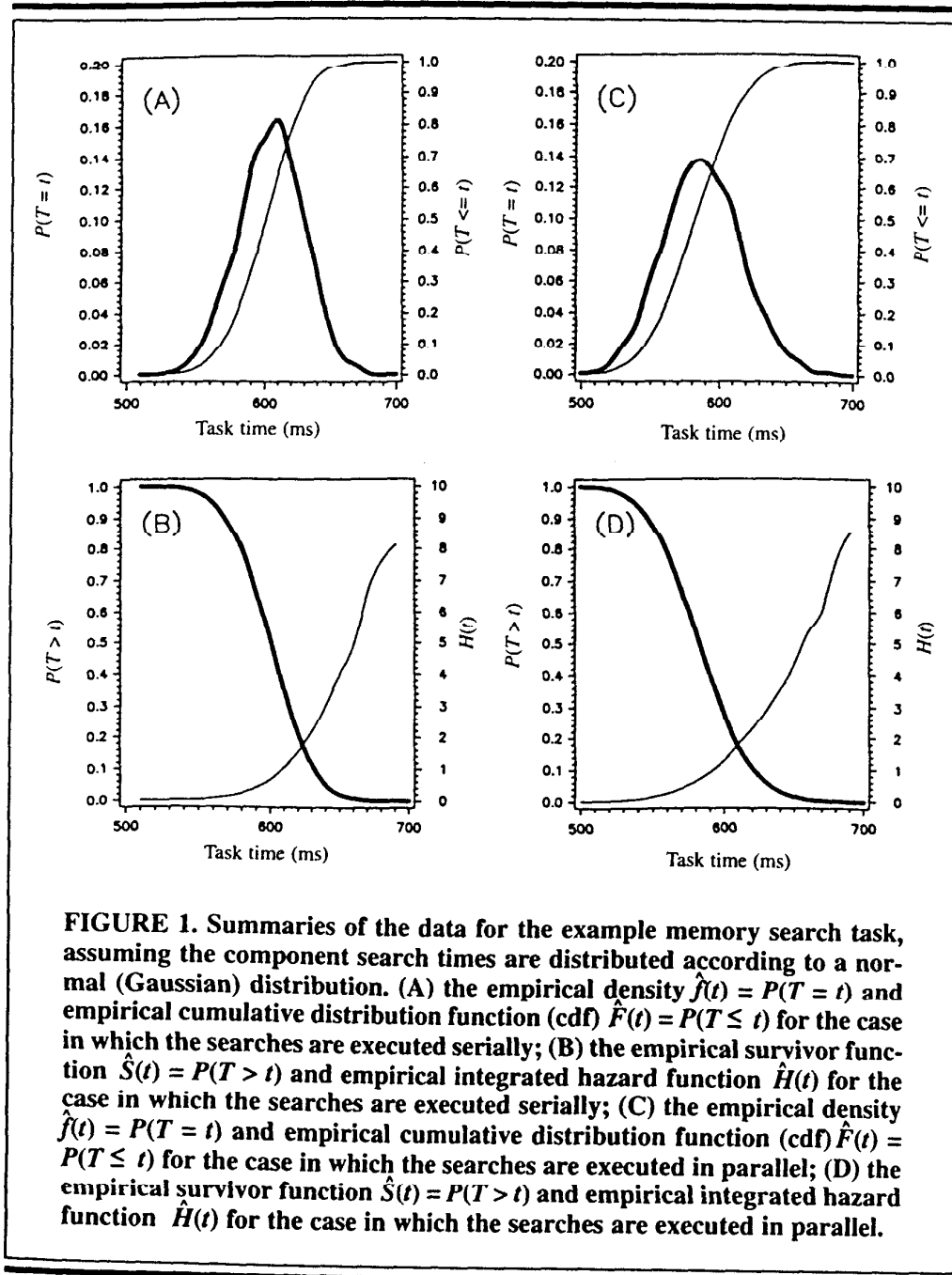
$$P(T \leq t^*) = \int_0^{t^*} f(t) dt \quad (2)$$

or the probability that the process we are interested in has finished at or before time t^* . Note that

$$P(T \leq \infty) = \int_0^{\infty} f(t) dt = 1,$$

just as the finite approximation sums to 1, as shown in the last equation.

²The distinction here depends on whether the variable being considered is discrete or continuous, with the former described in terms of probability mass function $p(t)$ and the latter described in terms of probability density functions $f(t)$ (a comprehensive discussion of issues related to this distinction and their connection to response times can be found in Luce, 1986; Townsend, 1992; Townsend & Ashby, 1983). Of course, even though the underlying probability function may be continuous, the sampled (observed) data will be a probability mass function that will approximate the density.



Equation 2 is referred to as the *cumulative distribution function*, or cdf. Equation 1 gives the empirical estimate of the cdf, and we will use $\hat{F}(t)$ to indicate the empirical estimate and $F(t)$ to indicate the general cdf. The lighter line in panel A of Figure 1 plots $\hat{F}(t) = P(T \leq t)$ for our example task.

The cdf has some nice properties for the present purposes. First, no matter

what the probability distribution looks like, the cdf will always be nondecreasing: In going from the shortest to the longest times, it will always be either increasing or flat, its smallest value will always be 0, and its largest value will always be 1. It does not matter what type of distribution (e.g., exponential, Gaussian, Rayleigh, etc.) it summarizes: The cdf will always look approximately like the one in panel A of Figure 1. Of course, the specific shape properties can differ. For instance, they could be linear (the uniform distribution), always negatively accelerated (curved down, as in the exponential distribution; described later), S-shaped (as in the normal distribution), or even always positively accelerated (curved upward), and so on.

Second, knowing the cdf gives us information about the mean. In contrast, knowing the mean does not necessarily give us information about the cdf (Townsend, 1990). For example, imagine that we have two experimental conditions that have allowed us to generate two cdfs, $\hat{F}_1(t)$ and $\hat{F}_2(t)$. Assume that, for this example, $\hat{F}_1(t)$ comes from the condition in which we see the faces upright and that $\hat{F}_2(t)$ comes from the condition in which we see the faces inverted in the security mirrors. Now assume that $\hat{F}_1(t) > \hat{F}_2(t)$ —that is, our cdfs are strictly ordered and do not cross over at any point. In this case, we automatically can infer that $\overline{RT}_1 < \overline{RT}_2$. What we would be saying is, essentially, that the finishing times in Condition 1 are stochastically shorter than those in Condition 2 at the level of the entire *RT* distribution, not just the level of the mean (although the latter is true by implication as well). It could, in comparison, be the case that $\overline{RT}_1 < \overline{RT}_2$ without there being an ordering in the cdfs. Consequently, a finding of an ordering at the level of the cdfs is a stronger finding (in that it allows us to infer more) than observing an ordering at the level of the means.

Knowledge of the cdf tells us, for any time value, the probability that the process finished at or before that time. A related function, and one that will get us one step closer to our desired measure of capacity, is one that can be obtained directly from the cdf. If $F(t)$ can be interpreted as a statement about when a process is done, then its contradiction (complement) will tell us about when the process is still going on. Since all of the probabilities associated with our time value, when taken together, sum to 1 [i.e., $\int_0^{\infty} f(t)dt = F(\infty) = 1$], the complement of $F(t)$ for any time t is simply $1 - F(t)$ and is known (somewhat morbidly, as it turns out) as the survivor function $S(t)$.³ Where $F(t)$ tells us the probability that a process is done at or before time t , $S(t)$ tells us the probability that the process is not done, that is, that the process is finished later than time t . In panel A of Figure 1, we can see that the probability that we have finished the task of turning on all of our equipment by the time 610 ms have passed is approximately 0.60. Said another way, $P(T \leq 610 \text{ ms}) \approx 0.60$. The dark line in panel B of Figure 1 plots

³The term is derived in part from the insurance industry, where it can be used to describe the probability that someone is still alive at age t .

the survivor function for this example, or $\hat{S}(t) = 1 - \hat{F}(t)$, and gives us the probability that the task is not done by some time t . Consistent with our preceding notation, we will use $\hat{S}(t)$ to indicate the empirical survivor function, and $S(t)$ to indicate the general survivor function. As we expect from the preceding definitions, $P(T > 610 \text{ ms}) \approx 1 - 0.60 = 0.40$.

Because the cdf and survivor function have complementary interpretations, it will come as no surprise that, when there is an ordering in $F(t)$, there is a complementary ordering in $S(t)$. If Process 1 is faster (at the level of the cdf) than Process 2, then $F_1(t) > F_2(t)$. In this case it will also be true that $S_1(t) < S_2(t)$, because at any given time the probability that Process 1 (the faster process) is not done will be less than the probability that Process 2 (the slower process) is not done. Said a bit more coarsely, a faster process gives a higher cdf and a lower survivor function, relative to some slower process (assuming that the adjective faster applies at the level of the RT distribution).

Like the cdf, the survivor function has a number of nice properties. First, knowledge of the survivor function gives us immediate knowledge of the mean. In fact, we can get the mean quite directly from the survivor function in the case of RT distributions:

$$E[T] = \int_0^{\infty} S(t) dt \quad (3)$$

(see Townsend & Ashby, 1983, p. 170). More importantly, having the survivor function puts us only one small step away from being able to specify our basic measure of capacity. To see how this is so, consider the central concept we are trying to make precise. We are trying to characterize capacity in terms of how much work a system is capable of doing in some unit of time. In terms of our example task of searching through the features we have held in memory, consider what might happen at any instant once we have started the search. If we are capable of doing a lot of work in a unit of time, then the likelihood of our finishing the task in the next instant should be quite high. If we are capable of doing only a small amount of work in some unit of time, then the likelihood of our finishing the task in the next instant should be quite low.

Engineers have quantified this ability to do some level of work in a unit of time by way of a function known as the intensity or hazard function.⁴ This function gives the probability of finishing the task in the next instant of time, given that the task is not yet completed. More formally, this general conditional probability function is specified as

⁴This terminology reflects the history of the use of this function to characterize failure rates for different types of devices.

$$h(t) = \frac{f(t)}{S(t)}. \quad (4)$$

The hazard function is analogous to the concept of power in physics (e.g., see Townsend & Ashby, 1983, pp. 77–79). This function has seen some use in cognitive psychology, in applications to situations in which it becomes important to understand the conditional probability of stopping once a process is started (e.g., studies of the ability to inhibit certain processes; Colonius, 1990; Logan & Cowan, 1984; Luce, 1986). But a problem with this function is that it can be quite tricky to estimate in empirical data (e.g., Smith, 1990). However, if instead of characterizing capacity in terms of how intensely the system is working at some instant—that is, its power—we consider how much work it has done up to some point, we find ourselves in an improved position. This notion is, then, analogous to the concept of dispensed energy, which is equal to the integral of power (Townsend & Ashby, 1978).

More colloquially, because the intensity or hazard function indicates how much work the system is capable of doing instantaneously at any moment, if we add up its value at each moment up to some arbitrary time t^* , we will have a measure of how much work the system did up to that point in time. In general terms, this would be given by

$$H(t^*) = \int_0^{t^*} h(t) dt, \quad (5)$$

which we call, appropriately enough, the integrated hazard function. But how are we better off at this point? Our definition is specified in terms of a function that we claimed was difficult to estimate from the data. A rough and ready solution comes from an identity in probability theory, namely,

$$H(t) = -\log[S(t)]. \quad (6)$$

What this identity allows us to do is get the integrated hazard function directly from the empirical survivor function,

$$\hat{H}(t) = -\log [\hat{S}(t)]$$

(Townsend & Ashby, 1983, pp. 26–27). That is, we just take the negative of the logarithm of the value for the empirical survivor function $\hat{S}(t)$ for each one of the time bins, and this gives us the value of the empirical integrated hazard function $\hat{H}(t)$ for each of these bins. Because of the values of the survivor function at the ends of the *RT* distribution, the value of the integrated hazard function at the shortest end will be 0—that is, $-\log(1)$, whereas the value at the longest end will be, theoretically, infinity—that is, $-\log(0)$. This makes sense: If the integrated hazard function gives us how much work the system has done up to some point in time and if we consider how much work it has done in an infinite amount of

time, then the result should be an infinite amount of work. The thin line in panel B of Figure 1 plots the integrated hazard function $\hat{H}(t)$ for our example task.

In practice, the upper limit of the integrated hazard function will be determined by the number of observations that compose the *RT* distribution and the number of time bins we choose to use. For example, assume we have taken 10 observations of our time to complete our example task. We decide to use 10,000 time bins to represent the *RT* distribution, and our 10 observations are distributed so that no more than 1 observation falls into any one bin. Then the next-to-last filled bin will correspond to a value of .90 for the cdf and .10 for the empirical survivor function, and the value of the empirical integrated hazard function at this point will be 1 ($= -\log[.1]$), after which it jumps to infinity. If instead we have 1,000 observations, under the same assumptions, the value of the empirical hazard function at the next-to-last filled time bin will be 3, after which it will go to infinity. Consequently, the more observations we are able to obtain for the condition in which we estimate the integrated hazard function, the better we will be able to estimate the upper range of that function.

The integrated hazard function, like the cdf and survivor functions, has the nice property that it always has the same basic shape: No matter what the form of the system is, and no matter what the form of the distribution of its finishing times is, it always starts at 0 and increases (perhaps with some flatness in between) forever, although for some distributions it will reach an upper limit. To illustrate this, we offer more information about the way we generated the data presented in panels A and B of Figure 1. To generate these data, we assumed that the times required to search for each item were normally distributed. The normal (Gaussian) pdf has the form

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(t-\mu)^2}{2\sigma^2}\right]$$

and describes a distribution with mean μ and standard deviation σ .⁵ The total time to complete the task of searching all three features was the sum of the three individual search times (i.e., serial exhaustive processing). We generated 100,000 samples of times for this task, used a bin size of 10 ms, and then used the frequencies in each of the bins to determine the empirical probability function, $\hat{f}(t)$, which then allowed us to determine the empirical cdf $\hat{F}(t)$, survivor function $\hat{S}(t)$, and integrated hazard function $\hat{H}(t)$.

Now compare these data with a set of observations generated under differ-

⁵Note that this distribution is defined for all real numbers, positive and negative: It associates probabilities with negative *RTs*. To deal with this, theorists often use a *truncated normal distribution* (e.g., Schweickert, Fisher, & Goldstein, 1992), that is, one that is defined to have a lower bound of 0 and is adjusted to preserve total probability.

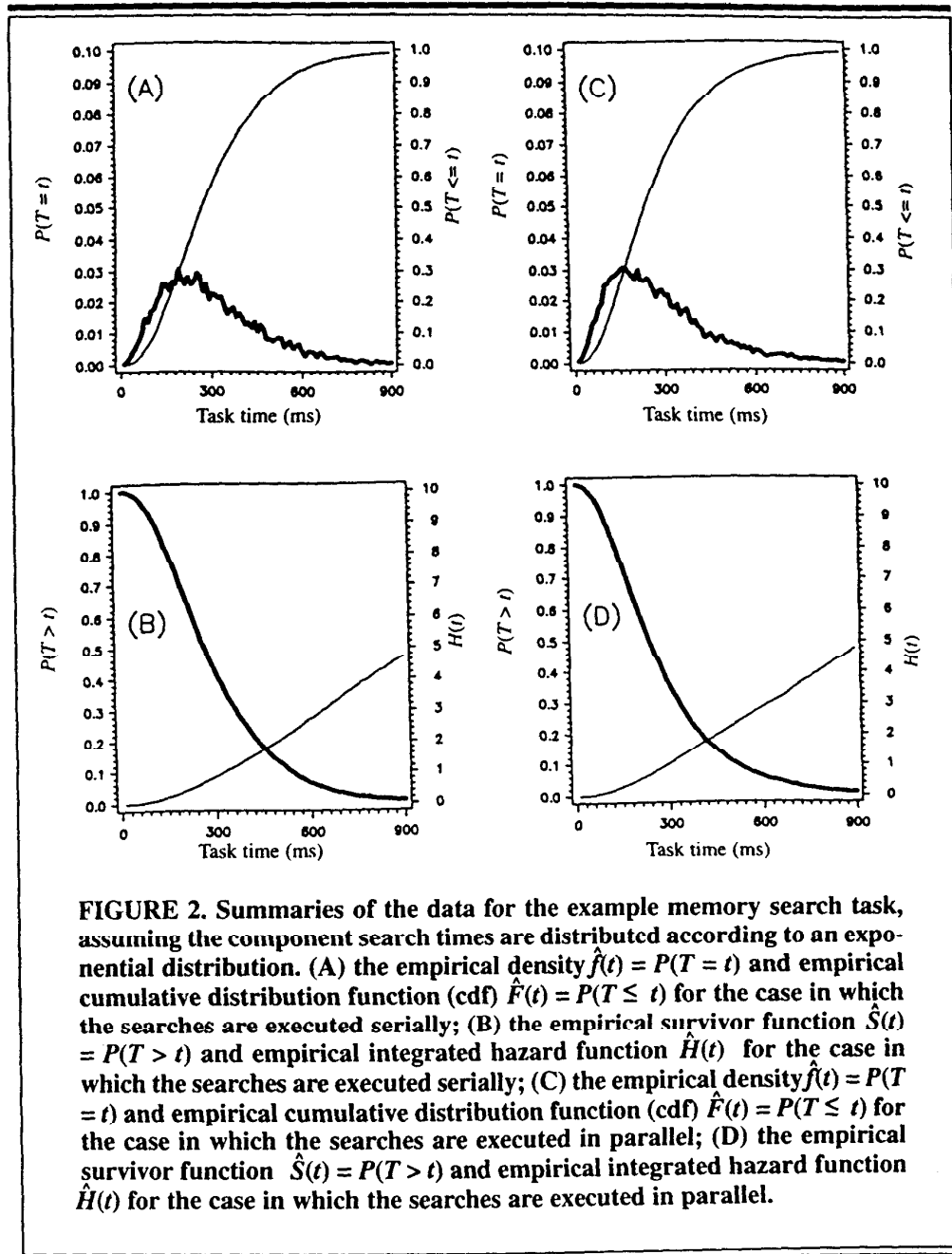


FIGURE 2. Summaries of the data for the example memory search task, assuming the component search times are distributed according to an exponential distribution. (A) the empirical density $\hat{f}(t) = P(T = t)$ and empirical cumulative distribution function (cdf) $\hat{F}(t) = P(T \leq t)$ for the case in which the searches are executed serially; (B) the empirical survivor function $\hat{S}(t) = P(T > t)$ and empirical integrated hazard function $\hat{H}(t)$ for the case in which the searches are executed serially; (C) the empirical density $\hat{f}(t) = P(T = t)$ and empirical cumulative distribution function (cdf) $\hat{F}(t) = P(T \leq t)$ for the case in which the searches are executed in parallel; (D) the empirical survivor function $\hat{S}(t) = P(T > t)$ and empirical integrated hazard function $\hat{H}(t)$ for the case in which the searches are executed in parallel.

ent assumptions. Assume now that instead of taking the sum of the search times, we find the maximum of the three search times. This would correspond to a parallel exhaustive search of the three features. The data generated for this model are presented in panels C and D of Figure 1. Finally, instead of assuming that the search times for the individual features are normally distributed, assume that they are distributed according to an exponential probability density function (one that

has a long history of use in modeling response times; e.g., Luce, 1986; Townsend & Ashby, 1983). The exponential pdf has the form $f(t) = \lambda e^{-\lambda t}$ and describes a distribution that has a mean of $1/\lambda$. Panels A and B of Figure 2 present the data for serial exhaustive processing, whereas panels C and D of Figure 2 present the data for parallel exhaustive processing when the component search times are exponentially distributed instead of normally distributed.

It is easy to see that the empirical probabilities differ greatly in form across Figures 1 and 2 (panel A in each). However, if you now examine the empirical cdfs, $\hat{F}(t)$, for each of these three cases, you will see that they are all increasing. As well, the empirical survivor functions, $\hat{S}(t)$, are all decreasing. The empirical integrated hazard functions, $\hat{H}(t)$, are all increasing but are not bounded, as are the cdfs. Interpreting each of these functions across the range of the *RT* distribution is straightforward. Thus, as derived earlier, for any point in time t^* , the integrated hazard functions are telling us how much work the system has accomplished to that time. The higher the integrated hazard function, the more work the system has been able to accomplish by that time.

Making Use of the Integrated Hazard Function

We now proceed to describe two ways in which the integrated hazard function can be used to characterize capacity. The first way requires specific experimental conditions and can be used to classify a system according to the fundamental taxonomy of capacity (limited, unlimited, super) at the level of the *distribution* of task completion times. The second way can be used across a wider range of experimental conditions but does not allow for classification of performance according to the fundamental taxonomy. Instead, it can be used to compare levels of capacity across experimental conditions, stimulus types, and so forth.

How does $H(t)$ relate to an implication true of distributional statistics formulated by the second author (Townsend, 1990)? In that methodological study, it was found that an ordering in hazard functions forced an ordering in cumulative distribution functions, which forced an ordering in means (along with other statistical orderings). Although not shown there, it can be demonstrated that an ordering on the hazard function forces an ordering on the integrated hazard function, which forces an ordering on the means. Thus, $H(t)$ resides at a moderately fine-grained level in the statistical inference tree of Townsend (1990).

C(t): The Capacity Coefficient

The first of our two applications of the integrated hazard function requires a specific set of experimental conditions that can be used in a variety of psychological tasks. To illustrate these conditions, assume that we modify our memory search task as follows. First, we say that the presence of one or more of the signs

of danger is sufficient for a response. Therefore, if we find two indicators (e.g., narrowed eyes and a sneer) rather than one (e.g., a sneer), we have redundant signals of trouble. Second, we assume that we can vary the load across trials, so that we have specific frequencies of two-target and each of the one-target (sneer alone, narrowed eyes alone) trials.

The task as we have set it up allows for self-terminating processing; that is, our search of memory can stop as soon as the first sign of danger (if one is present) is located. When processing is self-terminating and both signals are targets, then the situation is said to be “first terminating” (Colonius & Vorberg, 1994) or “minimum time.” Now if we assume that the items in memory are searched in parallel, then probability theory gives us another useful identity on the integrated hazard function for task completion times. Specifically, suppose we have n items that are being processed in parallel, and in an independent fashion, in a search in which the “fastest” item to finish allows for a response to be made. Then the integrated hazard function for the task when all n items are being processed is equal to the sum of the integrated hazard functions for processing each of the items separately (see Townsend & Ashby, 1983, pp. 248–250). That is,

$$H_n(t) = \sum_{i=1}^n H_{i,n}(t) \quad (7)$$

where i indexes the i th item in the set of n to be processed.

In our example task, since $n = 1$ or 2 , there are two danger signals that can be processed. Let $i = 1$ be the narrowed eyes alone, $i = 2$ be the sneer alone. Then the integrated hazard function for the task when one of the danger signals is present would be $H_1(t) = H_{1,1}(t)$ for the narrowed eyes alone or $H_2(t) = H_{2,1}(t)$ for the sneer alone. When both features are present in the stimulus, the general formula (Equation 7) becomes

$$H_2(t) = H_{1,2}(t) + H_{2,2}(t). \quad (8)$$

This will be true whenever the features are processed independently in parallel with a stopping rule that allows for self-termination. Now assume further that each of the feature channels is just as efficient when both features are present as when only its single feature is present—the so-called unlimited capacity assumption. Then Equation 8 becomes

$$H_2(t) = H_{1,1}(t) + H_{2,1}(t). \quad (9)$$

What Equation 9 is telling us is that, when capacity is unlimited, the amount of work that can be done with two inputs together is equal to the sum of the amounts that can be done with each of the inputs alone. Essentially, the amount of work that can be done per item is the same for the two-input case as it is for the one-input case: Even with an increase in workload from one feature to two, the amount of work that can be done (on a per-item basis) does not change. We form our measure

of capacity by dividing the integrated hazard function when both features are present, $H_2(t)$, by the sum of the individual, single-feature integrated hazard functions:

$$C(t) = \frac{H_2(t)}{H_{1,2}(t) + H_{2,2}(t)}. \quad (10)$$

More generally, for n targets being processed in parallel, under a self-terminating stopping rule,

$$C(t) = \frac{H_n(t)}{\sum_{i=1}^n H_{i,n}(t)}. \quad (11)$$

Now consider the interpretation of $C(t)$ in a little more detail. In the example we have been developing, $H_2(t) = H_{1,2}(t) + H_{2,2}(t)$, the numerator and denominator of Equation 10 would be equal, and the ratio would be equal to 1. As we mentioned earlier, if this is true, then this system is capable of doing twice as much work with two stimulus components as it is with one component: Having two elements to work with neither increases nor decreases its efficiency on either item separately, and this has a natural correspondence to the notion of unlimited capacity.

If, however, the presence of two stimulus components actually decreases the ability of the system to accomplish processing (relative to the case in which only one stimulus component is present), then the integrated hazard function in the numerator will be less than the sum of the two integrated hazard functions in the denominator, and the ratio in Equation 10 will be less than 1. This intuitively represents the condition of limited capacity. Finally, if having two stimulus components present increases the ability of the system to accomplish processing, then the integrated hazard function in the numerator will be greater than the sum of the two integrated hazard functions in the denominator, and the ratio in Equation 8 will be greater than 1. This outcome intuitively represents the condition of super capacity processing.

Using C(t): An Example

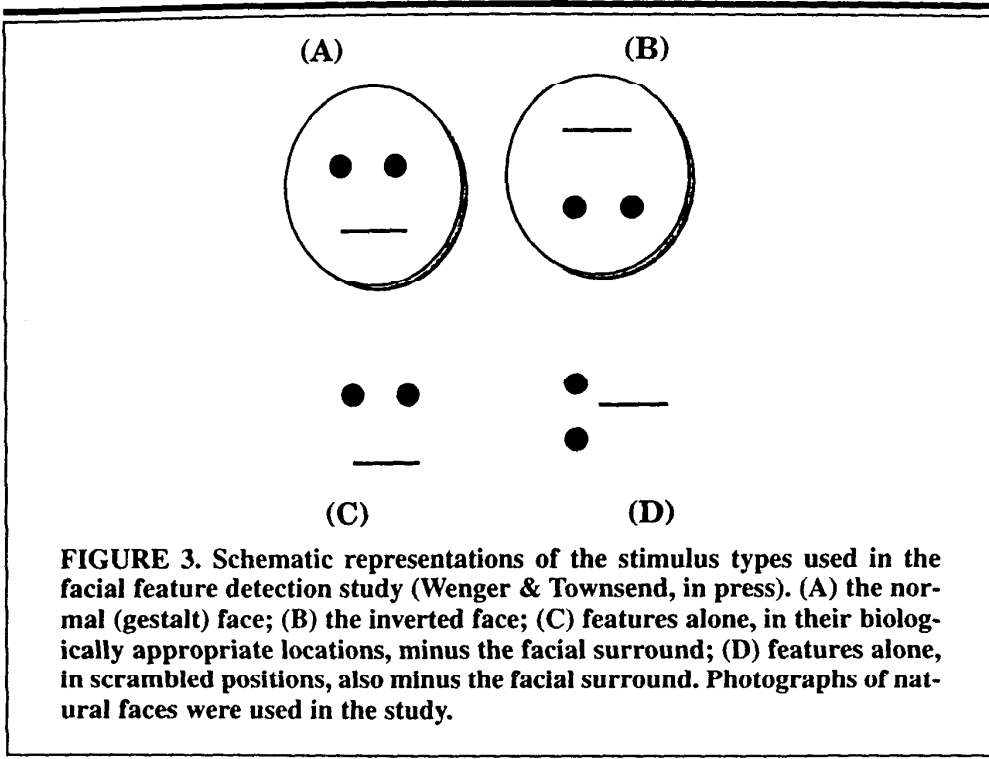
To illustrate the use of $C(t)$, consider the proposal that faces, as visual stimuli, might be processed as undifferentiated wholes rather than as sets of features (e.g., Farah, Wilson, Drain, & Tanaka, 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Implicit in this proposal is the possibility that having a full, well-configured face may be “better” (in some “gestalt” sense) than having some of the features. Said another way, having the entire face (i.e., all of the features) may be advantageous, relative to having just some of the component features (e.g., Tanaka & Farah; Tanaka & Sengco), particularly if they are in their biologically appropriate locations within the context of a facial surround. Because increasing

the number of features can be seen as an increase in workload, this proposal implies (among other things) that increasing workload should either have no effect or should improve performance. Essentially, this is a proposal for unlimited or super capacity processing.

We investigated this possibility in an earlier study (reported in detail in Wenger & Townsend, in press) in which we examined the ability to detect two anatomical features of a face: the eyes and mouth.⁶ A major goal of that study was to examine capacity (in addition to architecture and stopping rule) as a function of the number and degree of organization of the features. To this end, we examined performance across four stimulus types that (according to the literature on facial processing) might systematically alter the degree to which an observer would process the face as a well-organized whole (or gestalt). Figure 3 contains schematic representations of the stimulus types used (photographs of natural faces were used to construct the stimuli in Wenger & Townsend, in press). In the first (baseline) condition (panel A of Figure 3), both features, when present, were placed in their biologically appropriate positions within a facial surround. In the second condition (panel B of Figure 3), the features (when present) were in their biologically appropriate positions, but the entire stimulus was inverted, a manipulation that typically disrupts performance (e.g., Farah, Tanaka, & Drain, 1995; Yin, 1969). In the third condition (panel C of Figure 3), the features were present in the same relative positions as in the first condition, but the facial surround was absent. Finally, in the fourth condition (panel D of Figure 3), when the stimulus features were present, they were placed in biologically inappropriate locations (their eyes were rotated 90° counterclockwise as a pair and placed at the right side of the stimulus, and the mouth was moved up and centered vertically between the eyes). One type of trial presented both features, two types presented either one feature or the other but not both, and one type presented neither feature. Observers were instructed to respond affirmatively if at least one of the target features (pair of eyes or mouth) was present. Thus, the task was set up to allow for first-terminating processing of the features.

Four observers each performed this task with sufficient frequency so that a reliable estimate of the *RT* distribution was obtained for each stimulus type for each one of the observers. For present purposes, the more important result was the one related to the capacity coefficient (Equation 10). Figure 4 presents the *C(t)* data for one observer in this study (the data for the other three observers were quite similar). What was striking in these data was the consistent evidence for mild to moderate capacity limitations in processing, even with the gestalt (intact face) stimulus. In fact, for none of the observers was it the case that the intact

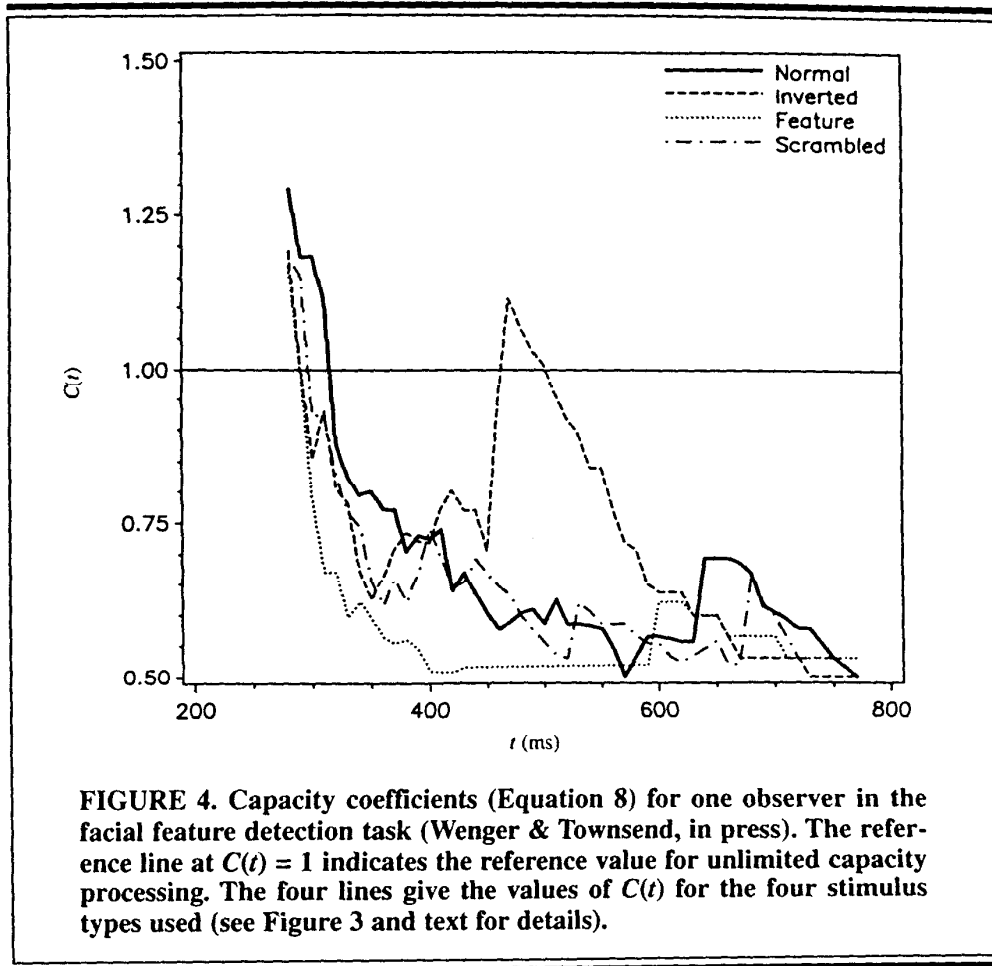
⁶This study also included other measures that provided converging evidence relevant to process architecture and stopping rule that cannot be detailed here (see Wenger & Townsend, in press).



face allowed for the highest capacity processing (see Wenger & Townsend, in press, for additional measures and details).

Making these findings regarding the capacity for processing the faces all the more intriguing was the fact that the data from all four observers supported a process architecture that is referred to as *coactivation* (e.g., Miller, 1982, 1991). Coactivation is a type of parallel processing in which the inputs are processed in parallel and then their outputs are combined in a single channel, which is then fed to a decisional operator.⁷ Such systems have been shown to be super capacity if the individual channels process their own inputs just as efficiently when there are two targets present as when there is just one (i.e., each individual channel is unlimited capacity; Townsend & Nozawa, 1995). The fact that other information points to coactivation (evidence that we cannot delve into here; see Townsend & Nozawa), but overall capacity is demonstrated to be limited—that is, $C(t) < 1$ —suggests that, in this experiment, the individual channels must be suffering quite a bit of degradation when moving from one target feature to two.

⁷As often happens with important concepts in psychology, the original notion of coactivation was primarily defined by certain experimental results, in this case by Miller (e.g., 1982). Later efforts, including some by Miller himself, offered a rigorous summation-of-activation interpretation of this concept (e.g., Colonius & Townsend, 1997; Diederich, 1995; Miller, 1991; Townsend & Nozawa, 1995).



The finding of consistent limitations in capacity for a coactive processing architecture is, in and of itself, intriguing. However, the point to emphasize for present purposes is the implication that the finding of capacity limitations holds for theories of face perception (e.g., Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997; see also discussions in Wenger & Townsend, in press). It appears that in a detection task of the present type, faces are not only processed under limited capacity but are also not necessarily superior in process capacity to less organized sets of features.

Comparing Capacity by Using $H(t)$: A Second Example

When experimenters vary the load (n) of things to process in a setting where first-terminating times are possible, they are in a position to compute $C(t)$ and thereby classify the system of interest according to the fundamental taxonomy of capacity. However, outside of such situations, it is still feasible to compare levels of capacity in distinct experimental conditions by way of the integrated haz-

ard function, $H(t)$. We pause here to note that the residual time components (i.e., those “extra” psychological mechanisms outside the processes under study) do not compromise the use of $H(t)$ to compare overall performance across conditions or observers.

Our application of $H(t)$ illustrates the second of our two approaches to characterizing capacity, in this case in a feature search task (details reported in Townsend & Wenger, 1999). Observers in this second study learned two faces, one arbitrarily designated as containing the target set of features (left eye, right eye, nose, and mouth), and the other arbitrarily designated as containing the nontarget set. On any given trial, observers were presented with a stimulus containing between two and four features. Of the features present in the test stimulus, between zero and four features were drawn from the target set, with the remainder being drawn from the nontarget set.

Figure 5 contains schematic examples of the different stimulus types; as in the preceding study (Wenger & Townsend, in press), natural photographs were used as the basis for the actual stimuli. The features that were present could be arranged in one of three configurations. In the *source-consistent gestalt* organization (panel A of Figure 5), all of the features present were drawn from the same set (either target or nontarget) and were located in their biologically appropriate positions. In the *source-inconsistent gestalt* organization (panel B of Figure 5), the features were heterogeneous with respect to the target and nontarget sets, and were located in their biologically appropriate positions. Both of these organizations resulted in very natural-looking faces, though the features would come from two different sources. In the *nongestalt* organization (panel C of Figure 5), the features were heterogeneous with respect to the target and nontarget sets and were arranged so that the features were not in their biologically appropriate positions.

In addition to this stimulus organization manipulation, we ran half of the blocks as first-terminating trials; that is, observers were instructed to give a detection response if *any* of the features present were from the target set. For the other half of the blocks, we instructed observers to use an exhaustive stopping rule. That is, they were told to give a detection response only when *all* of the features present were drawn from the target set.⁸ We ran a sufficient number of trials with each observer so that we could estimate the *RT* distributions for each stimulus type for each observer at each level of the design. We were then able to conduct all of our analyses on the data for each of the observers separately.

⁸In addition, we repeated all of the various trial conditions using two 4-letter words rather than two 4-feature faces as stimuli. For present purposes, we concentrate on the data from the trials that involved faces; the full set of results is available in Townsend and Wenger (1999). Although we present this study as one investigating visual search, there are more than likely contributions from memory involved in the performance of this task (e.g., see Hillstrom & Logan, 1998). Although important for modeling search performance, such issues do not complicate interpretation of the capacity data that we present here.

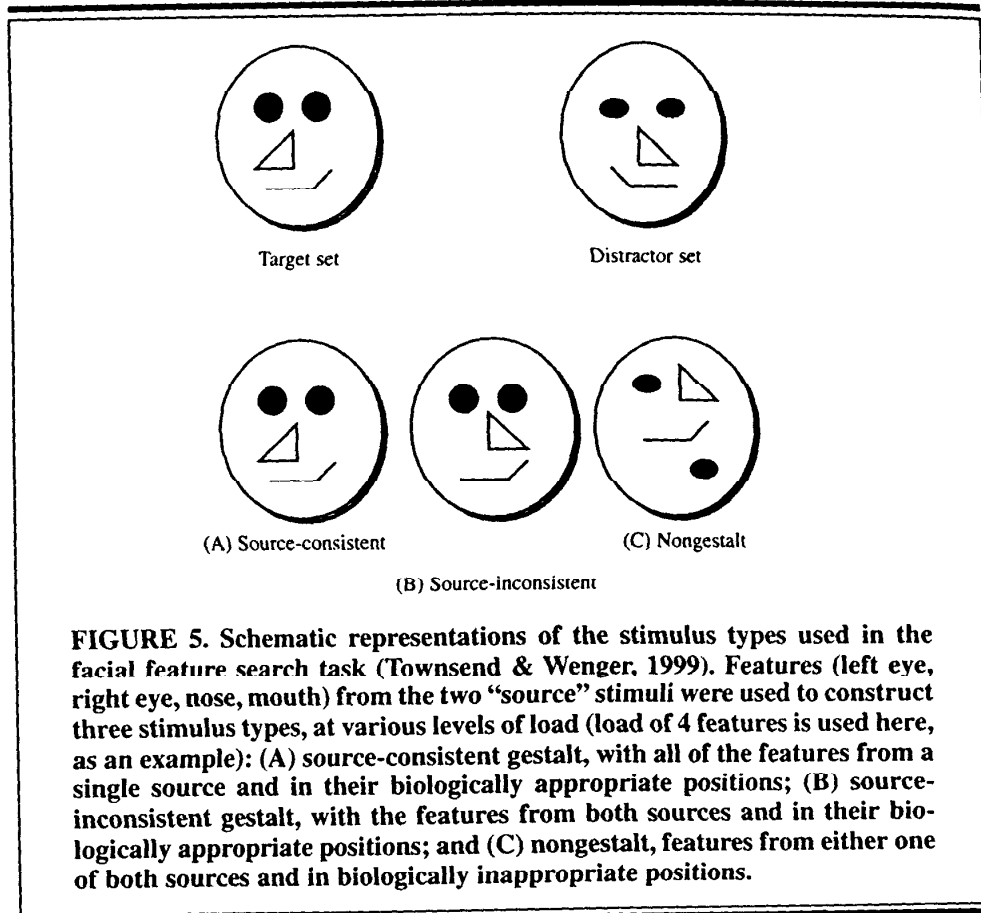


FIGURE 5. Schematic representations of the stimulus types used in the facial feature search task (Townsend & Wenger, 1999). Features (left eye, right eye, nose, mouth) from the two “source” stimuli were used to construct three stimulus types, at various levels of load (load of 4 features is used here, as an example): (A) source-consistent gestalt, with all of the features from a single source and in their biologically appropriate positions; (B) source-inconsistent gestalt, with the features from both sources and in their biologically appropriate positions; and (C) nongestalt, features from either one of both sources and in biologically inappropriate positions.

Our main concern in this study (Townsend & Wenger, 1999) was to ascertain whether stimulus organization had any impact on processing capacity and if so, what kind. Our primary tool was the use of integrated hazard functions for each of the different conditions to form a pair of ratios:

$$R_1 = \frac{H_{e,sc}(t)}{H_{e,ng}(t)} \quad (12)$$

$$R_2 = \frac{H_{e,si}(t)}{H_{e,ng}(t)} \quad (13)$$

In these ratios, the subscript e indexes the number of elements in the stimulus ($e = 2, 3, 4$), sc indicates the source-consistent gestalt condition, si denotes the source-inconsistent gestalt condition, and ng denotes the nongestalt condition. The first ratio (Equation 12) indicated the impact of the source-consistent gestalt organization, relative to the nongestalt organization, for a given number of ele-

ments. The second ratio (Equation 13) indicated the impact of the source-inconsistent gestalt organization, relative to the nongestalt organization, for a given number of elements. Both of these ratios allowed us to examine the effect of facial organization on visual search performance. If preserving facial organization aided capacity in processing, then both ratios would be predicted to be greater than 1, because, if this were true, the ability of the observers to accomplish work with some given number of stimulus features would be greater with the organization preserved (in the source-consistent and source-inconsistent gestalt trials) than with it violated (in the nongestalt trials). The difference would be reflected by having the integrated hazard functions in the first two conditions be greater than the integrated hazard functions in the last condition. Stated another way, $H_{e,sc}(t) > H_{e,n}(t)$ and $H_{e,si}(t) > H_{e,n}(t)$. The resulting ratios (Equations 12 and 13) would thus both be greater than 1.

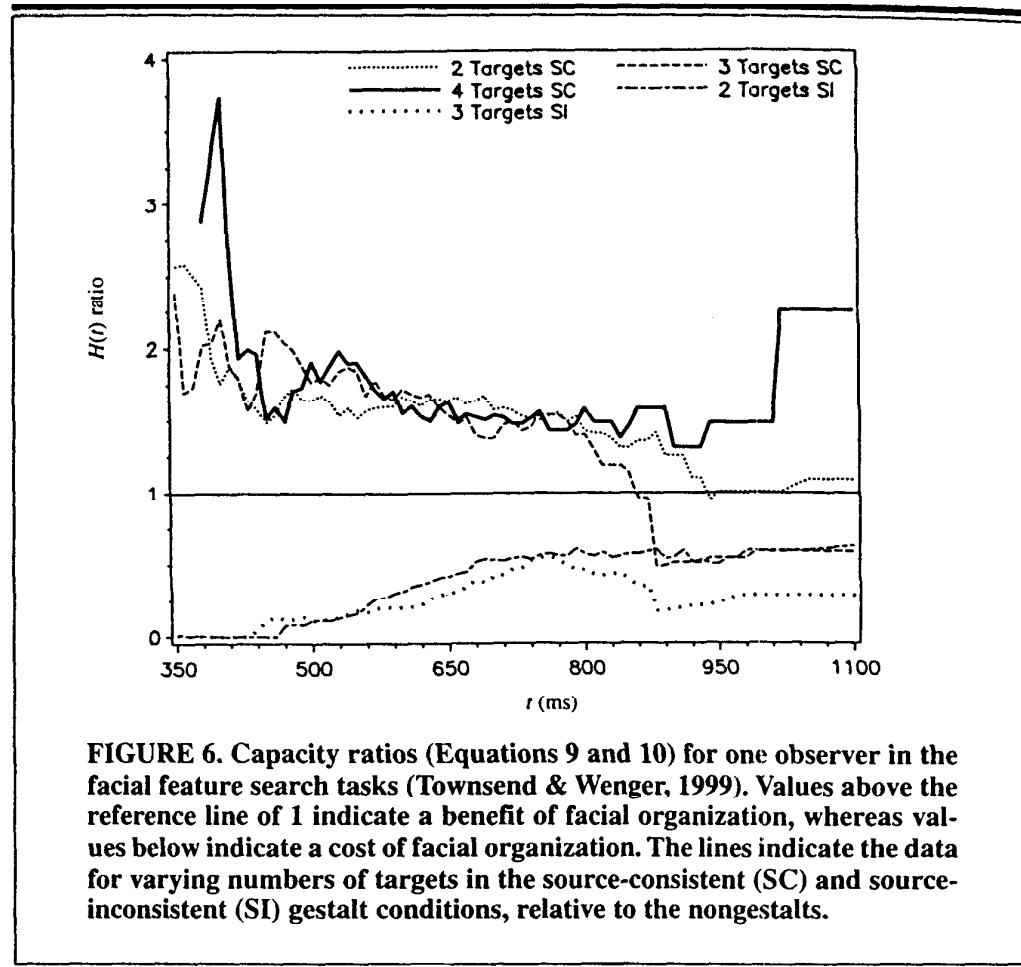
But there was also the possibility that facial organization might not always produce benefits in performance. This possibility was suggested by the findings we obtained in the detection task summarized earlier (see Wenger & Townsend, in press) and by previous work with search tasks in which faces were the target stimuli (see in particular, Kuehn & Jolicouer, 1994; Suzuki & Cavanagh, 1995).⁹ In particular, we were interested in the possibility that having a facial organization for a combination of target and nontarget features might actually impose a decrement in capacity. In this case, the second of our two ratios (the one comparing the source-inconsistent to the nongestalt) would be predicted to be less than 1.

Figure 6 plots the values of these two ratios from the first-terminating trials for one of the four observers in this study (the results were representative of all four observers; see Townsend & Wenger, 1999). As can be seen in Figure 6, the value of the first ratio was consistently greater than 1. That value indicated that facial organization, when the elements were all from the same source, increased processing capacity relative to the condition in which the same number of elements were present but were not in their biologically appropriate locations. However, the value of the second ratio was consistently less than 1. This indicated that facial organization could actually decrease processing capacity, relative to the nongestalt condition, when the elements were from two different sources. Such a result is a challenge to views of facial processing that suggest that facial organization must inevitably impart benefit. Indeed, it seems to do this only when the components are, in some sense, in agreement.

Comments on the Two Examples

It is important to note that $H(t)$ can be used to compare processing efficiency across any two conditions, as long as stopping rule and load are equated. In con-

⁹Note that these latter studies used a design qualitatively distinct from ours, although the general aims were similar.



trast, $C(t)$ by its very nature compares the integrated hazard functions obtained in a specific experimental milieu. Thus, if investigators are interested in comparing stimulus forms in terms of capacity, it will behoove them to ensure that both conditions use the same stopping rule and load. In the examples we have presented here, $C(t)$ measures effects across load within the same level of facial organization (in the first study; Wenger & Townsend, in press), whereas $H(t)$ was used to compare processing efficiency for the same load across facial organization (in the second study; Townsend & Wenger, 1999). Obviously, these are not equivalent. Thus, it is possible that load could have smaller effects in one type of stimuli than in another but that given a certain load, the first might possess larger $H(t)$ than the second.

Although an in-depth consideration of the results from our two studies is found in the cited papers, we should take a little space to outline our conclusions, particularly as they reflect on issues of capacity. The first investigation (Wenger & Townsend, in press) found that facial structure did not lead to super capacity or even unlimited capacity as measured by $C(t)$. Although the presence of both features rather than one led to improvement (i.e., two targets were processed faster

than one at the level of the RT distribution), there was less improvement than could have been expected on the basis of statistical facilitation in an unlimited capacity, independent parallel system. As noted earlier, it seems of considerable interest to learn exactly when and where (and, we hope, ultimately how) facial structure helps or hinders performance in a variety of perceptual and cognitive tasks. One simply does not know beforehand at what level facial structure will interfere with, and at what level it will facilitate, performance. For instance, few psychologists would likely have predicted the Wheeler-Reicher effect in which letters are recognized more accurately in the context of words than by themselves, even when ordinary contextual informed guessing is ruled out (Reicher, 1969; Wheeler, 1970). At this point, it is important to carry out a hierarchical set of studies to locate where hindrance and aid to processing in face perception occurs. Perhaps future work will bring a theory that makes a priori predictions.

We have already observed in the two studies just reviewed that some pivotal change in conditions has led from no help, and possibly mild hindrance, in capacity in the first study to conditions that help or hurt capacity in the second study. Although we concentrated on $C(t)$ in the first study and on $H(t)$ in the second study, both measurements were, in fact, made in both sets of experiments, and the conclusions were compatible with those noted here. What made the difference in the findings? Both experiments had well-trained observers and photographs of real faces. The only obvious differences were that the second study called for the observer to discriminate between distractor-face features and target-face features; the latter features were taken from a single coherent individual face, whereas the first study required the observer to detect the presence or absence of the designated features. It is possible that observers came to think of the normal face (panel A in Figure 3) as something akin to an individual and that their ability to recognize his or her features was improved relative to their ability to recognize less well-organized sets of the same features. On the other hand, it appears that within the source-inconsistent gestalt stimulus (panel B in Figure 5), facial organization actually caused performance worse than that based on randomized locations of the features. Although we are far from a comprehensive explanation for such findings, we should emphasize that the questions about such patterns of effects could not have been raised in the absence of measures of capacity, such as the ones we have presented.

A Do-it-Yourselfer's Guide to Constructing and Using Integrated Hazard Functions

At this point, we hope to have demonstrated that the empirical integrated hazard function $\hat{H}(t)$ can be used to characterize processing capacity in a fine-grained fashion in a variety of ways. By using the integrated hazard function, one can, in a direct way, capture important hypotheses about aspects of process capacity in a wide range of tasks. We conclude by detailing the steps one would need to execute in order to make use of the integrated hazard function.

To illustrate these steps, we will consider the performance of a system designed to detect either one or two features; the system is composed of a pair of processing channels. We will (consistent with the examples used in the preceding sections) assume that these features are the anatomical features of natural faces (e.g., the eyes and mouth) and that the task allows a detection response to be given when either or both of these features are present (i.e., allowing for self-termination).

Assume that we are interested in exploring the hypothesis that facial organization benefits capacity when the constituent features are drawn from the same face. In addition, assume that this same organization robs capacity when the features are from two different faces, as is consistent with the earlier reported findings. We will model these two situations by allowing a positive correlation in the processing rates of the two channels when the features are in agreement and a negative correlation in the rates when the features are not in agreement. Finally, assume that we have an experimental setup that will allow us to obtain a sufficient number of observations for each of the stimulus conditions to aid us in obtaining a robust estimate of the *RT* distribution. As a rule of thumb, we prefer to obtain 250 to 300 observations for each stimulus. Having collected the data, one would then do the following:

1. For each stimulus condition, remove the *RT*s corresponding to errors, anticipatory responses, equipment failures, and lapses of observer attention. As a rule of thumb, such errors should represent no more than 5–10% of the total responses.¹⁰
2. Determine the size of each *RT* bin and the number of bins that will be needed. We typically use 10-ms time bins and cover the range between 100 and 3,000 ms. The range and the size of the bins, of course, will vary according to the nature of the task being used.
3. Assign each *RT* in each of the stimulus conditions to a time bin. This can be done by dividing the obtained *RT* by the bin size, rounding to the nearest integer, then multiplying by the bin size. For example, with a bin size of 10 ms, an obtained *RT* of 379 ms would be assigned to the 380 ms time bin ($379/10 = 37.9$, round to $38 \times 10 = 380$).
4. Calculate the number of observations in each bin in each condition, along with the total number of observations in each condition.¹¹
5. Calculate the empirical probability of observing each time bin in each condition by dividing the total number of observations in the bin by the

¹⁰There are both theoretical and pragmatic reasons for setting such a threshold. First, it is not altogether clear the extent to which *RT* measures in lower accuracy conditions might be interpretable with respect to the issues discussed here (e.g., see Wenger & Townsend, in press). Second, with error rates higher than this, one significantly increases the amount of time required to collect a sufficient number of correct responses. Finally, one can conceivably make use of the capacity measures we have described for the distribution of error *RT*s.

¹¹In our applications, we construct all distributions and carry out all transformations using SAS (e.g., PROC FREQ).

total number of observations in the stimulus condition. This gives $\hat{f}(t)$ for each of the bins. At this point, one should check that the probabilities have been calculated correctly by summing all of the bin probabilities for a given condition; this sum should be 1.

6. Calculate the empirical cdf $\hat{F}(t)$ for each stimulus condition by accumulating the empirical probabilities from the lowest to the highest valued bin.
7. Calculate the empirical survivor function $\hat{S}(t)$ in each condition by subtracting the value of $\hat{F}(t)$ for each bin from 1.
8. Calculate the empirical integrated hazard function $\hat{H}(t)$ by taking the negative of the natural logarithm of the value of $\hat{S}(t)$ for each bin in each condition. To avoid numerical errors, set the value of $\hat{S}(t)$ for the lowest bin to some arbitrary value near 0 (i.e., .0001), and set the value of $\hat{S}(t)$ for the highest bin to some arbitrary value near 1 (i.e., .9999).

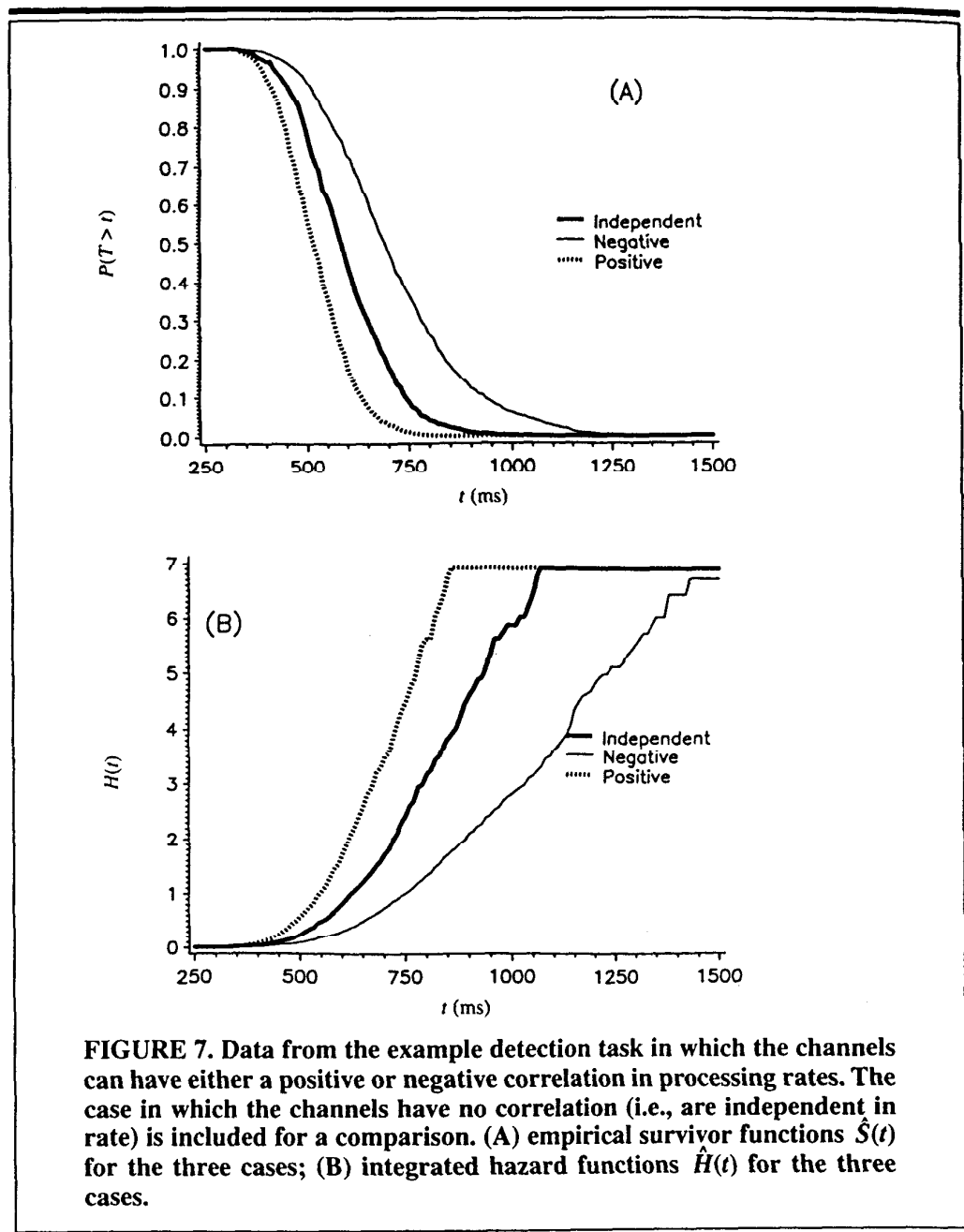
Figure 7 plots the summary functions $\hat{S}(t)$ and $\hat{H}(t)$ for this example task for the cases in which the channels have either a positive or a negative correlation. We have also included the case in which the channels have a 0 correlation (i.e., the rates are independent). We leave it to readers to explore how these data might be used as the basis for evaluating the effects of stimulus organization on processing capacity.

Conclusions

The notion of capacity has played a valuable role in theories about attention, perception, memory, and other cognitive operations. Indeed, the history of the use of this construct is a long one. However, it has not generally been the case that the construct has had a precise representation in measures of performance. In this article, we presented a natural way to represent the notion of capacity in response times, by way of the integrated hazard function. We demonstrated how the integrated hazard function could be used as the basis for characterizing capacity both in terms of a within-stimulus-type function of load as well as a general comparison of capacity between different experimental conditions of almost any kind. And we provided a set of easy steps for making use of these measures in experimental applications.

One obvious region of potential application lies in psychophysics. Psychophysics and sensory science almost exclusively use accuracy rather than RT in their investigations, particularly in comparisons of multiple stimulation vs. single stimulation (e.g., binocular or dichoptic vs. monocular stimulation).¹² As pointed out by Hughes and Townsend (1998; see also Townsend & Nozawa,

¹²Interestingly, one of the first RT studies was, however, carried out in a psychophysical environment (see Raab, 1962).



1995), accuracy by itself can be quite uninformative about underlying process structure, even in such pristine regions as sensory science. The measures suggested here can potentially be very beneficial and supplementary to accuracy statistics and to learning about the details of sensory architecture and process rules.

In our example studies, we varied the load within blocks but divided attention designs where the load or instructions (e.g., “only features a, b, c are relevant; d, e, f are to be ignored”) can use our measures just as easily. There are a

number of areas of possible application in cognition lying outside the usual purview of sensation and perception. For instance, we might expect the flat or near-flat mean *RT* functions that appear with extended practice in automatization experiments (e.g., Logan, 1988, 1992; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) to suggest extreme super capacity at a fine-grain level. In the domain of memory, Anderson (1976; Ross & Anderson, 1981; cf. Townsend, 1974) has tested for parallel processing in sentence retrieval experiments, whereas Wenger (1999) has examined questions of process architecture in memory retrieval in cognitive skill. It should prove interesting to assess capacity in the foregoing domains. Finally, although we have focused on capacity as being of interest in its own right in the present study, it is clear that our measures of capacity are synergistic with the other dimensions, such as architecture, stopping rule, and dependence relationships.

REFERENCES

- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). San Diego, CA: Academic Press.
- Bonnell, A. M., & Hafter, E. R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception & Psychophysics*, *60*, 179–190.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.
- Cheal, M., & Gregory, M. (1997). Evidence of limited capacity and noise reduction with single-element displays in the locating-cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 51–71.
- Colonius, H. (1990). A note on the stop-signal paradigm, or how to observe the unobservable. *Psychological Review*, *97*, 309–312.
- Colonius, H., & Townsend, J. T. (1997). Activation-state representation of models for the redundant signals effect. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 245–254). Mahwah, NJ: Erlbaum.
- Colonius, H., & Vorberg, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology*, *38*, 35–58.
- de Haan, E., Lutz, C., & Noest, A. (1998). Nonspatial visual attention explained by spatial attention plus limited storage. *Perception*, *25*, 591–608.
- Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, *39*, 197–215.
- Farah, M. J., Tanaka, J. N., & Drain, M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception & Performance*, *21*, 628–634.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review*, *105*, 482–498.
- Henderson, J. M. (1996). Spatial precues affect target discrimination in the absence of visual noise. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 780–787.
- Hillstrom, A. P., & Logan, G. D. (1998). Decomposing visual search: Evidence of multiple item-specific skills. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1385–1398.
- Hughes, H. C., & Townsend, J. T. (1998). Varieties of binocular interaction in human vision. *Psychological Science*, *9*, 53–60.

- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kantowitz, B. H., & Knight, J. L. (1976). On experimenter-limited processes. *Psychological Review*, *83*, 502–507.
- Kuehn, S. M., & Jolicouer, P. (1994). Impact of the quality of the image, orientation, and similarity of the stimuli on visual search for faces. *Perception*, *23*, 95–122.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 451–468.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.
- Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 883–914.
- Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295–327.
- Luce, R. D. (1986). *Reaction times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Luck, S. J., Hillyard, S. A., Mouloua, M., & Hawkins, H. L. (1996). Mechanisms of visual-spatial attention: Resource allocation or uncertainty reduction. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 725–737.
- Marsh, R. L., & Hicks, J. L. (1998). Event-based prospective memory and executive control of working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 336–349.
- Miller, J. O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279.
- 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.
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instruction. *Quarterly Journal of Experimental Psychology*, *11*, 56–60.
- Navon, D. (1984). Resources—a theoretical soup stone? *Psychological Review*, *91*, 216–234.
- Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. *Psychological Review*, *86*, 214–255.
- Norman, D. (1968). Toward a theory of memory and attention. *Psychological Review*, *75*, 522–536.
- Nyberg, L., Nilsson, L. G., Olofsson, U., & Baeckman, L. (1998). Effects of division of attention during encoding and retrieval on age differences in episodic memory. *Experimental Aging Research*, *23*, 137–143.
- Parzen, E. (1960). *Modern probability theory and its applications*. New York: Wiley.
- Payne, D. G., & Wenger, M. J. (1998). *Cognitive psychology*. Boston: Houghton-Mifflin.
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590.
- 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.
- Ross, B. H., & Anderson, J. R. (1981). A test of parallel versus serial processing applied to memory retrieval. *Journal of Mathematical Psychology*, *24*, 183–223.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.
- Schweickert, R., & Boggs, G. J. (1984). Models of central capacity and concurrency. *Jour-*

- nal of Mathematical Psychology*, 28, 223–281.
- Schweickert, R., Fisher, D. L., & Goldstein, W. M. (1992). *General latent network theory: Structural and quantitative analysis of networks of cognitive processes* (Technical Rep. No. 92-1). West Lafayette, IN: Purdue University Mathematical Psychology Program.
- Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception & Performance*, 14, 591–600.
- Shiffrin, R. M. (1975). The locus and role of attention in memory systems. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. 5, pp. 168–193). San Diego, CA: Academic Press.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: Vol. 2. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127–190.
- Smith, P. L. (1990). Obtaining meaningful results from Fourier deconvolution of reaction time data. *Psychological Bulletin*, 108, 533–550.
- Sperling, G. (1960). The information available in a brief visual display. *Psychological Monographs*, 74(Whole No. 498).
- Sperling, G. (1963). A model for visual memory tasks. *Human Factors*, 5, 19–31.
- 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*, 46A, 225–245.
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, 25, 583–592.
- Tellinghuisen, D. J., Zimba, L. D., & Robin, D. A. (1997). Endogenous visuospatial precuing effects as a function of age and task demands. *Perception & Psychophysics*, 58, 947–958.
- Theios, J. (1973). Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7, pp. 43–85). San Diego, CA: Academic Press.
- 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. 163–168). Hillsdale, NJ: Erlbaum.
- Townsend, J. T. (1981). Some characteristics of visual whole report behavior. *Acta Psychologica*, 47, 149–173.
- Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108, 551–567.
- Townsend, J. T. (1992). On the proper scales for reaction time. In H.-G. Geissler, S. W. Link, & J. T. Townsend (Eds.), *Cognition, information processing, and psychophysics: Basic issues* (pp. 105–120). 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, UK: Cambridge University press.
- 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., & Wenger, M. J. (1999). *On the costs and benefits of faces and words*. Manuscript in preparation.
- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 77, 533–546.
- Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18, 293–297.
- Wenger, M. J. (1999). On the whats and hows of retrieval in the acquisition of a simple cognitive skill. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 1137–1160.
- Wenger, M. J., & Townsend, J. T. (in press). Faces as gestalt stimuli: Process characteristics. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition*. Hillsdale, NJ: Erlbaum.
- Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1, 59–85.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81, 141–145.

APPENDIX Important Notation

\overline{RT}	mean RT
T, t	the random variable for task or process time and any particular value of that random variable
$p(t), f(t)$	theoretical probability mass and density functions on the random variable for task or process time
$\hat{f}(t)$	empirical probability density function on task or process time
$F(t), \hat{F}(t)$	theoretical and empirical cumulative distribution function, $F(t) = P(T \leq t)$, or the probability that the task or process is done at or before time t
$S(t), \hat{S}(t)$	theoretical and empirical survivor function, $S(t) = P(T > t)$, or the probability that the task or process is not done by time t
$E[T]$	expected value of the random variable T , the mean value for the duration of the task or process
$h(t)$	intensity or hazard function, $h(t) = f(t)/S(t) = f(t)/[1 - F(t)]$, or the conditional probability that the process or task will finish in the next instant, given that it is not yet finished
$H(t), \hat{H}(t)$	theoretical and empirical integrated hazard function,
	$H(t^*) = \int_0^{t^*} h(t) dt = -\log[S(t)]$
$C(t)$	capacity coefficient for parallel self-terminating processing, for n items,

$$C(t) = \frac{H_n(t)}{\sum_{i=1}^n H_{i,n}(t)}$$

A value of 1 indicates unlimited capacity, a value less than 1 indicates limited capacity, and a value greater than 1 indicates super capacity.

Manuscript received October 30, 1998
Revision accepted for publication February 5, 1999