COGNITIVE THEORY
Volume 3

edited by

N. John Castellan, Jr.
Frank Restle

with contributing editors

Michael H. Birnbaum
Stephen W. Link
George R. Potts

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1978 Hillsdale, New Jersey

DISTRIBUTED BY THE HALSTED PRESS DIVISION OF

JOHN WILEY & SONS
New York Toronto London Sydney

Lawrence Erlbaum Associates, Inc., Publishers
62 Maria Drive
Hillsdale, New Jersey 07642

Distributed solely by Halsted Press Division

Library of Congress Cataloging in Publication Data (Revised)

Main entry under title:

Cognitive theory.

Includes bibliographies and indexes.
1. Cognition. 2. Speech perception.
3. Judgment. 4. Memory. I. Restle, Frank.
BF311.C553 153.4 75-14293
ISBN 0-470-71732-7 (v. 1)
0-470-99025-2 (v. 2)
0-470-26375-X (v. 3)

Printed in the United States of America

8

Methods of Modeling Capacity in Simple Processing Systems

James T. Townsend
F. Gregory Ashby
Purdue University

INTRODUCTION

The study of capacity, empirically and theoretically, is of considerable importance in cognitive psychology (see, e.g., Kahneman, 1973; Townsend, 1974a). Yet there are relatively few mathematical techniques within psychology designed to assess or analyze capacity, much less a systematic theory that interrelates the various and sundry uses of the concept.

Although we cannot hope to completely attain the latter goal in this chapter, we attempt to initiate development of some of the many possible branches that a mathematical knowledge-tree or theory-tree of capacity might possess.

One appropriate milieu in which to discuss and construct these branches could well be that of pattern recognition (Townsend, 1971) and multisymbol search situations (e.g., Sternberg, 1966; Townsend & Roos, 1973). The matching or comparison of patterned information from the environment with information that has previously been stored internally is a process that attracts a rather incredible amount of attention from psychologists as well as engineers and applied mathematicians. There is good reason for this interest. It is difficult to think of a psychological experiment that does not involve such processes in at least some phase of its conditions, although these may or may not be the target processes of significance to a particular study. Even the rote serial learning of a list of nonsense syllables necessitates the activation of comparisons of the input patterns with long-term and/or short-term memory. The burgeoning of the engineering and mathematical literature on pattern recognition in the last few years indicates the tremendous utility that diverse artificial recognizers can or will possess, affecting a broad spectrum of our existence in everything from biomedical devices (e.g., recognizers of disease syndromes and prosthetic reading aids for the blind) to high-altitude geophysical scanning techniques and foreigh language translation.

Our goals in this chapter, then, are first to limn a theoretical perspective of pattern matching, concentrating on a linear correlational model of feature comparison and second, but with greater emphasis, to investigate some aspects of the architecture and dynamics of capacity. The pattern matching model allows us to give a detection-type analysis within which concepts of noise, capacity, and latency, can be introduced. The subsequent decision function is described only qualitatively, because it is discussed extensively elsewhere (e.g., Smith & Spoehr, 1974; Massaro & Schmuller, 1975; Townsend & Ashby, unpublished manuscript). Similarly, except as concerns the notion of capacity, remarks about matching a single target symbol with a multisymbol display or memory set are fairly cursory, because we are ready to add nothing especially novel about other issues (for an analysis of some of these, see Townsend, 1974a).

The analyses of capacity concentrate on two approaches. One takes a simple linear system point of view, and the other is developed within the confines of stochastic latency theory. The first approach concentrates on several interesting measures of capacity, of which some are relatively pure measures of expended "energy"; others are based on accuracy or accuracy plus latency. Results from the second approach may be directly employed when accuracy is high or when the errors are presumed to be unrelated to the process in question. They may be more indirectly employed in order to give the state of completion of a set of elements at any point of time and, hence, can be employed in conjunction with decision assumptions to produce latency-accuracy relationships in time-pressured or brief-display conditions.

Several of the new capacity measures that these approaches yield are primarily useful in model construction and in comparing the capacity structure of two or more models. Others are, in addition, observable, that is, computable from easily obtained statistics on accuracy and latency. These latter may prove helpful not only in relating internal model structure and dynamics to behavioral data but also in serving as summary capacity statistics.

It is not difficult to see that these capacity concepts have a natural interpretation in terms of the pattern-matching situation under discussion. However, they are much more general than this particular application might suggest. The pattern-matching milieu, though, is of interest in its own right and serves as a convenient vehicle to explicate our theorizing about capacity.

Some of the concepts we introduce (such as "power," "energy," and "linear filter") come from engineering theory and have been employed fruitfully by others, most particularly in modeling acoustical detection mechanisms, although linear system mathematics is now burgeoning in visual psychophysics (both spatially and temporally) and, naturally, in human performance as well. However, we seek to exploit these notions in more cognitive ways and to explore some new mathematical descriptions of capacity.[1]

Our attitude toward "capacity" is in line with the philosophy formulated earlier: that capacity is a construct that can and should be applied to different levels of processing and is one that can appear in various guises and possess somewhat different implications depending on the specific experimental and theoretical circumstances (Townsend, 1974a). It would be a mistake at this point, we think, to try to fit capacity into a narrow preconceived mold–such would appear to provide only a Procrustean solution. However, this attitude does not preclude the importance of investigation of relationships between differing conceptions of capacity; for example, between capacity as "power" and capacity as the ability to deal with computational complexity. Rather, we think such research is much needed.

A LINEAR SYSTEM MODEL OF FEATURE MATCHING

Before embarking on the theoretical development, it may be apposite to mention that there are characteristics that a full mathematical model or theory will ultimately have to possess, with respect to which the present model is mute. We have in mind particularly the possibility of hierarchical mechanisms that seem to be present in visual symbolic processing (e.g., Estes, 1972; Hoffman, 1975; Weisstein, 1973). For example, it may well be that location information is extracted relatively early, that presence of identical patterns in the display is detected soon thereafter, and so on. The development here does not exclude such mechanisms, which it is thought can be intercalated in later work. We do feel, however, that processing to the degree represented herein may be sufficient for elementary visual recognition tasks.

Hereafter we use the term *symbol* to refer to a pattern that has been assigned a response and learned by the organism, whom we shall call O. Our problem is to put down a plausible account of how O goes about recognizing one of the N possible symbols (it is convenient to take for granted that O knows the size and members of the presentation set). It would be possible but tedious for this particular enterprise to totally formalize the development. Instead, we work through the main ideas with emphasis on the mathematics that is requisite for getting across needed concepts and predictions.

For starting reference observe Fig. 1, which illustrates one version of a now familiar relatively coarse breakdown of processing subsystems germane to the present discussion. A subset of features is assumed to be extracted from the

[1]Some of Lappin's developments (chapter 6 of this text) also make use of linear system concepts.

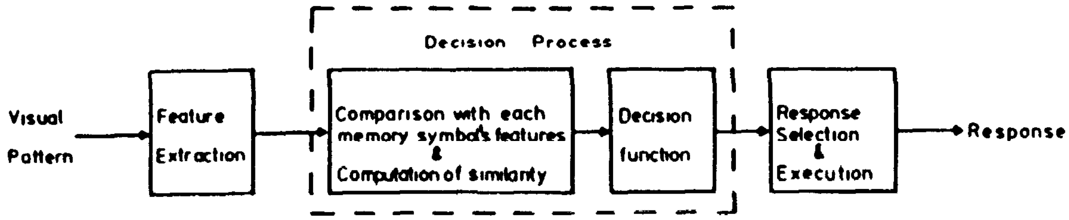

FIG. 1 Schematic of a model for feature matching.

symbol and then submitted to the decision process. Here two tasks are performed. First, this internal extracted feature list is matched in some way against the stored (presumably in long-term memory) list of all possible features. The fact that such a list exists follows from our assumption that O knows the size and members of the presentation set. We envision the second phase of the decision process as determining the identity of the presented symbol from the results of the matching phase by first assigning the various similarities a measure. This information is passed to response selection, which instructs the response execution phase as to the motor response to be associated with the presented symbol.

Townsend and Ashby (unpublished manuscript) explore a wide variety of models based on feature extraction principles. These models are very useful in the investigation of general properties of extraction, matching, and decision, but they are mute with regard to the question of exactly how the "seen" features are extracted from the stimulus, as well as to the question of how the extracted features are compared with the set of stored features that characterize the pattern set.

The extraction process is taken as a given here [but see, e.g., Weisstein (1973) for a discussion of issues at this level], and our present emphasis is on the comparison process.

To begin, we wish to propose to treat each feature as a separate signal, specifiable in principle by a waveform in time and space. The spatial representation of this waveform of course depends on the modality of the presented symbol. For example, in an auditory discrimination task, any presented symbol (e.g., any simple or complex tone) can be described as a function of time, with the dependent variable being amplitude. But the visual modality has two spatial dimensions, and thus the resulting waveform must be three-dimensional. Figure 2 shows a simple example. Assume that presentation of a straight line feature is characterized by a uniform intensity square-wave in two dimensions [as in Fig. 2 (A)], along with a time function indicating the entire temporal course of the waveform. Figure 2 (B) represents a hypothetical brief presentation of such a line. It must be remembered that the amplitude or intensity within the (x, y) plane is changing as a function of time. By taking infinitely thin slices orthogonal to the t axis, we can view this density at any desired point. Figure 2 (C) illustrates three such slices: one taken just after stimulus *presentation*, one just before, and one just after stimulus *offset*. It can be seen that $f(x, y)$, the amplitude surface, is initially small, increases, and then decreases following stimulus offset.

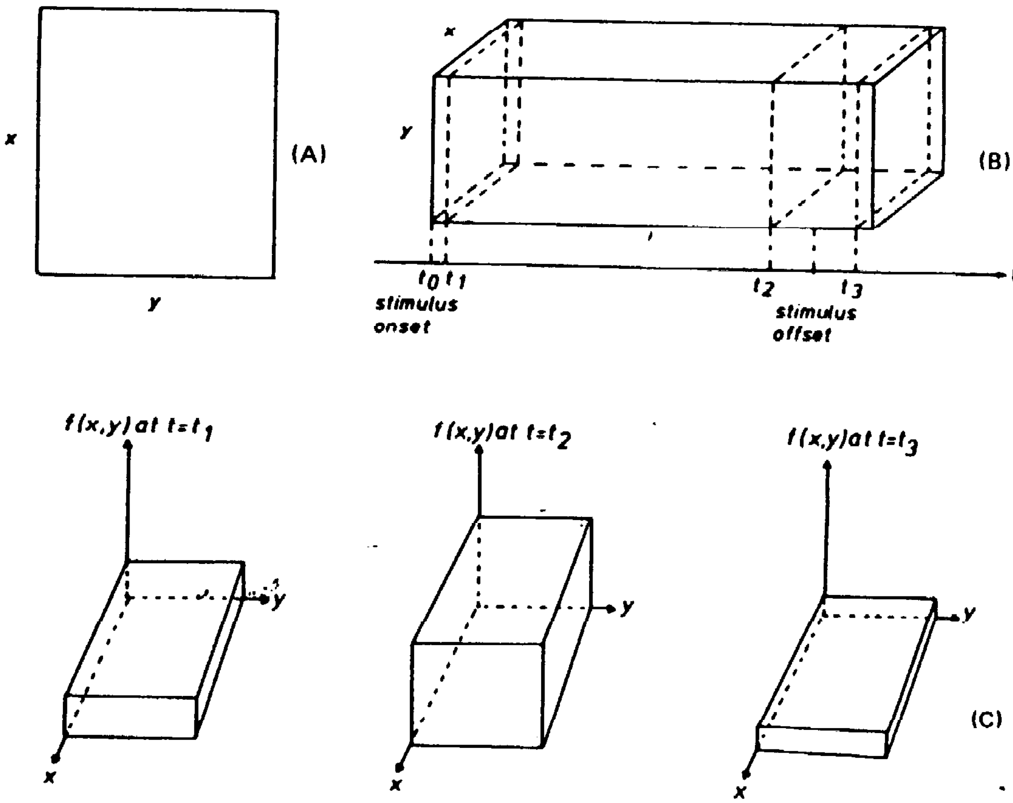

FIG. 2 An example of a three-dimensional visual waveform. (A) The characterization of a straight line feature as a square wave in two spatial dimensions. (B) The time course of a hypothetical brief presentation of this feature. (C) The amplitude surface within the (x, y) plane at three arbitrarily selected time points. These surfaces are obtained from (B) by taking slices orthogonal to the time axis at points t_1, t_2, and t_3 and then plotting amplitude, $f(x, y)$, as a function of position in the (x, y) plane.

Notice that in all cases the surface of $f(x, y)$ is flat; that is, the amplitude intensity is uniform and therefore independent of position in the (x, y) plane. In this respect Fig. (C) represents a perfect observer. In reality, we would expect some irregularities in the $f(x, y)$ surface, which could arise from such effects as neural lateral interaction and quantal fluctuations (with very weak signals).

To return to the feature-matching process, it seems reasonable to assume that O subjects this internal signal (feature) to comparisons with like signals, probably with whatever waveforms would exist in the same physical location (relative to the overall spatial extent of the symbol). We assume that O conducts this comparison by implementing a *cross-correlation* in two dimensions.[2] That is,

[2] Analogous notions of comparison arise when the pattern is represented as a Fourier (or other) transform of the original intensity function or in holographic models. For instance, comparison of holograms via interference patterns can be considered as a means of correlation or comparison (see, e.g., Smith, 1969). It is the logical structure with which we are primarily occupied here, not the ultimate resolution of the question of how visual input is coded.

we view O as taking a time series of spatial samples which are held in a visual operating memory, in the position of a given feature and then correlating these against the stored concomitants associated with the various memory symbols. Such a discrete series, sampled at the appropriate time interval, is capable of adequately describing virtually any real signal of finite duration.

From the sampling theorem of signal analysis, we know that the samples must be taken at time intervals no more than $1/2W$ seconds apart, where W is the signal bandwidth. Reconstruction of the signal using discrete sampling procedures can therefore only succeed with signals of finite bandwidth. However, this restriction appears in our case to be trivial, because it seems highly probable that O operates as a low-pass filter completely attenuating all frequency components of the input signal greater than some finite value W_0 (Cornsweet, 1970).

In such a model, reasonably long display times primarily provide for a lengthened comparison period (i.e., more samples), with the "figure" itself being quite stable except for "on" and "off" transient effects. That is, if the intensity is high and display period lengthy, the time course of the feature signal is pretty much the same. This state of affairs may be compared with that in audition where the main variation of interest is always in the time dimension (excluding spatial localization). However, it is to be expected that with short exposure times, low intensities, or small visual angles, noise will undoubtedly depend on the characteristics of the symbol, background, and any attendant mask. In the present development, we assume the standard white noise description.

It can now be assumed that the correlation comparisons within features are performed serially, although a limited capacity parallel mechanism would do as well and is no less viable presently. It is to be expected, however, that processing *across* features is parallel.

Let the simple feature sample be designated by $r_i = r_{i1}, \ldots, r_{iK}$, where r_{ij}, $j = 1, K$ is the jth of K samples (equally spaced every Δt time units) for the input from feature position i. The correlation result for memory feature s_r is, of course,

$$y_K = \sum_{j=1}^{K} r_{ij} s_{rj} . \tag{1}$$

In the continuous analogue to (1), we would have

$$y(T) = \int_0^T r_i(t) s_r(t)\, dt , \tag{2}$$

with T as the total interval of sampling. Note that to improve readability, we have treated r_{ij} as a single value rather than the two dimensional array we earlier argued for. A more detailed representation of (2) is

$$y(T) = \int_0^T \int_{\{x\}} \int_{\{y\}} r_i(x, y, t) s_r(x, y, t)\, dy\, dx\, dt .$$

In the discussion that follows, this more precise description of the cross-correlation device provides, in general, no added insight over the form of Eq. (2) and thus is dropped in favor of the simpler notation.

The correlation mechanism of (1) or (2) can easily be pictured as the result of (optimal) matched filtering. Without getting into the details of linear system theory (e.g., Padulo & Arbib, 1974), it can be noted that the output $y(T)$ of a real linear filter at time T can be written as

$$y(T) = \int_{\tau=0}^{T} h(T - \tau) r(\tau)\, d\tau , \tag{3}$$

where r is again the input, and h is the weighting the system gives the input back, from when it began [at time $\tau = 0$: $r(0)$ was the input, and $h(T)$ is the importance it has T time units later] up to the present [at time $\tau = T$: $r(T)$ is the input, and $h(0)$ is the weighting given this present input in determining the output].

When this filter is "matched" for a certain signal s_i, $h(T - \tau)$ becomes $s_i(\tau)$ or, in the discrete sampling case, s_i; and the correlation mechanism is obtained. Of course, we are likening s_i, or $s_i(\tau)$, to a tape between times 0 and $T = K\Delta t$, which is then compared with a similar memory tape; although the possibility should not be entirely discounted that the correlation is done in real time rather than later as our iconic-like tape.

Now, given the present setup, it is easy to show that with white Gaussian noise $n(t)$ or n_j is introduced, where $\mu = 0$ and $\sigma^2 = 1$ [i.e., written $N(0, 1)$], then

$$r_{ij} = \begin{cases} as_{ij} + n_j & \text{if feature } i \text{ is present}, a > 0, \\ n_j & \text{if feature } i \text{ is absent}, \end{cases} \tag{4}$$

where the separate n_j, $j = 1, K$ are uncorrelated and normally distributed, and a represents the strength of the input signal relative to the internal memory signal. The correlation result (alternatively the distribution of output y) is then

$$y_K = \sum_{j=1}^{K} r_{ij} s_{ij} \sim \begin{cases} N(a \sum_{j=1}^{K} s_{ij}^2, \sum_{j=1}^{K} s_{ij}^2) & \text{if feature } i \text{ is present}, \\ \text{or} & \\ N(0, \sum_{j=1}^{K} s_{ij}^2) & \text{if feature } i \text{ is absent}, \end{cases} \tag{5}$$

where the symbol $\sim$ means "is distributed as."

From this familiar type of description, predictions of signal-to-noise ratio, selection of an optimal decision criteria, and hit and false-alarm frequencies can be computed. The reader should keep in mind those aspects of the foregoing structure contained in the input symbol and also in a given memory symbol.

The previous cases [Eq. (5)] were only those when the feature i was present in the memory symbol and was or was not in the input symbol. The other two cases, where there is no feature present in the memory symbol, are not so obvious. Assumptions must be made as to what happens when the input (with or without the feature) is correlated with "nothing." Although there are a number of possible tacks one could take, let us presume that "nothing" is representable as a constant background effect (or gain) and therefore as a constant factor b in the correlation; that is, $h(t) = bs_i$, or $h(t) = b$, depending on whether feature i exists in memory or does not.

Then, we acquire the complement to Eq. (5), which completes the set of four possibilities:

$$
y_K = \sum_{j=1}^{K} h(K\Delta t - j\Delta t) r(j\Delta t) \sim
\begin{array}{l|l|l}
 & \multicolumn{2}{c}{\textit{Feature } i} \\
 & \textit{Input} & \textit{Memory} \\
\left\{\begin{array}{l}
N\left(ba \sum_{j=1}^{K} s_{ij}^2,\ b^2 \sum_{j=1}^{K} s_{ij}^2\right) \\[2ex]
N\left(0,\ b^2 \sum_{j=1}^{K} s_{ij}^2\right) \\[2ex]
N\left(ba \sum_{j=1}^{K} s_{ij},\ Kb^2\right) \\[2ex]
N(0,\ Kb^2)
\end{array}\right.
&
\begin{array}{l}
\text{if present} \\[2ex] \text{if absent} \\[2ex] \text{if present} \\[2ex] \text{if absent}
\end{array}
&
\begin{array}{l}
\text{present (6a)} \\[2ex] \text{present (6b)} \\[2ex] \text{absent (6c)} \\[2ex] \text{absent (6d)}
\end{array}
\end{array}
$$

A b is now also a part of the formulas where the memory letter contains feature i and represents the arbitrary but nonzero gain or indiscriminate amplification (i.e., the DC component).

The signal-to-noise ratio is highest, of course in (6a), but interestingly there is a nonzero "signal" output in case (6c), although not in (6b). However, whether case (6b) or (6c) produces the most correlation results greater than a decision criterion c, also depends on the variances, $b^2 \Sigma s^2$ and KB^2, or in particular, on Σs^2 versus K.

If the average power in the signal feature $(\Sigma s^2)/K = 1$, then the variances are equal, and the frequency that feature i is detected to be present is greater in (6c) than in (6b) for any value of c. Notice also that (6b) and (6d) also depend on the variance (or equivalently, the total energy in the noise). If $(\Sigma s^2)/K = 1$ then they predict equal detection ratios.

The foregoing is highly tentative. In some cases it may be that an additional noise source should be associated with the correlation processor. This would lead to a greater symmetry in (6) with respect to cases (6b) and (6c). However, also attendant would be minor mathematical complications in that the resultant noise would not be Gaussian; and in any event, with well-learned memory symbols, (6) may be reasonable.

A quite different approach would be to posit a simple neural setup with a counter neuron being fed by a central memory-feature "i" neuron (member of a set of neurons associated with a given letter) and an input neuron from the periphery. The memory neuron would fire faster when feature i was contained in the memory letter, and the input neuron would fire faster when feature i was contained in the letter presented. Impulses from the two neurons could be additive with regard to the counter neuron. In this way, additive sources of noise could come from both types of neuron and affect overall accuracy. However, this additive approach loses the well-documented structure of detection within a linear filter framework that the former description possesses. McGill (unpublished manuscript) has shown that additive analogues can be worked out. Nevertheless, there is yet to appear an adequate account of spatial frequency effects in terms of counter theory. Anderson (1973) has explored a neural theory closely aligned with linear system theory.

At any rate, it is time to think about what happens regarding the decision function. We cannot go into it in much depth here; as noted earlier a number of current and new decision models are discussed in a recent report on letter recognition (Townsend & Ashby, unpublished manuscript). Suffice it to say that we may think of the decision function as exacting a probability measure on the matching and nonmatching features occurring with each memory symbol.

The result of the feature comparisons on *each* memory symbol can be characterized by three sets of features made up of: (1) those overlapping the particular memory symbol and the input set; (2) those in the memory symbol but not in the input set; and (3) those in the input set but not in the memory symbol. Potentially, each unique set of features in (1), (2), and (3) in combination with each specific memory symbol can have different effects on the decision function. Basically, we conceive of the first part of the decision function as exacting a measure on these sets that determines their weight in the final selection of a response.

In typical instances the process is simplified by assuming that the three types of matches have uniform effects on the measure, independent of the unique extracted-set-memory-symbol combinations. For example, in several current models, a single feature of type (3) is sufficient to deliver the memory symbol to the second part of the decision stage (see Fig. 1) with an assigned measure of zero; that symbol will than have no chance of being reported unless all other symbols also have measure zero attached to them on that trial.

The second phase of the decision function process maps the memory symbols, on the basis of the assigned measures, to a guessing set. The response is then selected from this set by use of response bias probabilities.

We have thus far developed our general feature matching model for the situation when only a single symbol is presented. Before presenting our results on mathematical capacity theory, we expand our discussion to include multisymbol search experiments.

Multisymbol search paradigms typically designate a single symbol as the target and, on some proportion of the trials, embed it in a set of other symbols; on the remainder of the trials, the target is absent. The multisymbol set may appear in a display after a target has been designated [*early target* (ET) design, e.g., Atkinson, Holmgren, & Juola, 1969, Townsend & Roos, 1973], or it may be in a short-term memory and the target or probe display after the multisymbol memory set [*late target* (LT) design, Sternberg, 1966]. In either case the task of the observer is to indicate, usually with a button press, whether or not the target was present. Of course, it is a short step to situations where one of a set of possible targets is always in the multisymbol set with a forced choice among the critical set of symbols (e.g., Estes & Taylor, 1964) and to situations where various aspects of the irrelevant symbols and targets are manipulated.

We discussed single symbol pattern recognition in terms of feature extraction via a filtering or matching of the input with the feature sets of the symbols in memory, presumably a long-term memory in the case, for example, of recognizing letters from one's native alphabet. The generalization of our model to multisymbol paradigms is straightforward. For LT tasks the details are analogous to those we have developed, because we assumed that O recognizes a present symbol by comparing it to N symbols stored in memory, although of course the LT set is presumably contained in a short- rather than long-term memory. When an ET task with visual display is considered, it may be that the target symbol acts as a filter against all the incoming symbols, most likely in parallel after which a decision is made on the basis of overall feature similarity, just as in the case of single symbol recognition. Here, however, the decision is made on the basis of similarity with the target symbol. In LT tasks, if comparisons were carried out in an acoustic form system (see Townsend & Roos, 1973), the matching would be between acoustic features, but if the comparisons were made in a visual form system, the matching would again be among visual features. As in the earlier case, it may be that the decision function proceeds in a more or less automatic fashion. For instance, the measures of similarity (based on feature overlap and differences) could be accruing continuously, and when (and if) one of them mounts higher than some criterion, a "yes" response is made. Alternatively, the decision function could be calculated after all feature matching is completed and all similarity measures computed.

We turn now to the dynamic description of capacity in pattern-matching tasks.

THE ARCHITECTURE OF CAPACITY

Capacity as Power or Energy

As is well known and as we saw earlier in the present context, capacity can be expressed in terms of the output signal-to-noise ratio, which in turn reflects the quality of the filter. If the filter is exactly the signal, which is correlated against the input, then the signal-to-noise ratio is maximized and is in fact the *input* signal-to-noise ratio accumulated over the sampling interval and the bandwidth of the signal.

We return to the signal-to-noise ratio (which is, of course, just the d' of signal detectability theory) later. Now, however, we wish to employ the correlational linear system model to go beyond d'. That is, we attempt to represent capacity at the microlevel of the correlator, unperturbed by noise, by developing the constructs of power and energy.

Power and energy, as defined in the following, although apparently not heretofore exploited in cognitive contexts (as opposed to uses in ideal detector theory as, for example, in acoustic psychophysics), are, of course, tried and true constructs from electrical and communications engineering. A cognitive development of these concepts seems potentially useful. For instance, individuals may differ in the power or energy of their systems rather than in the amount of noise present. Further, in some cases, notions of power and energy may be helpful in modeling variations in capacity devoted to a detection or recognition task (as in shared-attention experiments, e.g., Kantowitz & Knight, 1976). Some altered or expanded form of certain of these ideas may also find employment in the study of capacity in simple and complex motor performance situations (e.g., see Pew, 1974).

Here, though, we are concerned with pattern matching, where we feel capacity is closely related to latency. In order to bring latency into our deliberations, we find it necessary to invent some new potential measures. The first that we suggest are primarily for model building and comparison, as are power and energy, and our measures utilize these quantities again. Then, in the next subsection, the possibility that power and energy might affect the time taken by the system to perform its function is entertained, and a way in which this might be effected stochastically is presented. Finally, some summary measures of *effective* capacity, which attempt to capture accuracy and latency factors when the capacity supply is constant and which should often be observable, are developed.

First, we want to make some cursory remarks on time- and event-limited sampling. Let us consider the situation when O is forced by the experimenter to limit the latency of responses by deadline procedures. There seems to be basically two ways in which this might be accomplished (with many possible variations on these themes). First, O might directly fix the time in which he or she processes, or he or she might limit the amount of information he or she acquires (e.g., Luce & Green, 1972). In terms of the correlation scheme previously presented, both kinds of control could possibly limit the number of the time samples (e.g., to be less than K in the formulas in the second section).

In time-varying signals, this would have the effect of disallowing a complete correlation with the signal component of the input and thus lead to increased error. On the other hand, in basically stationary visual patterns with reasonably long display intervals, as when the rise and decay times shown in Fig. 2 are brief compared to the length of the signal, longer sampling intervals essentially permit repeated observations of the same information (although with varying noise intensity, of course). Therefore, increasing the time of observation in this case would lead to increment in detectability according to well-known formulas (e.g., Green & Swets, Chap. 9, 1966). We must truncate our discussion of these points in order to address the more neglected question of how to represent capacity of the functioning system per se, but we return to the relation of accuracy to processing duration in the next subsection.

As noted earlier, a system or subsystem employing a correlation function is optimal under rather wide conditions. Furthermore, once the so-called impulse response function $h(t)$ takes on the form giving the correlation result, additional amplification, for example by means of increasing b in Eq. (6), does not lead to any improvement in signal-to-noise ratio, because the noise is amplified also. Hence, a major part of any capacity allotment to a specific subsystem, as to one of the feature detectors (whether the capacity allotment is manipulable or not), must be the setting up and maintenance of the correct form of $h(t)$. It is clear, though, that below some lower limit of amplification, the system must fail to perform its work adequately. The best way to formally represent the capacity required for maintenance of $h(t)$, or how performance should suffer when the capacity allocated to amplification falls below the acceptable limit, is not at all obvious presently.

However, we may for the moment set aside the problems of how (and how much of) the available capacity is allocated to the maintenance of $h(t)$, what the lower limits of amplification are, and the consequences of not meeting these lower limits. We may still ask if there exists any potentially useful measure of the capacity used in the correlator. The answer is in the affirmative.

One measure of possible interest is the *power* or rate of energy disbursement. The power applied at a particular time can be found from $[h(t)]^2$ (which we sometimes write $h^2(t)$ to lessen the number of parentheses) for any t from O to T. For some purposes the power allocated to a particular frequency might be of interest, in which case $|H(f)|^2$ would be appropriate, where $H(f)$ is the Fourier transform of $h(t)$ and is known as the *system function* (e.g., Padulo & Arbib, 1974, p. 426).

If the *total* capacity used up during the entire correlation period (i.e., the energy) is of interest, then we must sum or integrate the power dissipated over the observation interval or frequency bandwidth covered by the correlator. Thus the total capacity in the type of system developed in the preceding can be described in "energy" terms as

$$C(T) = \int_0^T h^2(t)\,dt = \int_{-WT}^{WT} |H(f)|^2\,df.$$

One of the nice properties of the "System function" approach is that the system function of a serial system consisting of n linear subsystems can be written as the product of the component system functions. Thus, if $H_n(f)$ denotes the overall system function, and $H(f, i)$ is the individual system function i of the ith subsystem, then

$$H_n(f) = \prod_{i=1}^{n} H(f, i).$$

The power applied to frequency $f = f'$ then, for example, would be

$$H_n^2(f') = \left[\prod_{i=1}^{n} H(f', i)\right]^2$$

and the total capacity in energy terms would be the integral of this expression over the appropriate bandwidth. Further, note that the power in terms of the time dimension can be obtained by taking the inverse Fourier transform of $H_n(f)$ and squaring. We should remark in passing that there are analogous structures for discrete-time linear systems as well.

We reemphasize here that the usefulness of say, $[h(t)]^2$, does not depend on there being a "real" power open to our inspection. Rather, $[h(t)]^2$ is a hypothetical process or function (what used to be called an *intervening variable*). This function represents the magnitude of the proposed hypothetical correlation mechanism at any point in time, in a way that allows us to employ some of the logic, intuition, and mathematical relationships of analogous concepts in physics and engineering.

Suggested Measures That Include Speed and Accuracy[3]

We as yet have no structure to reflect the type of *latency* changes that might stem from allocations of more or less capacity to the system given by $h(t)$ or $H(f)$. At any particular level of analysis, an observer might or might not be able to alter the allocation of capacity to a system (e.g., the feature-within-a-letter level may be too fine to permit such control), but even if the capacity is constant, the effect on latency must be shown for a complete system description. Although latency structure of a deterministic nature could be constructed, we may as well begin with the more interesting general problem where for a given allocation or amount of capacity, the latency varies probabilistically from trial to trial.

We want: (1) to have a measure of capacity that reflects not only, say, the power or energy applied on a particular trial (as we have previously) but also the

[3]The measures incipiently explored in this section are oriented around power, energy and signal-to-noise ratio interleaved with notions of stochastic latency theory. We refer the reader to other recent mathematical approaches to speed-accuracy relations, particularly Green and Luce (1973), Link (1975), Pike (1973), Swensson and Thomas (1974), Theios et al. (1973), and Thomas (1971). The chapters by Joseph Lappin, Stephen Link, and Robert Pachella, Keith Smith, and K. Stanovich, in the present text are all related to speed-accuracy issues.

speed with which it was applied; and (2) to have in hand a means of describing how capacity affects the latency of a system.

We must wade through some mathematical detail, even though our treatment is restricted to outline form, in order to develop a rationale for objective (1). So as not to lose sight of the forest for the trees, let us set down the suggested measure for (1) and then return to a slightly more specific justification. The measure investigated is Q/T, the total energy produced by a certain system, relative to the duration of time taken to produce it. Note that this measure increases with energy and with speed. As we shall see, in stochastic systems, it makes sense to think of Q as a random quanitity, for any given time T, and also, though less obviously perhaps, to consider T as a random variable for any given energy Q. Thus we wish to define, in general, the notion of an average or expected value of the ratio Q/T. In a particular psychological experiment, on each trial O is confronted with a task requiring some energy Q in order to be completed. The time variable T is immediately useful in expressing how long any particular system takes to produce the requisite energy Q.

An approach to this probelm that presents itself is to start with the time course $h(t)$ of the particular system under discussion. Now, in general, because we consider stochastic systems, there is a *set* or *ensemble* of such functions and a distribution describing the frequency with which one or more particular $h(t)$s actually occurs. This frequency distribution may be called $f[h(t)]$. What we are really interested in, however, is the power, and later, the energy, so we ask about $[h(t)]^2$ as a function of time, which is, as we saw earlier, the power curve of a deterministic system with impulse response function $h(t)$. The frequency distribution on $h(t)$ in a stochastic system immediately generates a new distribution on the power curves as functions of time, so that we also have the distribution $f[h^2(t)]$ (obviously this is a new f, not the old one for $h(t)$; we use f to keep the notation simple). For brevity, let $h^2(t) = P(t) =$ power, and $f[h^2(t)] = f[P(t)]$.

Let us return to consider our potential capacity measure Q/T. We would like to develop it stochastically from the preceding distribution on $P(t)$. Now, the amount of energy Q produced by our stochastic system at a particular time T is a random quanitity. Similarly, the duration T taken by the system to produce some amount of energy Q is also random or probabilistic. Hence, in order to achieve something like an average of Q/T over all possible values of Q and T, we require knowledge about the probability distribution on Q and T.

To begin, we define a necessary and important quantity

$$\bar{f}(Q, T) = \int_{[P']} f(P)\,dP,$$

where P' is the set of power curves that integrate to exactly energy Q in exactly time T. This formula clearly gives the total frequency with which the system has produced energy Q at time T. Note that we have suppressed the argument t in $P(t)$ in the main integral, because it is the probability density on the set of *functions* P, which is of interest there. We have placed a bar over f because it is not a true joint probability function, because it will not be 1 when integrated over all possible values of Q and T, each going from O to infinity.

On the other hand, by holding one variable fixed (either Q or T), we can define a conditional probability function on either Q or T that is always positive and integrates to one. For instance, when T is fixed and Q can vary, we can generate $f(Q|T)$ by sweeping out the variable Q in $\bar{f}(Q, T)$. For any value $Q = Q'$ that we choose, $f(Q', T)$ is the probability function (either a density or mass) of $f(Q \mid T)$ at the point $Q = Q'$; that is, $f(Q = Q' \mid T)$ is the "frequency" that, given processing is completed at time T, the system has produced energy Q' (or more precisely that it has produced energy between $Q' - \Delta Q$ and Q', where ΔQ is arbitrarily small).

Analogously, by holding Q fixed and sweeping out T, we can produce $f(T \mid Q)$, which is the density on the finishing times given that the system has expended exactly energy Q. This density is, of course, determined by the ensemble of power curves, because different curves produce energy Q at different times. In fact, if Q is large, there may exist curves that never produce it. It is obvious that such curves can not contribute to $f(T \mid Q)$. Hence $f(T \mid Q) = \bar{f}(Q, T)$ is defined (Q fixed and T varying) if and only if *every* power curve in the ensemble $P(t)$ of the system produces Q at *some* time point T.

Having now developed the necessary tools, let us, via some simple examples, see how a system's ensemble of power curves can affect its capacity as reflected in the statistic $E(Q/t)$, the expectation of the ratio of Q to T. First consider a system A that has but one power curve in its ensemble, $P(t) = P$. Thus in this system instantaneous power is always constant.

It may be assumed that the experimenter determines the processing load L of the system and that the total energy Q, which the system expends during processing, is some function of this load. For simplicity we assume $Q = L$ in these examples, but in reality it may be necessary to assume that L determines, say, some probability distribution on expended energy. Now processing is completed when energy, Q, is expended, and because system A has only one power curve, for any given Q, we will always be able to specify the completion time, T_A, exactly. T_A is just the time when the accumulated energy under the power curve, $P(t) = P$, is exactly Q:

$$\int_0^{T_A} P\,dt = Q,$$

or

$$PT_A = Q,$$

and finally

$$T_A = \frac{Q}{P}.$$

Thus for any Q we can quickly calculate T_A.

We can now determine $E(Q/T)$. Because $Q = L$, Q is a constant, and thus $E(Q/T) = QE(1/T \mid Q)$, and therefore

$$E(\tfrac{1}{T} \mid Q) = \int_0^\infty \tfrac{1}{T} f(T \mid Q)\, dt.$$

However, once we are given Q, we know that $T = T_A$ is certain. Therefore $f(T \mid Q)$ must have all of its mass at the point $T = T_A$; that is, $f(T \mid Q) = \delta(T - T_A)$, where $\delta(T)$ is the Dirac delta function at time T. The calculation of the preceding expectation is thus trivial:

$$E(\tfrac{1}{T} \mid Q) = \int_0^\infty \tfrac{1}{T} \delta(T - T_A)\, dt = \tfrac{1}{T_A}$$

Substituting in $T_A = Q/P$ yields $E(1/T \mid Q) = P/Q$, and therefore our measure of capacity is $QE(1/T \mid Q) = P$. The capacity of system A is, then, constant and thus independent of both expended energy and processing time.

To contrast this, consider system B, also with only one power curve, but one that is not a constant but that decays exponentially over time; $P(t) = Pe^{-t}$. For a given Q, we again know the completion time, T_B, exactly. Specifically, T_B is such that

$$\int_0^{T_B} Pe^{-t}\, dt = Q$$

and therefore

$$1 - e^{-T_B} = \frac{Q}{P}$$

or

$$e^{-T_B} = 1 - \frac{Q}{P}$$

Taking natural logarithms of both sides yields

$$T_B = \ln\left(\frac{P}{P - Q}\right).$$

Now again $f(T \mid Q)$ has all its mass at a single point in time; this time at $T = T_B$ [i.e., $\delta(T - T_B)$]. Thus $E(1/T \mid Q) = 1/T_B$, and, hence

$$QE(\tfrac{1}{T} \mid Q) = \frac{Q}{\ln\left(\frac{P}{P - Q}\right)}$$

First, note that this funciton is undefined for $Q \geqslant P$. This is not surprising when we note that if the system were to run for infinite time, it would only expend energy $Q = P$, because

$$\int_0^\infty Pe^{-t}\,dt = P.$$

Hence $f(T \mid Q)$ is not defined for $Q > P$, and, in fact, such a system could not process loads requiring energy greater than P.

Second, we note that as Q approaches zero (from the right), the measure approaches P. Alternatively, as Q gets large, the measure generally decreases toward zero. Therefore, without a doubt, capacity is greater in system A as measured by $Q \cdot E\,(1/T)$.

Notice, too, that the capacity in system B decreases because of its power curve; for larger values of Q, system B takes disproportionately more time to produce that Q than does system A. Thus system Bs capacity is less for larger Q than for smaller Q. This corresponds to the decreasing behavior of $QE\,(1/T)$ as a function of Q.

Finally consider system C, with two power curves in its ensemble. With probability 2/3, the curve $P\,(t) = Pe^{-t}$ (system Bs power curve) is selected, and with probability 1/3, the curve $P\,(t) = P$ is selected (system As curve). In this case $f\,(T \mid Q)$ is composed of two points, one of mass 2/3 at $T = T_B$ and one of mass 1/3 at $T = T_A$. Thus

$$E\left(\frac{1}{t} \mid Q\right) = \int_0^\infty \frac{1}{T}\left[\frac{1}{3}\,\delta(T-T_A) + \frac{2}{3}\,\delta(T-T_B)\right]dt = \left(\frac{1}{3}\cdot\frac{1}{T_A}\right)+\left(\frac{2}{3}\cdot\frac{1}{T_B}\right)$$

Once a power curve is selected, Q determines the completion time in exactly the same manner as it does in systems A and B. Our measure of capacity is therefore

$$QE\left(\frac{1}{T} \mid Q\right) = \left(\frac{1}{3}\cdot P\right) + \left(\frac{2}{3}\cdot\frac{Q}{\ln\,[P/(P-Q)]}\right)$$

just a weighted average of the capacities of systems A and B. More complex examples are easy to construct. The reader is invited to investigate the case where: (1) Q, the processing energy required, is distributed exponentially given the load L, that is, $g\,(Q|L) = (1/L)e^{-Q/L}$; and (2) there are two power curves in the ensemble—P, a constant that occurs with probability r, and the other being Pt which occurs with probability $(1 - r)$. From these facts a joint density on Q and T can be generated and $E\,(Q/T)$ and other statistics computed.

In the foregoing example, the quantity $E\,(Q/T)$ reduced to $Q \cdot E\,(1/T)$, because Q was viewed as being determined by the task load selected by the experimenter. Two other measures may often be preferable due to their easy computability, and although they may not always give equivalent answers, they do tend to represent the same kind of information. One of these measures simply employs $Q/E(T) = Q/\overline{T}$, that is, the fixed value of energy over the mean of the time taken to produce that energy (strictly speaking, we should have written $E(T \mid Q)$ in the denominator). The other related measure uses the median of T conditional on Q in the denominator, med $(T \mid Q) = T^*$.

As a simple example of these latter measures, suppose that $f(T \mid Q)$, determined from $\overline{f}(Q, T)$ by holding Q fixed and generating the distribution on T, turned out to be exponential: $f(T \mid Q) = a(Q)e^{-a(Q)T}$, where $a(Q)$ is the exponential rate parameter, which, in general, depends on Q but is of course fixed for a given Q. We then have immediately

$$\frac{Q}{\overline{T}} = a(Q) \cdot Q \qquad \text{and} \qquad \frac{Q}{T^*} = \frac{a(Q) \cdot Q}{\log 2}$$

Note that these two measures are linearly related to each other. Note also that, in general, we can expect $a(Q)$ to decrease as a function of Q, because we believe longer times are needed to produce larger values of Q. As in the earlier examples of Systems A, B, and C, exactly how these measures act as Q varies depends on how the energy-producing ability of the ensemble of power curves changes with Q. This ability is summarized in this example in $a(Q)$. Thus, if $a(Q) = 1/Q$, then our present capacity measures are constant, indicating that the energy producing capacity of our system keeps up with increased needs as expressed by Q.

Before passing to the next topic, the possibility should be mentioned again that a (a priori) distribution may be placed on Q. This makes sense whenever the experimenter is varying the load and the dependent variables are not segregated according to load. As mentioned earlier, in some circumstances it might instead be that there is an internal distribution on the energy required for processing a given load from trial to trial even though that load is constant. In any case, such a distribution on Q immediately induces a true joint distribution on Q and T and allows the computation of such capacity statistics as $E(Q/T)$ and $E(Q)/E(t)$.

Furthermore, in some cases, there may exist an a priori distribution $f(T)$, on T. As we shall see, such a distribution could be generated by inherent randomness in the human processing system or by trial-to-trial variations in Os criterion on time or information sampling. In fact, it turns out that postulation of an $f(T)$ is necessary to predict speed-accuracy trade-offs when a single power curve is assumed and the processing load and externally imposed accuracy-time constraints are fixed. This necessity arises, for example, in certain interpretations of the Lappin and Disch results (1972).

We finally mention, in anticipation of discussion in the following on speed-accuracy trade-off, that the foregoing examples have been constructed with the idea of processing a given load exactly once. We consider further the possibility that the load may be processed more than once, or, in general, that more information can be gained with longer processing.

Lest we forget, at the beginning of this section, we cited *two* requirements for our study of the effects of both speed and accuracy on capacity. As of now we have only considered the first of these, which was to find measures of capacity that take into account both the energy produced as well as the time occupied in producing it. The second requirement, which is in a sense the obverse of the first, is to find a means of exhibiting the influence of capacity on the latency probability distribution. It is to this problem that we now turn.

Keeping in mind that we are discussing energy and the speed with which the system produces it, rather than effective capacity, which includes an observable accuracy component (to be taken up later), it becomes clear that we can represent increased capacity by way of a general increase in the height of the power curves in the ensemble of a given system. For instance, P could be elevated in the preceding examples (Systems A, B, and C). This naturally increases the average speed with which any particular energy is output. However, it may be that in some cases, it will be possible to describe the effect on the distribution of T *directly* by some means.

There are a number of possibilities that come to mind, but our present conception, one that seems to offer some insight and intriguing results, is to enter a capacity factor into the probability distribution by way of the so-called intensity function (Gumbel, 1958), also known as the hazard function, the age-specific failure rate, and the conditional density function (e.g., McGill & Gibbon, 1965; Cox, 1962). The intensity function gives the likelihood that something being processed will be completed in the immediate future, given that it has not yet been completed up to the present time; it can be written as

$$I(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{\bar{F}(t)}$$

where $f(t)$ is the density function of the processing distribution, and $F(t)$ is the cumulative distribution function. The quantity $1 - F(t) = \bar{F}(t)$ also goes by various names, the most common of which is probably *survivor function*, due to its use in failure theory. There, $\bar{F}(t)$ is the probability that a component has not yet failed, or, synonomously, that it has "survived" to time t.

Let us suppose that the capacity factor enters into the intensity function as a multiplicative factor so that

$$I(t;c) = cI(t),$$

where $I(t)$ is the intensity function when $c = 1$. The interesting result follows immediately that the survivor function is

$$\bar{F}(t;c) = \exp[-\int_0^t I(t;c)\,dt] = \exp\left\{\ln[\bar{F}(t)]^c\right\} = [\bar{F}(t)]^c,$$

that is, the survivor function is a power function with exponent c. This means that the family of survivor functions, or, equivalently, the family of distribution functions $F(t; c) = 1 - \bar{F}(t; c)$, are ordered according to c:

$$\bar{F}(t; c_1) > \bar{F}(t; c_2) \qquad \text{if and only if } c_2 > c_1$$

or

$$F(t; c_1) < F(t; c_2) \qquad \text{if and only if } c_2 > c_1$$

Thus it can be seen that a larger c means shorter latencies. It is obvious that the foregoing relation implies a consequent ordering of the means and medians of the distributions. In the case of the median, for instance, the dependence of the median on the gain is represented by

$$T^*(c) = \bar{F}^{-1}(\sqrt[c]{.5}).$$

so the median is the inverse survivor function of the cth root of ½. Clearly, the median moves to the left (decreases) as c increases, in a way that can be trivially calculated once the distribution or survivor function is known.

It is pertinent to note here that the intensity function is stronger or more elemental than the survivor or distribution functions in the sense that an ordering of the intensity functions implies an ordering of the distribution or survivor functions, but the reverse is not true; one distribution function may always (for any t) be greater than another without the same being true for their associated intensity functions (a formalization follows in the next section).

Now we can see that with the energy Q set at some value, but possibly also dependent on c, the measure of capacity given by the ratio of Q to the median time T^* is

$$\frac{Q}{T^*} = \frac{Q}{F^{-1}(\sqrt[c]{.5})},$$

The intensity function of exponential distributions is just the rate of the distribution, so that in our especially simple example of a system producing $f(T \mid Q)$ as an exponential function of T, seen previously, the new distribution, including the influence of capacity via a multiple of the intensity function, becomes

$$f(T \mid Q) = ca(Q)e^{-ca(Q)T}$$

Following from this, the capacity measure taken as the ratio of Q to the median of T is given as

$$\frac{Q}{T^*} = \frac{c \cdot a(Q) \cdot Q}{\log(2)}$$

This quantity increases linearly with c as it stands, indicating that capacity is improving, just as it increases with Q if $a(Q) = a$ and c are constant. The former reveals capacity improvement because of faster speed, the latter because of increased energy production for given average speed.

Of course, although the effect of capacity on latency through the intensity function can be quite helpful, the circumstances under which manipulations of the distribution on the power curve ensemble (or the ensemble itself) produce changes in the intensity function, particularly multiplicatively, are unknown. It is likely that a fairly narrow class of transformations on the ensemble lead to intensity function orderings. To be sure, it is not even clear what class of ensembles and associated distributions leads to particular $f(T \mid Q)$s, for example, the exponential.

We have thus far considered capacity as a function of expended energy and of processing time. We now turn our attention to possible measures that might reflect what we call *effective capacity*. These are typically based on observable reaction times and an accuracy statistic. It may be anticipated that such measures can also be closely related to trade-offs of speed and accuracy.

The approach taken here is a special case of the foregoing general development, with additional assumptions linking capacity as energy to observable accuracy statistics. Specifically, suppose that the power curve ensemble consists of a single function that is constant at power P over all time. Strictly speaking, $h^2(t)$ is not exactly constant, of course, but, in general, we can write

$$\frac{1}{T}\int_0^T s^2(t)\,dt = \bar{P}$$

and hence the total signal energy equals $\bar{P}T$, that is, the (average power) · (time). Actually, in the ideal visual situation, where in each time interval ΔT, a spatial correlation is performed on the constant pattern (with independently varying Gaussian noise), the variations in h are *spatial*, but h is constant in time for each ΔT. This ideal situation produces $h^2(t) = P(t) = P$; that is, the power allotted to the cross-correlation between the input and the memory item is constant for each correlation.

Under these assumptions we can write an expression for the expended energy in terms of time T:

$$Q_h(T) = P \cdot (T - t_0)$$

where $Q_h(T)$ represents the energy produced by the filter $h(t)$ up to time T, P is the constant power, and t_0 is the dead time before the system can produce any effective energy. It may be reasonable to identify the effect of this power with the well-known quantity $(d')^2$, which in matched filtering, such as we are considering, is defined as the total signal energy relative to the noise power density. This is due to the well-known aforementioned relation (Green & Swets, Chap. 9, 1966) between repeated observations of a constant stimulus in Gaussian noise and increases in detectability. More precisely, it is easy to see that

$$[d'(T)]^2 = \frac{\int_0^T s^2(t)\,dt}{\frac{N_0}{2}} = \frac{Q_h}{\frac{N_0}{2}}$$

where $N_0/2$ is the noise power per unit frequency, which with white Gaussian noise is constant. Substituting in $Q_h = P \cdot (T - t_0)$, we arrive at

$$[d'(T)]^2 = \frac{2}{N_0} P \cdot (T - t_0) = (A \cdot P \cdot T) - (A \cdot P \cdot t_0)$$

Where A is the reciprocal of the constant noise power. Hence $[d'(T)]^2$ is a linear function of time with a slope that is proportional to the power. Note that all three terms $(A, P,$ and $t_0)$ that determine the intercept must always be nonnegative. Hence we have an additional test of the foregoing assumptions; not only must $(d')^2$ be a linear function of T, but also the $(d')^2$ intercept must be negative.

Note that we actually plot $(d')^2$ against RT rather than T and that in general RT is unequal to T. Specifically, we assume RT $= T + t_1$, where t_1 is the time contributed by all other stages in the entire processing system (e.g., response selection and execution). This inconvenience causes us no great problems, because it means only that $[d'(t)]^2$ lifts off the time axis when RT $= t_1 + t_0$ rather than when $T = t_0$. Thus we can consider the $(d')^2$ plot against RT as a simple shift of the "true" plot against T.

One interpretation of the speed-accuracy curve $[d'(T)]^2$ as a function of T (or RT) is that a suitable transformation of the curve should yield capacity viewed as power or something proportional to power. In the present instance, the appropriate transformation is of course

$$\frac{[d'(T)]^2 + APt_0}{\mathrm{RT}} = AP$$

where APt_0 is the intercept and A is a constant of proportionality from the foregoing formula of $[d'(T)]^2$.

The simple form of this function suggests a transformation to attempt to get at the underlying capacity expressed as power, or at least something that is proportional to this power. This "power" measure can be considered as a measure of effective capacity per unit time. In the present case, it is the slope of the $[d'(T)]^2$ curve.

These concepts have some general validity: If the power of a system, expressed in terms suitable to that system, is constant over time, and if accuracy increases as a function of increased time, then it is always possible in principle to *transform* the resulting speed-accuracy curve to a measure reflecting the constant underlying "power." However, it is critical to note that if the cognitive or perceptual power $P(t)$ is *not* constant over time, then such an attempt to find a constant measure of capacity is fruitless. Indeed, a speed accuracy curve then represents the variation in $P(t)$ as well as the more often assumed variation in an internal criterion on time or information. It still may be possible, with a suitable model, to uncover $P(t)$ as a function of time.

Despite the foregoing caveats, it might nevertheless be of interest to look for accuracy measures that are linear functions of time. Some other investigators have proceeded in this or an analogous fashion. That is, under the assumption of a constant system, discriminability per unit time (this implies the assumption of but one power curve in the ensemble for a given experimental condition), a measure proportional to this discriminability should be a linear function of reaction time. The results of these studies, which have plotted accuracy measures, such as $(d')^2$, as functions of reaction-time data, are overall rather encouraging.

Taylor, Lindsay, and Forbes (1967) have studied, in a series of experiments, the relationship between $(d')^2$ and reaction time. These studies involved detection tasks embedded in a 2 X 2 design. In the most conclusive of these, they transformed data of Schouten and Bekker (1967) to $(d')^2$ versus mean reaction time. The results strongly supported linearity.

Lappin and Disch (1972), instructing their subjects to respond to two easily discriminable stimuli at a speed that maintained a 25% error rate, studied the relationship of median RT to five different accuracy measures, including d' and $(d')^2$, each coming from a distinct model of the system. All of these measures provided reasonably good linear fits to the data, and in every instance the Q_h intercept was indeed negative. On the other hand, it may strike one as curious that none of the measures were strongly falsified, because they arose from separate theoretical hypotheses. Several potential explanations for this state of affairs come to mind.

First, it is possible that there was too much variability or noise in the data to separate the predictions of the models. Second, it could be that these measures are really similar functions of the data. For instance, one of the measures was $(-\ln \eta)$, where η is the stimulus similarity parameter of Luce's choice theory. Luce (1963) has shown this parameter is quite similar to d' of signal detectability theory. Also tested by Lappin and Disch (1972) was transmitted information that Taylor, Lindsay, and Forbes (1967) have pointed out is related to $(d')^2$.

The third possibility is that all of the functions are approximately linear only over the restricted range of the experiment. The set of RT data collected by the authors ranged from about 100 msec to about 300 msec, and thus the fitted linear curves were quite steep. Possibly, if widely varying experimenter-imposed deadlines had been employed, there would have been greater predictor variation.

The Lappin and Disch study can be considered a micro-trade-off experiment, where the speed variation arises out of internal processing variations rather than direct experimenter manipulation. As anticipated earlier, it is exactly in this type of situation that it is necessary to assume the existence of an a priori distribution $f(T)$, because a power ensemble of one curve can produce no variation in reaction time related to accuracy. Note that $f(T)$ does not enter into the $[d'(T)]^2$ plots directly; it only provides for variation in T, which is accompanied by more or less energy acquired from the (single) power curve. The distribution $f(T)$ could be generated by a varying criterion on the part of O on how much time or information he or she employs each trial. Another possibility is that $f(T)$ might be generated by assuming that the time the stimulus is available is distributed as $f(T)$, and that for any given time T, the information is of a constant integrity.

Green and Luce (1973), by varying response deadlines in a yes-no detection task, generated d' versus $\bar{T}$ curves, which appear quite linear over a larger RT range (ca. 200 to 1200 msec). Again, in all cases, the d' intercepts were negative. In one condition deadlines were placed on both signal and noise trials; in another condition a deadline was placed only on signal-present trials. On noise trials Os were not penalized for long response times. Results indicated that the d' versus $\bar{T}$ slopes were consistently much steeper in the signal-only deadline condition. This is perhaps interpretable as an increase in P due to release of capacity from elsewhere in the total system.

If, according to our earlier development, we can write an approximate relation of the form [rather than the exact relation in terms of $(d')^2$] as,

$$d'(T) \doteq (A \cdot P \cdot T) - (A \cdot P \cdot t_0),$$

we see at once that Ps increase should not alter the $d'(T) = 0$ intercept. In fact, the points where the d' curves cross the time axis appear little changed across the two conditions. In addition, increasing P should push the $d'(T)$ intercept (the $T = 0$ intercept) farther away from zero, because the term APt_0 increases. This prediction also was clearly supported by the Green and Luce data.

As the duration and/or intensity of the display is turned up and the signal-to-noise ratio increases to where performance approaches perfection, a measure based on speed and accuracy may become less adequate; for example, errors may be largely unrelated to the perceptual and decision processes under study. It may then be preferable to employ one or more quantities to measure capacity related to latency only rather than to both latency and accuracy.

CAPACITY MEASURES AND DYNAMICS BASED ON SPEED ONLY

General Introduction

In this section we consider approaches to pattern-matching tasks that emphasize speed alone and are broad in the sense that they are based on nonparametric properties of the processing distributions. One can utilize these characteristics or derivative properties in tests of fundamental principles of processing and in delimiting the range of potential explanatory models. One of the first attempts at such a development was by Sternberg (1973), who proposed a test of self-terminating models based on some properties of the RT distribution. Sternberg proceeded by way of assumptions about how search or comparison varies with the location of an item in a hypothetical internal storage array and by then deriving predictions about how the RT distributions should be affected by changes in this array size.

Although Sternberg apparently derived these properties with the idea of seriality in mind, his assumptions are not necessarily confined to this class of models. However, predictions emanating from those assumptions are contrary to many plausible parallel models. Specifically, Sternberg assumes that the processing system consists of a series of stages, each acting as though it has its own source of capacity that is unaffected by how many other stages are in the chain. This is essentially an assumption of unlimited capacity within each stage (see Townsend, 1974a, 1974b), and his method is therefore applicable to only a limited subset of the class of potential models.

Our own view is that a prediction about the ordering of distributions is first and foremost a prediction that arises from the architecture of capacity and only secondarily concerns issues such as self-terminating versus exhaustive or parallel versus serial. Thus we present what we hope is a more general theoretical approach. It is based on the examination of capacity effects at different levels of processing: in terms of hazard functions, distribution functions, and a measure of central tendency.

We take special care to make clear that although we discuss these approaches in terms of the multisymbol matching task, the notions are much more general; in fact, they can be applied to any situation in which a system is processing some finite number of things. That is, we are concerned with the effect of increasing the number of elements on reaction time, and although we speak here of these elements as symbols (e.g., letters), they are in no way restricted to symbols. Rather, they could be features within letters, as in the work of Taylor (1976), or each element could conceivably be a group of letters (e.g., words).

Throughout this discussion it is assumed that error rates are low and that those errors that do occur are independent of the comparison process and hence can be ignored. There are several germane points in this regard. The first is that in some cases conclusions are unchanged or strengthened when errors change. For instance, if error rate and latency both increase when the number of symbols increases, then obviously evidence of limitations in capacity is obtained. Only when latency and error information go in opposite directions will ordinal conclusions about capacity be ambiguous.

The second point is that it is still not clear that the errors acquired in a typical memory or brief visual display search experiment with very high accuracy are intimately connected to the comparison process itself. Such intuitions are supported by findings such as those of Swanson and Briggs (1969) and Pachella

(1974) that an 8% or 9% error increase in an LT task revealed only a change in intercept (i.e., slope remained constant). Changes in intercept but not slope have been said to imply corresponding changes in base time (stimulus encoding, response selection, and response execution) but not comparison time (e.g., Swanson & Briggs, 1969; Sternberg, 1967). Thus it may be that error rates up to about 10% (with error rates greater than 10%, Pachella, 1974, began finding slope changes) do not greatly influence the comparison process and therefore may be safely ignored when modeling this stage.

Third, the type of structure developed here (in addition to that introduced in the previous sections) can itself be of aid in predicting accuracy changes in certain cases. For example, the derivations in the following yield probability distributions for completion times, so that if the time allowed for processing is abbreviated, either by experimenter control (as with shortened display times in the case of visual search) or by observer-determined deadlines, the latency and accuracy effects can be predicted.

We make these remarks not because we are unconvinced of the importance of empirical and theoretical work in the relations of speed to accuracy, quite the contrary, as the earlier sections should demonstrate. It is only by continued work in these regions that a fuller understanding of the precise relationships will be achieved.

Development of the Stochastic Inequalities

A peculiarly intriguing approach to the study of stochastic systems is the analysis (or development) of the time intervals between successive completions (the intercompletion times) in terms of capacity written via the intensity function on the intercompletion time. The overall intensity function for a given stage (and therefore a given intercompletion time) is just that of a given symbol in serial processing but in parallel processing and some hybrid models will be itself a combination of intensity functions of individual symbols undergoing processing. Of course, qualitative or quantitative differences in capacity can be embedded in the orderings of the respective intensity functions.

It is useful to compare the use of the intensity function to represent capacity differences with two other candidates. One places orderings directly on the distribution functions and another on an inequality of random times. It is easy to show that if, say, the intensity function for symbol A is always greater than that for symbol B, then the distribution function for symbol A (or A and B may be intercompletion times or subsystems, etc.) is always greater than that for symbol B. Suppose that

$$I_A(t) = \frac{f_A(t)}{\bar{F}_A(t)} > \frac{f_B(t)}{\bar{F}_B(t)} = I_B(t)$$

then

$$\int_0^t \frac{f_A(t)}{\overline{F}_A(t)}\,dt = -\int_0^t \left[\frac{d\log \overline{F}_A(t)}{dt}\right] dt > -\int_0^t \left[\frac{d\log \overline{F}_B(t)}{dt}\right] dt$$

$$= \int_0^t \frac{f_B(t)}{\overline{F}_B(t)}\,dt \qquad ,$$

which by integration leads to

$$-\log \overline{F}_A(t) > -\log \overline{F}_B(t),$$

and by exponentiation implies

$$\overline{F}_A(t) < \overline{F}_B(t) \qquad \text{i.e.,} \quad F_A(t) > F_B(t),$$

which was the claim. This is natural, because the intensity inequality states that at any time t when A is not yet completed, the likelihood that it is completed in the next instant is greater than the comparable likelihood for B. The inequality in the distribution functions makes the weaker statement that for any t the probability that A was completed at some time less than or equal to t is greater than the comparable probability for B. Put succinctly, an ordering of intensity functions orders the distribution functions, but not conversely; if the distribution functions are ordered, the intensity functions may or may not be.

As an example of this latter proposition, consider the two density functions illustrated in Fig. 3, where the random variable A is distributed Rayleigh (i.e., $f_A(t) = te^{-t^2/2}$ for $t > 0$ and zero otherwise), and B is distributed exponential with rate $\lambda = 60$ but displaced $t = 40$ units to the right (i.e., $f_B(t) = 60e^{-60(t-40)}$, $t > 40$; zero otherwise). We might think of this 40-unit time lag as a dead time or an irreducible minimum response time. Now our first requirement is that the distribution functions are ordered; thus we ask,

$$F_A(t) \overset{?}{>} F_B(t) \qquad \text{for all } t > 0,$$

which is equivalent to asking

$$1 - e^{-t^2/2} \overset{?}{>} 1 - e^{-60(t-40)}$$

or

$$e^{-t^2/2} \overset{?}{<} e^{-60(t-40)}$$

Taking natural logarithms of both sides yields

$$\frac{-t^2}{2} \overset{?}{<} -60(t-40)$$

and finally

$$t^2 - 120t + 4800 \overset{?}{>} 0 \qquad \text{for all } t > 40.$$

This expression obviously describes a parabola opening upward. Setting the derivative equal to zero and solving for t implies that its minimum is at $t = 60$. But we see that the value of the parabola as it crosses $t = 60$ is

$$(60)^2 - (120)(60) + 4800 = 1200 > 0.$$

Because the minimum is greater than zero, we know all other points must be, and therefore, as desired,

$$F_A(t) > F_B(t) \qquad \text{for all } t > 0.$$

Now the intensity function of A is given by

$$I_A(t) = \frac{te^{-t^2/2}}{e^{-t^2/2}} = t \qquad \text{for all } t > 0$$

and

$$I_B(t) = \frac{60e^{-60(t-40)}}{e^{-60(t-40)}} = 60 \qquad \text{for all } t > 40.$$

Thus, when $40 < t < 60$, we see that $I_B(t) > I_A(t)$, and thus the counterexample

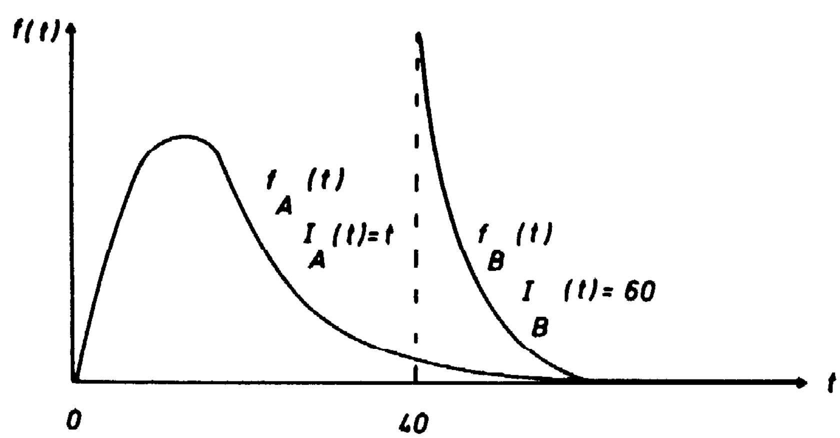

FIG. 3 Counterexample to the proposition that $F_A(t) > F_B(t) \Rightarrow I_A(t) > I_B(t)$, for all $t > 0$.

is complete. An ordering on the distributions does not imply an ordering on the intensity functions.

Yet a third type of ordering can easily be stated in terms of the underlying random variables. Let t_B and t_A be the respective random variables for As and Bs completion. Then this type of ordering can be put as

$$P(t_A < t_B) > \frac{1}{2}$$

That is, the probability that the time to complete A is less than that to finish B is greater than ½. This type of ordering is the basis of several nonparametric tests, which are probably often employed with little thought to exactly what a significant statistical result really has to say about the underlying populations.

As it happens, this inequality is implied by an inequality in the distribution functions and therefore by an inequality in the intensity functions. The proof is quick. If $F_A(t) = F_B(t)$ for all $t \geqslant 0$, then we see that

$$P(t_A < t_B) = \int_0^\infty f_B(t)F_A(t)\,dt = \int_0^\infty f_B(t)F_B(t)\,dt = \frac{1}{2}.$$

When $F_A(t) > F_B(t)$ for all $t > 0$ as assumed, we can compare

$$P(t_A < t_B) = \int_0^\infty f_B(t)F_A(t) \qquad \text{versus} \qquad \int_0^\infty f_B(t)F_B(t)\,dt = \frac{1}{2}.$$

This is easily transformed to

$$\int_0^\infty f_B(t)F_A(t)\,dt - \int_0^\infty f_B(t)F_B(t)\,dt \qquad \text{versus} \qquad 0,$$

or

$$\int_0^\infty f_B(t)[F_A(t) - F_B(t)]\,dt > 0,$$

by virtue of our assumption $F_A(t) > F_B(t)$ for all positive t, and we are done, because if the last expression is greater than 0, $P(t_A < t_B) > ½$.

On the other hand, such an ordering on the random variables does not imply an ordering on the distributions. An appropriate counterexample is found when both distributions are normal, with the mean of the A-distribution being less than that of the B-distribution. Because $P(t_A < t_B) > ½$ is equivalent to $P(t_A - t_B < 0) > ½$, we see from symmetry, and the fact that the mean of the difference variable $(t_A - t_B)$ is less than 0 and its mean equals its median, that the inequality $P(t_A < t_B) > ½$ holds. However, we can set the variance of the A-distribution sufficiently large (relative to the B-variance) that an obvious crossover in the distribution functions occurs, therefore violating an inequality in the latter.

This random-variable (or probability) ordering is moreover even weaker than the foregoing indicates, because it does not imply an ordering on either the means or the medians. For instance, assume that t_A is distributed Cauchy for $t > 0$ (i.e., $f_A(t) = 2/[\pi(1+t^2)]$ for $t > 0$ and 0 elsewhere) and that $f_B(t)$ is any distribution with a finite mean and displaced far enough to the right so that $P(t_A < t_B) > ½$. Now the expected value of A is infinite, and therefore

$$P(t_A < t_B) > \frac{1}{2} \not\Longrightarrow E_A(t) < E_B(t).$$

As for the medians, consider the discrete case in Fig. 4. Here

$$P(t_A < t_B) = (.49) \cdot (1.0) + (.02) \cdot (.49) + (.49) \cdot (.49) = .73$$

and therefore $P(t_A < t_B) > ½$. Now it is obvious that $\text{Med}(t_A) > \text{Med}(t_B)$, and thus

$$P(t_A < t_B) > \frac{1}{2} \not\Longrightarrow \text{Med}(t_A) < \text{Med}(t_B).$$

We have thus established the following dominance relationship in the three types of "stochastic" inequalities

$$[I_A(t) > I_B(t)] \Longrightarrow [F_A(t) > F_B(t)] \Longrightarrow [P(t_A < t_B) > \frac{1}{2}]$$

for all t in the domain of the probability functions. We refer to these, from strongest to the weakest, as a "intensity," "distribution," and "probability" inequality

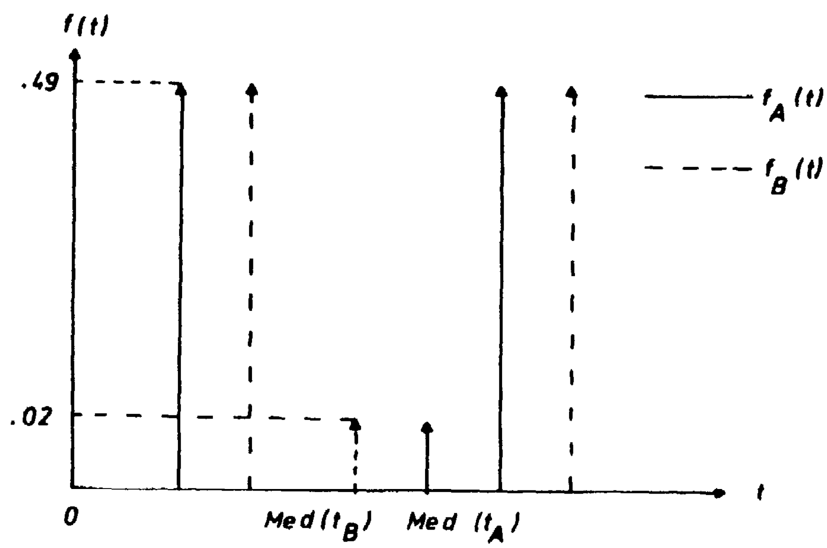

FIG. 4 Counterexample to the proposition that $P(t_A < t_B) > ½ \Rightarrow \text{Med}(t_A) < \text{Med}(t_B)$, for all $t > 0$.

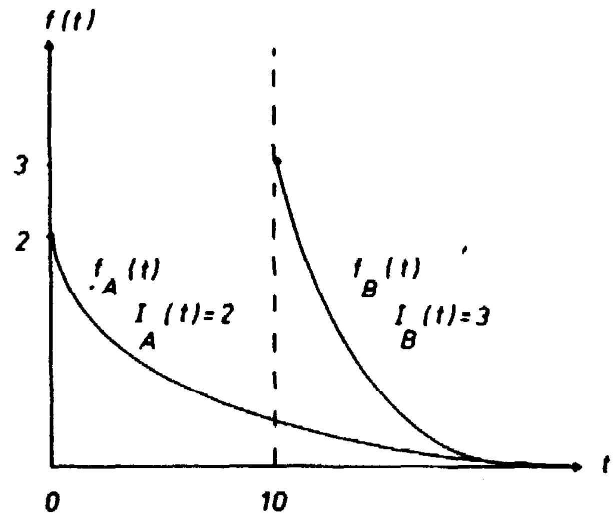

FIG. 5 Counterexample to the proposition that $P(t_A < t_B) > ½ \Rightarrow I_A(t) > I_B(t)$, for all $t > 0$.

or ordering. The rubric of the last leaves something to be desired in terms of specificity but perhaps will suffice for now.

Further, via counterexamples, we have seen that

$$[P(t_A < t_B) > \frac{1}{2}] \not\Longrightarrow [F_A(t) > F_B(t)] \not\Longrightarrow [I_A(t) > I_B(t)]$$

In this reverse case, the transitivity [that $P(t_A < t_B) \geq ½ \not\Rightarrow I_A(t) > I_B(t)$] must be demonstrated. As one would expect, the counterexample is easily constructed.

Assume that t_A and t_B are both distributed exponentially with rates two and three, respectively, but that $f_B(t)$ is displaced to the right $t = 10$ units as in Fig. 5. Now $P(t_A < 10) \cong 1$, and therefore it is obvious that $P(t_A < t_B) > ½$. But $I_A(t) = 2$, and $I_B(t) = 3$, and hence $I_A(t) < I_B(t)$, which completes the counterexample.

It is to be emphasized that the foregoing dominance relationship does not rule out that, for instance, $I_A(t) > I_B(t)$ and $F_A(t) > F_B(t)$ both hold simultaneously. Instead, this may frequently occur; but if one knows only that $F_A(t) > F_B(t)$, one must check $I_A(t) \overset{?}{>} I_B(t)$ to be sure that the ordering holds there also.

CONSTRUCTION AND ANALYSIS OF MODELS VIA INTENSITY FUNCTION DYNAMICS

In the rest of the chapter, we utilize some of the theoretical developments of the preceding section to study microscopic capacity influences on the stage of some very general classes of models. The focus is on intensity functions as measures of capacity, although one can, of course, start instead at the weaker levels associated with the distribution or probability inequalities. We begin by positing a common variety parallel model and then generalizing our results to a much broader class of systems.

Suppose as noted that processing is parallel and also independent on total completion times of the individual symbols (see Townsend, 1974a). This means that the processing times on the various symbols are independent events, which of course has different ramifications in parallel as opposed to serial systems. As we shall see, the investigation of the capacity question, which arises as the number of symbols is varied, places no constraints on the rates at the various individual symbol positions. In general, they may all be distinct.

Consider, then, the intensity functions that describe the capacity of the system during the *first* stage of processing. For various possible numbers of symbols, we have

$$
\begin{aligned}
I_1(t) &= \frac{f_{11}(t)}{\bar{F}_{11}(t)} \\
I_2(t) &= \frac{f_{21}(t)\bar{F}_{22}(t) + f_{22}(t)\bar{F}_{21}(t)}{\bar{F}_{21}(t)\bar{F}_{22}(t)} = \frac{f_{21}(t)}{\bar{F}_{21}(t)} + \frac{f_{22}(t)}{\bar{F}_{22}(t)} \\
&= I_{21}(t) + I_{22}(t), \\
&\;\;\vdots \\
I_n(t) &= \frac{\sum_{i=1}^{n} f_{ni}(t) \prod_{j \neq i}^{n} \bar{F}_{nj}(t)}{\prod_{i=1}^{n} \bar{F}_{ni}(t)} = \sum_{i=1}^{n} \left[\frac{f_{ni}(t)}{\bar{F}_{ni}(t)}\right] = \sum_{i=1}^{n} I_{ni}(t),
\end{aligned}
\tag{7}
$$

where the first subscript denotes the value of n (display size) and the second the serial (spatial or temporal) position of a symbol. These intensity functions reveal the likelihood that a symbol is completed at the instant after time t, given that it is not yet finished (at time t). There are several things to notice about (7). One is that the preceding expression for a fixed value of n determines the capacity of the individual symbols (during *all* stages of processing), because independence necessitates that the f_{ni} do not change during a single trial, and also that the overall capacity of the system is therefore given. Another aspect of (7) is that this overall capacity, as designated during Stage 1, before any symbols have been completed, can be written as the sum of the individual intensity functions and therefore the sum of the individual capacities. This seems to be an exceedingly felicitous and natural property of representation of capacity by way of the intensity functions. However, it is important to take note that when the "excess" survivor functions, the $\bar{F}_{nj}(t)$, $(j \neq 1)$, are cancelled out of the numerator and denominator, the resultant expression is no longer a true intensity function, because it integrates to n instead of to 1.[4] This should cause no difficulties as long as properties of intensity functions are not inappropriately applied to the simplified versions of (7). Comparisons of overall capacities during Stage 1 can be made between the models of different systems by employment of the sums of the individual intensities.

We emphasize the capacity for Stage 1, even though all capacity consequences are fixed by stating (7), because the orderings for potentially observable reaction times are in general implied only for the Stage 1 intercompletion time, which is, of course, equivalent to the minimum finishing time for the n symbols (processing time of first element finished). Following the development from an earlier paper (Townsend, 1974b), and assuming an inequality in the overall intensity of capacity as suggested in (7), for two systems working on n_1 and n_2 symbols, respectively, we have, as earlier

$$I_n^{(1)}(t) = \sum_{i=1}^{n_1} \frac{f_{n_1 i}^{(1)}(t)}{\bar{F}_{n_1 i}^{1}(t)} > \sum_{i=1}^{n_2} \frac{f_{n_2 i}^{(2)}(t)}{\bar{F}_{n_2 i}^{2}(t)} = I_n^{(2)}(t)$$

[4]To be precise we should write

$$I_n(t') = f(t' \mid t' \geq t) = \frac{\sum_{i=1}^{n} f_{ni}(t') \prod_{j \neq i}^{n} \bar{F}_{nj}(t')}{\prod_{j=1}^{n} \bar{F}_{nj}(t)}$$

Now we can see that when $t' = t$, cancellation is perfectly legitimate, *at that time point* t. And, $I_{in}(t)$ *does* equal the intensity function of the ith position being processed alone at time t. However, for actual description of $I_n(t)$ as t' varies over values greater than t, both $I_n(t)$ and $I_{ni}(t)$ must contain the entire probability measure, i.e.,

$$I_{ni}(t') = \frac{f_{ni}(t') \prod_{j \neq 1}^{n} \bar{F}_{nj}(t')}{\prod_{j=1}^{n} \bar{F}_{nj}(t)} = I_{ni}(t' \mid t).$$

In any case, the instantaneous rate description, which is given when $t' = t$, is properly written as the sum with cancellations included and so validly represents a summing of capacity at $t' = t$. Further, the minimum completion time distribution predictions (or any others for that matter) do not depend on this mode of expressing capacity.

By integration we achieve

$$\int_0^t I_n^{(1)}(t)\,dt = -\sum_{i=1}^{n_1} \log \overline{F}_{n_1 i}^{(1)}(t) > -\sum_{i=1}^{n_2} \log \overline{F}_{n_2 i}^{(2)}(t) = \int_0^t I_n^{(2)}(t)\,dt$$

which leads to

$$\overline{F}_{n_1}^{(1)}(\min t) = \prod_{i=1}^{n_1} \overline{F}_{n_1 i}^{(1)}(t) < \prod_{i=1}^{n_2} \overline{F}_{n_2 i}^{(2)}(t) = \overline{F}_{n_2}^{(2)}(\min t)$$

or

$$F_{n_1}^{(1)}(\min t) > F_{n_2}^{(2)}(\min t).$$

As indicated by the notation, these products are just the survivor functions for the respective minimum completion (or Stage 1 intercompletion) times, and so it is ascertained that for any value of t, it is always more likely that System 1 has completed its first symbol than that System 2 has.

On the other hand, this measure of overall capacity is not very general, because it holds no strict implications for other processing times. For example, an inequality may hold in the foregoing measure, yet the distribution functions for the *exhaustive* processing times need not be ordered such that the System 1 times are (in terms of distribution) shorter than those for System 2. To end up with distribution or probability inequalities to any particular level of processing (e.g., minimum processing time, as in the foregoing, self-terminating with a single target, and exhaustive are the most commonly used), one must write the intensity function for that level. Thus, for *exhaustive* processing in the present independent parallel model, the intensity function is just

$$I_n(\max t) = \frac{\sum_{i=1}^{n} f_{ni}(t) \prod_{\substack{j=1 \\ \neq i}}^{n} F_{nj}(t)}{1 - \prod_{i=1}^{n} F_{ni}(t)} .$$

On the other hand, the *self-terminating* intensity when the target is specified to be in position j, is simply $[f_{nj}(t)/\overline{F}_{nj}(t)]$. Interestingly, the average of the self-terminating intensities is *not* the intensity function of the self-terminating processing time. Rather, the latter, which is the quantity we need, is

$$I_n(S-T, t) = \frac{\frac{1}{n}\sum_{i=1}^{n} f_{ni}(t)}{1 - \frac{1}{n}\sum_{i=1}^{n} F_{ni}(t)}$$

which brings with it the necessary implications for the self-terminating distribution function, which is, of course,

$$F_n(S - T, t) = \frac{1}{n} \sum_{i=1}^{n} F_{ni}(t).$$

The intensity function can be written for any given level of processing in which one has interest. But the mathematics developed in the preceding for the minimum completion time turn out to be more general in the sense that at any stage of processing, we can conditionalize on the past events (e.g., items completed) and times (e.g., times of completion of the items), and then the basic structure of the minimum completion time lifts over into the conditional *intercompletion time.* Thus one can investigate the capacity at any stage of processing, given any particular previous set of completions and times (on that trial), and *compare* two or more models at that processing stage. Alternatively, one can *build* any particular capacity architecture, in the fullest possible sense, by considering each possible stage (or class of stages) and imposing his or her capacity specifications there.

As a simple example of this *comparison* property, consider the standard serial model (with a single order of processing) based on a gamma distribution with k stages at each symbol versus a standard independent parallel model whose distributions are gamma with k stages each, and assume that all rate parameters are equal. Then, a comparison of the intensity functions shows that the one for the serial model's successive stages remains constant, whereas the one for the parallel model decreases. A manifestation of this is the lengthening intercompletion times (probabilistically) in the parallel case as opposed to the constancy of the intercompletion times of the serial model.

As an example of the *building* concept, we may note that many, perhaps most, of the models we might think up on the spur of the moment to attempt to represent psychological phenomena probably would possess constant or decreasing intensity functions, for reasons associated with the two models immediately preceding. Suppose, though, that because of warm-up effects in serial processing, for instance, that it was expected that intercompletion times (which are equivalent to processing times in serial models) should actually decrease in a certain fashion. This property could be incorporated directly into the model by specification of the pertinent intensity functions. Or assume a series of problems, each successive member of which is composed of an increasingly smaller set of subproblems. Again, this structure could be implemented by way of the intensity functions–as a reduction in the number of stages in successive gamma processes, for instance. Of course, such a description is by no means restricted to gamma processes–stage mechanisms (or successive intercompletion time distributions) can be associated with any set of distributions whatsoever.

Let us briefly take up a special case of a conditional intensity function in intercompletion times. Consider a parallel model that assumes within-stage independence (e.g., Townsend, 1976). Suppose that $n = 2$ and that we wish to peruse the intensity for the intercompletion time for item a conditioned on b having been completed at time $t'_{b1} = t_{b1}$. Let the state of as processing at time t_{b1} be designated by $S_a = j$, where the state of being completed for a is $k > j$ (e.g., k may be the number of features making up a and j the number completed by time t_b). Let τ_{a2} (τ_{b1}) be the second (first) intercompletion time when b is completed first (note that t_{b1} is the same as τ_{b1}; that is, the first intercompletion time is equal to the first processing time). For the moment, let f_{ij}, $\bar{F}_{ij}$ be the density and survivor functions for the *intercompletion times* (of item $i = a, b$, during stage j), and let g_i, $\bar{G}_i$ be the density and survivor functions for the *processing times* per se (for item i). The intensity function we wish to observe is

$$I(\tau_{a2} \mid S_a = j \quad \text{at time } t'_{b1} = t_{b1}) = \frac{f(\tau_{a2} \,\&\, S_a = j \,\&\, t'_{b1} = t_{b1})}{P(\tau'_a \geqslant \tau_a \,\&\, S = j \,\&\, t'_{b1} = t_{b1})}$$

$$= \frac{f_{a2}(\tau_{a2} \mid S_a = j \quad \text{at } t_{b1})P_{a1}(S_a = j \quad \text{at } t_{b1})g_{b1}(t_{b1})}{P_{a1}(S_a = j \quad \text{at } t_{b1})\bar{F}_{a2}(\tau_{a2} \mid S_a = j \quad \text{at } t_{b1})g_{b1}(t_{b1})}$$

$$= \frac{f_{a2}(\tau_{a2} \mid S_a = j \quad \text{at } t_{b1})}{\bar{F}_{a2}(\tau_{a2} \mid S_a = j \quad \text{at } t_{b1})} .$$

The first fs (without subscripts) are, of course, meant to represent the appropriate general probability functions.

The very last formula gives the intensity function in terms of the Stage 2 conditional intercompletion time density (in the numerator) and the Stage 2 conditional survivor function for as intercompletion time. Note that we have a general case here where the second stage intercompletion time distribution can depend not only on the state the uncompleted item was in at time t_b but also on the exact time that item b finished. Clearly, by manipulation of this expression for the different stages and times and order of completion, any conceivable stochastic capacity structure can be embedded in the model.

If grosser intensities are desired, these can be obtained by averaging. For example, suppose that one would like to have the above intensity conditioned on t_b but *not* on the state j. Then we could derive

$$I_{a2}(\tau_{a2} \mid t'_{b1} = t_{b1}) = \frac{f(\tau_{a2} \,\&\, t'_{b1} = t_{b1})}{P(\tau'_{a2} \geqslant \tau_a \,\&\, t'_b = t_b)}$$

$$= \frac{\sum_{j=0}^{k-1} P_{a1}(S_a = j \quad \text{at } t_{b1})f_{a2}(\tau_{a2} \mid S_a = j \quad \text{at } t_{b1})}{\sum_{j=0}^{k-1} P_{a1}(S_a = j \quad \text{at } t_{b1})\bar{F}_{a2}(\tau_{a2} \mid S_a = j \quad \text{at } t_{b1})} .$$

A special case of this model is when both g_a and g_b are gamma with k features and parameter ν, and ν does not change across stages. The two intensity functions (conditioned and not conditioned on state S_a, respectively) are then elementary:

$$I_{a2}(\tau_{a2} \mid S_a = j \quad \text{at } t'_{b1} = t_{b1}) = \frac{\dfrac{\nu(\nu\tau_{a2})^{k-j-1}\, e^{-\nu\tau_{a2}}}{(k-j-1)!}}{\displaystyle\sum_{i=0}^{k-j-1} \frac{(\nu\tau_{a2})^i\, e^{-\nu\tau_{a2}}}{i!}}$$

and

$$I_{a2}(\tau_{a2} \mid t'_{b1}) = \frac{\displaystyle\sum_{i=0}^{k-1} \frac{(\nu t_{b1})^i\, e^{-\nu t_{b1}}}{i!} \cdot \frac{\nu(\nu\tau_{a2})^{k-i-1}\, e^{-\nu\tau_{a2}}}{(k-i-1)!}}{\displaystyle\sum_{i=0}^{k-1} \frac{(\nu t_{b1})^i\, e^{-\nu t_{b1}}}{i!} \sum_{j=0}^{k-i-1} \frac{(\nu\tau_{a2})^j\, e^{-\nu\tau_{a2}}}{j!}}$$

A more general model than this parallel gamma model is obtained by allowing ν to be dependent on the stage and on the item, so that ν in the numerator of $I_{a2}(\tau_{a2} \mid S_a = j \;\&\; t'_{b1} = t_{b1})$ would become $\nu(a; 2) = \nu_{a2}$, and the rate in the denominator (representing the processing rate during Stage 1 for a) would be $\nu\,(a; 1) = \nu_{a1}$. A very general model would result from allowing ν to be also dependent on the times of completion and on the state the item is in, thus turning the ν for a at Stage 2 into $\nu\,(a; 2; j; t_{b1})$.

The discussion has been emphasizing the description of the intensity function by way of the associated densities, survivor functions, etc. We can, however, choose to take the intensity function as the "given" and derive the other ways of describing the distribution (distribution function, density function, etc.) from the intensity function. The nice aspect of this is that almost *any* function of time, $u(t) > 0$ for all $t \geqslant 0$, can be *defined* as an intensity function and the associated distribution function produced by

$$F(t) = 1 - e^{\int_0^t u(t')dt'}$$

or the survivor function simply as

$$\bar{F}(t) = e^{-\int_0^t u(t')dt'}$$

It is clearly necessary that $\lim_{t\to+\infty} \int_0^{\infty} u(t')\,dt' = +\infty$. Of course, we can also make $u(t) = I(t)$ depend on the state of processing as well as previous processing times. Thus, if in our present developments, $I_{a2}(\tau_{a2} \mid t'_{b1} = t_{b1})$ is specified, then the survivor function is $\exp\,[-\int_0^{\tau_{a2}} I_{a2}(\tau_{a2} | t'_{b1} = t_{b1}) d\tau'_{a2}]$, and the density function for this particular intercompletion time is just

$$I_{a2}(\tau_{a2} \mid t'_{b1} = t_{b1}) \exp \left[- \int_0^{\tau_{a2}} I_{a2}(\tau'_{a2} \mid t'_{b1}) \, d\tau'_{a2} \right]$$

One way of viewing the latter expression is that we can conceive of any stochastic process (e.g., on intercompletion times) as a sort of time-varying exponential process. Of course, the distribution is truly exponential if and only if the intensity function I is constant (not varying with time).

Using the intercompletion time distributions is particularly appropriate when the successive intercompletion times are independent. This occurs with exponential distributions but is by *no means* confined to them. To be sure, some systems or their models may be more easily described by the intensity functions on the completion times themselves, and capacity can still be compared between two models, for example, by selecting the analogous states of processing in the two models. We find in this connection, not surprisingly, that independent parallel models are nicely depicted by intensity functions of the individual *total completion time* (this is the time from $t = 0$, when the system begins processing anything, to when the item under study is completed; Townsend, 1974a). Such intensity functions for an element may depend on time and state for that element but cannot depend on what is going on within the other items. Serial systems, on the other hand, are usually more simply described via the intercompletion time distributions, because in serial processing the intercompletion times are the actual processing times on individual items.

SUMMARY OF MAJOR RESULTS

We first developed a feature-matching and letter-recognition model based on a linear correlator with output

$$y_K = \sum_{j=1}^{K} r_{ij} s_{rj}$$

where r_{ij} is the jth discrete sample of feature i, and s_{rj} represents the analogous sample from memory feature s_r. Distributions of y_K were then obtained when the input signal contained white Gaussian noise and the memory signal was riding a DC component of amplitude b. There it was found that y_K is distributed as

$$\begin{array}{l|l|l} & \multicolumn{2}{c}{\textit{Feature } i} \\ & \textit{Input} & \textit{Memory} \\ \left\{\begin{array}{l} N(ba \sum_{j=1}^{K} s_{ij}^2, b^2 \sum_{j=1}^{K} s_{ij}^2) \\[2ex] N(0, b^2 \sum_{j=1}^{K} s_{ij}^2) \\[2ex] N(ba \sum_{j=1}^{K} s_{ij}, Kb^2) \\[2ex] N(0, Kb^2) \end{array}\right. & \begin{array}{l} \text{if present} \\[2ex] \text{if absent} \\[2ex] \text{if present} \\[2ex] \text{if absent} \end{array} & \begin{array}{l} \text{present} \\[2ex] \text{present} \\[2ex] \text{absent} \\[2ex] \text{absent} \end{array} \end{array},$$

where a represents the strength of the input signal relative to the memory signal. Signal-to-noise ratios for these cases were discussed.

The concepts of power and energy were developed within both the time and frequency domain. The notion of a set or ensemble of power curves and a distribution on these as representing the ability of a system to produce energy played a central role in these parts of the chapter. Some measures of capacity were suggested that included both the energy, Q, consumed by the system, and the time, T, taken by the system to expend that energy. These measures were of the form Q/T. Note that increasing either the energy or the speed increases these measures. Also suggested as a possible measure of effective capacity, which is proportional to the power, P, allocated to the filter, was

$$\frac{[d'(T)]^2 + APt_0}{\mathrm{RT}} = AP$$

where APt_0 is the intercept of $[d'(T)]^2$ as a function of RT, and A is a constant of proportionality. These results were discussed in light of some recent empirical studies.

Then emphasizing speed alone, we developed three very general measures of capacity; that is, the measures were in the form of orderings on intensity (hazard) functions, distribution functions, and the underlying random variables. It was shown that a dominance relationship exists among these measures, that is

$$[I_A(t) > I_B(t)] \Longrightarrow [F_A(t) > F_B(t)] \Longrightarrow [P(t_A < t_B) > \tfrac{1}{2}] \qquad \text{for all } t > 0$$

but that the reverse ordering does not necessarily hold. It was then demonstrated (using the intensity function ordering) how one can use these measures to: (1) investigate the capacity at any stage of processing; and/or (2) construct any particular capacity architecture in a model one wishes.

ACKNOWLEDGMENTS

The final stages of preparation of this manuscript were completed at the Technical University of Braunschweig during a Visiting Professorship of the first author sponsored by the Deutsche Forschungsgemeinschaft. Both authors were on leave there and are indebted to the Deutsche Forschungsgemeinschaft and the Institute for Psychology. Joseph Lappin provided a critical reading, and John Castellan and Stephen Link gave helpful editorial assistance. We also thank Leslie Waugh Townsend for assistance in preparation of this chapter.

REFERENCES

Anderson, J. A. A theory for the recognition of items from short memorized lists. *Psychological Review*, 1973, *80*, 417-438.

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

Cornsweet, T. N. *Visual perception*. New York: Academic Press, 1970.

Cox, D. R. *Renewal theory*. London: Methuen, 1962.

Estes, W. K. An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.). *Coding processes in human memory*. New York: Winston-Wiley, 1972.

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

Green, D. M., & Luce, R. D. Speed-accuracy trade-off in auditory detection. In S. Kornblum (Ed.), *Attention and performance, IV*, New York: Academic Press, 1973.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

Gumbel, E. J. *Statistics of extremes*, New York: Columbia University Press, 1958.

Hoffman, J. E. Hierarchical stages in the processing of visual information. *Perception & Psychophysics*, 1975, *18*, 348-354.

Kahneman, D. K. *Attention and effort*. Englewood Cliffs: Prentice-Hall, 1973.

Kantowitz, B. H., & Knight, J. L., Jr. Testing tapping timesharing, II: Auditory secondary task. *Acta Psychologica*, 1976, *40*, 343-362.

Lappin, J. S., & Disch, K. The latency operating characteristic: I. Effects of stimulus probability on choice reaction time. *Journal of Experimental Psychology*, 1972, *92*, 419-427.

Link, S. W. The relative judgment theory of two choice response times. *Journal of Mathematical Psychology*, 1975, *12*, 114-135.

Luce, R. D. Detection and recognition. In R. D. Luce, R. B. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. I). New York: Wiley, 1963.

Luce, R. D., & Green, D. M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, *19*, 14-57.

Massaro, D. W., & Schmuller, J. Visual features, pre-perceptual storage, and processing time in reading. In *Understanding Language*. New York: Academic Press, 1975.

McGill, W. J. *Introduction to counter theory in psychophysics*, Unpublished manuscript, 1966.

McGill, W. J., & Gibbon, J. The general-gamma distribution and reaction times. *Journal of Mathematical Psychology*, 1965, *2*, 1-18.

✓ Pachella, R. G. The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Potomac, Md.: Lawrence Erlbaum Associates, 1974.

Padulo, L., & Arbib, M. A. *System theory*. Phila., Pa.: Saunders, 1974.

Pew, R. W. Human perceptual-motor performance. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Potomac, Md.: Lawrence Erlbaum Associates, 1974.

Pike, R. Response latency models for signal detection. *Psychological Review*, 1973, *80*, 53-68.

Schouten, J. F., & Bekker, J. A. M. Reaction time and accuracy. *Acta Psychologica*, 1967, *27*, 223-229.

Smith, H. M. *Principles of holography*. New York: Wiley-Interscience, 1969.

Smith, E. E., & Spoehr, K. T. The perception of printed English: A theoretical perspective. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and Cognition*. Potomac, Md.: Lawrence Erlbaum Associates, 1974.

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

Sternberg, S. Two operations in character recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 1967, *2*, 45-53.

Sternberg, S. *Evidence against self-terminating memory search from properties of RT distributions*. Paper presented at the annual meeting of the Psychonomic Society, St. Louis, November, 1973.

Swanson, J. M., & Briggs, J. E. Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, 1969, *81*, 223-239.

Swensson, R. G., & Thomas, R. E. Fixed and optimal stopping models for two-choice discrimination times. *Journal of Mathematical Psychology*, 1974, *11*, 213-236.

Taylor, D. A. Holistic and analytic processes in the comparison of letters. *Perception & Psychophysics*, 1976, *20*, 187-190.

Taylor, M. M., Lindsay, P. H., & Forbes, S. M. Quantification of shared capacity processing in auditory and visual discrimination. *Acta Psychologica*, 1967, *27*, 223-229.

Theios, J., Smith, P. G., Haviland, S. E., Traupmann, J., & Moy, M. C. Memory scanning as a serial self-terminating process. *Journal of Experimental Psychology*, 1973, *97*, 323-336.

Thomas, E. A. C. Sufficient conditions for monotonic hazard rate. An application to latency-probability curves. *Journal of Mathematical Psychology*, 1971, *8*, 303-332.

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

Townsend, J. T. 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*, Potomac, Md.: Lawrence Erlbaum Associates, 1974. (a)

Townsend, J. T. *Analysis of search reaction time distributions and independent parallel models*. Paper presented at seventh annual Mathematical Psychology Meeting, Ann Arbor, August, 1974. (b)

Townsend, J. T. A stochastic theory of matching processes. *Journal of Mathematical Psychology*, 1976, *14*, 1-52.

Townsend, J. T., & Ashby, F. G. *Toward a theory of letter recognition: Testing contemporary feature models*. Unpublished manuscript, 1976.

Townsend, J. T., & Roos, R. N. Search reaction time for single targets in multiletter stimuli with brief visual displays. *Memory and Cognition*, 1973, *1*, 319-332.

Weisstein, N. Beyond the yellow-Volkswagen detector and grandmother cell: A general strategy for the exploration of operations in human pattern recognition. In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology: The Loyola Symposium*, New York: Winston-Wiley, 1973.