JOURNAL OF MATHEMATICAL PSYCHOLOGY 37, 526–555 (1993)

# Response Time Distributions in Memory Scanning

F. GREGORY ASHBY

*University of California at Santa Barbara*

JENN-YUN TEIN

*Arizona State University*

AND

J. D. BALAKRISHNAN

*University of California at Santa Barbara*

This article reports the results of a memory scanning experiment (S. Sternberg, 1966, *Science*, **153**, 652–654) in which each of four subjects participated in about 1500 experimental trials per memory set size. These large samples made it possible to test a number of important nonparametric (i.e., model-free) properties of the response time (RT) distributions. These properties place severe constraints on the various memory scanning models and they provide a deeper description of the data than summary statistics or goodness-of-fit values. Five conclusions stood out. First, increasing the size of the memory set induced the strongest possible form of stochastic dominance on both target present and target absent trials. Second, the RT hazard functions were nonmonotonic, thereby falsifying a large class of serial search models. Third, strong evidence was obtained against an exhaustive search. Fourth, some evidence was found that adding an item to the memory set inserts a stage with exponentially distributed duration into the processing chain, at least on target absent trials. Fifth, the data supported the hypothesis that three of the subjects stored the representations of the memory set items in a visual short-term memory system and the fourth subject used an acoustic short-term system. To our knowledge, the only extant model of memory scanning that is consistent with all these results assumes that search is parallel, self-terminating, and of very limited capacity. 

In almost any domain of experimental psychology, the time taken to perform a task is likely to carry some important information. Consequently, the study of response times has been of central concern for many years and a large body of

This research was supported in part by National Science Foundation Grants BNS88-19403 and DBS92-09411 to the first author. Some of this research was conducted at Ohio State University. We thank Ben Murdock, Roger Ratcliff, and James Townsend for their helpful comments and Bruce Bloxom for making available the software for quadratic spline estimation. Correspondence concerning this article and requests for reprints should be addressed to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, CA 93106.



0022-2496/93 $5.00

empirical and theoretical findings has been amassed (e.g., Luce, 1986; Townsend & Ashby, 1983; Welford, 1980). As might be expected, however, theoretical developments have progressed more rapidly than empirical ones, in the sense that a great many theoretical results and predictions have been derived that have never been properly tested. In many cases, this is because these results depend on complicated properties of the response time (RT) distributions, and the accurate estimation of these distributions will typically require unusually large sample sizes.

Response time distributions have been carefully estimated in simple reaction time tasks (e.g., Burbeck & Luce, 1982; Green & Luce, 1971; Kohfeld, Santee, & Wallace, 1981), but less empirical work has been done along these lines in the more cognitive tasks, such as memory scanning, visual search, and same–different (although see Ashby, 1982; Balakrishnan & Ashby, 1992; Ratcliff, 1978). Possibly this shortcoming is because, in general, the more cognitive the task the greater the number of potential independent variables. Rather than carefully estimate RT distributions under one specific set of conditions, cognitive psychologists have preferred to explore the effects of new independent variables.

This article reports the results of a memory scanning experiment (Sternberg, 1966) in which each of four subjects participated in about 1500 experimental trials per memory set size. These large sample sizes allow fairly accurate estimates of the RT distributions and therefore, make it possible to test empirically a number of important distributional properties. A great many mathematical models of memory scanning have been proposed (see, e.g., Townsend & Ashby, 1983, for a review), and a carefull comparison of all these is beyond the scope of this article. Instead, we focus on nonparametric (i.e., model-free) properties of the RT distributions. These properties place severe constraints on the various models, so they serve to falsify large classes of competing models. They also provide a deeper description of the data than summary statistics or goodness-of-fit values, so they can be used to evaluate models that might be developed sometime in the future.

## The Memory Scanning Experiment

On each trial of a standard memory scanning task (Sternberg, 1966), the subject is first shown a list of alphanumeric characters, called the memory set. Following a brief delay, a single character, called the target, is shown to the subject, whose task is to respond YES or NO as quickly as possible, depending on whether the target was or was not a member of the memory set. In the varied set procedure, the memory set is changed on each trial. Stimulus conditions are arranged so that perfect accuracy is possible and RT is the dependent variable of primary interest. The most popular independent variable in these experiments is memory set size.

Perhaps the largest class of memory scanning models assume that the items in the memory set are stored in some short-term memory buffer and that the subject systematically searches this buffer for the target item. The models in this class, which differ greatly in their assumptions about the search process, can be categorized along a number of different dimensions. For example, serial models assume a sequential search, whereas parallel models assume that search is simultaneous. Self-terminating models assume the search is halted as soon as the target is discovered, whereas exhaustive models assume the entire short-term memory buffer is searched on every trial. Discrete-stage models assume that later processing stages do not begin until earlier stages are completed, whereas continuous-flow models assume that information flows from one stage to another in a continuous fashion. A number of the properties that are tested below were derived by assuming one of these types of processing. While it may be tempting to interpret a violation of the property as evidence in support of the contrasting category, it must be remembered that none of these categorizations is all inclusive. It is fairly easy to formulate models that are neither serial nor parallel, neither self-terminating nor exhaustive, and neither discrete-stage nor continuous-flow.

Most of the search models of memory scanning make predictions about response time but not about response accuracy. In the simpler, two alternative forced choice paradigm, models predicting both dependent variables are common. Many of these are based on diffusion processes, random walks, or counting processes (i.e., the so-called accumulator models). Because the memory scanning task uses two responses but more than two stimulus alternatives (at least when the memory set size is at least two), the cost of naively applying most of these models to the memory scanning task is the loss of a detailed processing interpretation. For example, traditional random walk models would associate one barrier with a NO response and one barrier with a YES response. On target present trials, the drift would tend toward the YES barrier. The effect of increasing memory set size would be to decrease the drift rate. Although such a model might be able to account for the major empirical results, it would reveal little about the underlying search process, because it makes no assumptions about why drift rate decreases with memory set size. One would do almost as well with a model that simply assumed RT increases with memory set size.

Search models that predict both response time and response accuracy have been formulated. Perhaps the most successful of these is Ratcliff's (1978) diffusion model. This is a parallel, self-terminating model that assumes *each comparison* can be modelled by a diffusion process. Although no nonparametric tests of the diffusion model are known, because it assumes a parallel, self-terminating search it is constrained by the serial versus parallel and the self-terminating versus exhaustive tests considered below.

An important alternative to the search models of memory scanning is provided by the so-called strength models (e.g., Baddeley & Ecob, 1973; Cavanagh, 1976; Murdock, 1985; Wickelgren & Norman, 1966). Strength models assume direct access to memory, so search time is instantaneous (or at least a constant). When the memory trace of the target item is accessed, its trace strength is compared to a criterion value. If it exceeds the criterion then a positive response is given, and if it falls below the criterion then a negative response is given. RT is assumed to vary inversely with the absolute value of the difference between the trace strength of the target item and the criterion value. Strength models make predictions about both response accuracy and response latency, but again, no nonparametric tests are known.

Monsell (1978) showed that on both YES and NO trials, RT is affected by how recently the target item appeared as a stimulus—a result he interpreted as favoring strength models over search models. Such a result does falsify the simplest class of search models, but it is straightforward to construct models in which search is through a list of items larger than the memory set and in which search rate depends on trace strength. This more general class of search models should have no more difficulty with the Monsell results than do the strength models.

A detailed description of the various theoretical properties that were tested is given after the Method section.

## METHOD

*Subjects.* Three undergraduate students and one of the experimenters participated in the study. The students were paid assistants in the laboratory, and each of the four had normal or corrected-to-normal vision.

*Stimuli and Apparatus.* The stimulus set consisted of the 11 lower-case consonants: c, d, f, j, k, l, n, p, s, v, and z. Subjects were seated in a small booth with soft lighting and viewed an HP2000 CRT display at approximately eye-level. Stimulus characters were generated by a Megatek graphics generator with a resolution of $4096 \times 4096$. The order and timing of the displays were controlled by computer. A two-button response board was connected to the computer, providing both the accuracy and response time measurements.

*Procedure.* The design followed the varied set paradigm outlined by Sternberg (1966). Before each trial, a new memory set containing $n$ letters (for $n = 2$, 3, 4, or 5) was selected at random and without replacement from the total set of 11 stimulus letters. The letters in the memory set were then presented simultaneously and in a linear array for a duration of $n$ s. The display corresponding to the largest memory set size subtended a visual angle of about 3.25°. After a delay of 2 s, the target letter was presented. The target was a member of the memory set on 50% of the trials. The subject responded YES by pressing a button on the right of the response board, and NO by pressing a button on the left. Subjects were instructed to respond as quickly as possible without sacrificing accuracy, and were allowed a maximum of 5 s to respond.

Two complete practice sessions (330 trials) were followed by 15 experimental sessions for each subject. Each experimental session began with 10 warm-up trials, followed by 330 experimental trials divided into 6 blocks (55 trials per block). Accuracy feedback (percent correct) was provided at the end of each block.

## Theoretical Properties, Results, and Discussion

Let $\mathbf{RT}_k$ denote the response time on a trial in which the memory set size is $k$. Denote the mean or expected value of $\mathbf{RT}_k$ by $E(\mathbf{RT}_k)$, the probability density function by $f_k(t)$, and the cumulative probability distribution function $P(\mathbf{RT}_k \leqslant t)$ by $F_k(t)$.

### *Error Rates*

The error rates are given in Table 1. Note that they are uniformly small. For Subject 1, overall error rate was about 2.2% and for Subjects 2–4, the overall error rates were less than 2.0%. For example, when the memory set size was 2, Subject 2 was incorrect on only 1 of 639 target present trials and on only 7 of 612 target absent trials. These small sample sizes make it impossible to accurately estimate the RT distributions on trials when subjects were incorrect. In fact, even incorrect mean RT cannot be accurately estimated. Therefore, all RT analyses were restricted to trials on which a correct response was made in less than 3 s.

Schweickert (1985) showed that if the probability of a correct response is the product of the probabilities that each stage in an independent serial process is executed correctly, then experimental factors that affect different stages will have additive effects on log percent correct. In memory scanning experiments, it is

TABLE 1

Error Rates by Memory Set Size and Trial Type

| | | Trial type | |
|---|---|---|---|
| Subject | Set size | Target absent | Target present |
| 1 | 2 | 2.4% | 1.3% |
| | 3 | 2.8% | 1.6% |
| | 4 | 2.6% | 0.8% |
| | 5 | 3.0% | 2.1% |
| 2 | 2 | 1.1% | 0.2% |
| | 3 | 2.4% | 0.5% |
| | 4 | 4.6% | 0.8% |
| | 5 | 4.2% | 1.4% |
| 3 | 2 | 1.6% | 0.8% |
| | 3 | 1.9% | 1.3% |
| | 4 | 1.8% | 1.6% |
| | 5 | 1.9% | 1.9% |
| 4 | 2 | 1.3% | 0.7% |
| | 3 | 1.4% | 0.9% |
| | 4 | 2.4% | 1.7% |
| | 5 | 1.2% | 2.0% |

natural to test this prediction on the factors of memory set size and response type (i.e., YES versus NO). Although a number of nonmonotonicities exist, Table 1 indicates that in accord with this prediction, error rates tend to increase with memory set size. Unfortunately, however, small error rates cause large standard errors on log percent correct and so the test is of dubious value in the present case.

## *The Effect of Increasing the Processing Load*

### *Theoretical Properties*

One of the most popular assumptions in RT theory is that adding an item to the memory set should increase processing time and thus that $\mathbf{RT}_k$ should be stochastically greater than $\mathbf{RT}_{k-1}$. Theoretically, there are many levels at which such a stochastic dominance can be established (Townsend, 1990; Townsend & Ashby, 1978, 1983). Among the weakest is at the level of the mean or expected RT. That is, we could conclude that larger memory sets require at least as much processing as smaller memory sets if

$$E(\mathbf{RT}_k) \geqslant E(\mathbf{RT}_{k-1}) \tag{1}$$

for all values of $k$ and for both target present and target absent conditions. This prediction has been supported in virtually every varied set memory scanning study that has been reported (e.g., Sternberg, 1975). In fact, a common finding is that mean RT increases linearly with memory set size.

A stronger form of stochastic dominance is an ordering at the level of the cumulative distribution functions:

$$F_{k-1}(t) \geqslant F_k(t), \qquad \text{for all} \quad t > 0. \tag{2}$$

Such a dominance is stronger in the sense that an ordering of the cumulative distribution functions implies an ordering of the means, but an ordering of the means does not guarantee an ordering of the cumulative distribution functions. Townsend and Ashby (1983) found strong empirical support for the Eq. (2) ordering in the memory scanning data of Townsend and Roos (1973).

An even stronger form of stochastic dominance involves the so-called hazard function, which is defined by

$$h_k(t) = \frac{f_k(t)}{1 - F_k(t)}. \tag{3}$$

At each particular time $t$, the hazard function gives the conditonal probability density that a response will be given in the next instant, given that one has not yet occurred. The hazard function, also called the age specific failure rate, is used heavily in reliability theory as a model for the life span of a system component (e.g., Barlow & Proschan, 1965). In this context, $t$ represents the time at which the system first fails. For many physical systems, such as a light bulb, the hazard function initially rises because of the possibility that the component was defective when initially installed. After it has functioned successfully for a fixed time, the possibility that it is defective is diminished and so the hazard function begins to fall. Eventually, because of natural aging, the hazard function gradually begins to rise again. An ordering of the hazard functions

$$h_{k-1}(t) \geqslant h_k(t), \qquad \text{for all} \quad t > 0 \tag{4}$$

implies an ordering of the distribution functions, but the reverse is not true (e.g., Townsend & Ashby, 1978, 1983). To our knowledge, the Eq. (4) ordering has not been empirically tested for memory scanning data (although see Ratcliff, 1988; Townsend, 1990).

The hazard function completely characterizes a distribution, in the sense that the density function and cumulative distribution functions can be expressed in terms of the hazard function via

$$f(t) = h(t) \exp\left[-\int_0^t h(x)\, dx\right] \tag{5}$$

and

$$F(t) = 1 - \exp\left[-\int_0^t h(x)\, dx\right]. \tag{6}$$

Even so, hazard functions are of interest in their own right. For example, they are associated with many important theoretical properties (for a list, see Bloxom, 1984; Luce, 1986; Townsend & Ashby, 1983). In addition, in many cases the hazard function is more visually informative than the density function. This is especially true when the tails of the distributions are important (Luce, 1986, pp. 18–19). For example, the density and distribution functions of Rayleigh and Gamma distributions appear very similar (both densities are positive valued, unimodal, and positively skewed), but the hazard functions are strikingly different. The Rayleigh hazard rate increases linearly with $t$ whereas the Gamma hazard rate initially increases but then asymptotes at a constant value.[1]

Finally, an even stronger form of dominance occurs if the likelihood ratio

$$l_k(t) = f_k(t)/f_{k-1}(t) \tag{7}$$

is nondecreasing in $t$. Once again the dominance is stronger in the sense of the unidirectional implication (e.g., Ross, 1983; Townsend, 1990; Townsend & Ashby, 1983). As with hazard functions, likelihood ratios in memory scanning experiments have never been examined (although see Townsend, 1990).

[1] Of course, a large sample size is needed to discriminate between two distributions that differ primarily in the tail.

To summarize the dominance relations, the following implications hold:

$$
\begin{aligned}
&l_k(t) \text{ nondecreasing} \quad \text{for all} \quad t \geq 0 \\
&\Rightarrow h_{k-1}(t) \geq h_k(t) \quad \text{for all} \quad t \geq 0 \\
&\Rightarrow F_{k-1}(t) \geq F_k(t) \quad \text{for all} \quad t \geq 0 \\
&\Rightarrow E(\mathbf{RT}_k) \geq E(\mathbf{RT}_{k-1}).
\end{aligned}
$$

*Results*

*Mean RT Ordering.* The mean RT versus memory set size functions are presented in Fig. 1 for correct responses only. For Subjects 1, 2, and 4 the average standard error associated with each mean RT is 6.1 ms and for Subject 3 it is 11.4 ms. With one exception, each subject shows the expected mean RT increase with memory set size for both target present and target absent trials (all differences significant with $p < 0.01$). The exception occurs for Subject 3 on target present trials when the set size is 3 or 4. This slight mean RT decrease is not statistically significant ($p > 0.2$).

*Cumulative Distribution Function Ordering.* The estimated cumulative distribution fuctions for target present trials are shown in Fig. 2. Note that there are relatively few cross-overs, so it appears that the functions are ordered by set size. The distribution function estimates on target absent trials appear even more regular and ordered than those on target present trials. Further support for

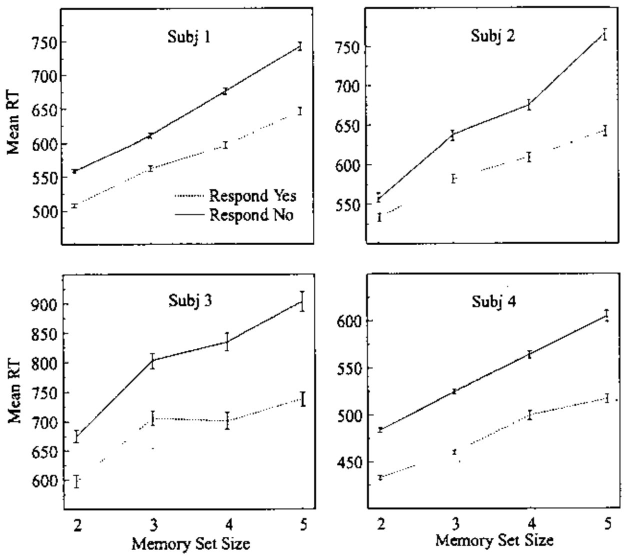


FIG. 1. Target present and target absent mean RT as a function of memory set size for each of the four subjects.

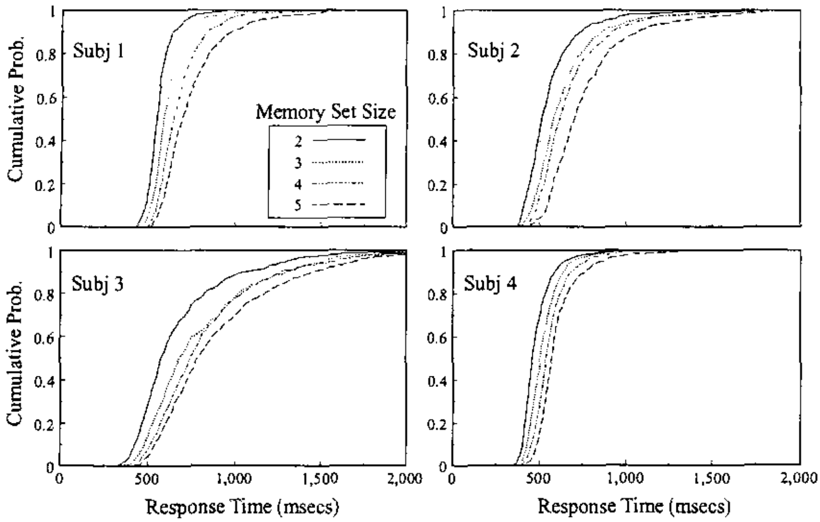


FIG. 2. RT cumulative distribution function estimates on target present trials for each of the four subjects.

stochastic dominance at this level is obtained by Kolmogorov–Smirnov tests of the null hypothesis $H_0$: $F_{k-1}(t) \geqslant F_k(t)$ for all $t > 0$, against the alternative $H_1$: $F_{k-1}(t) < F_k(t)$ for some $t > 0$, for each pair of consecutive memory set sizes (e.g., Walsh, 1965). For these data the null hypothesis of a distribution function ordering was not rejected in any case (with $\alpha = 0.01$).

*Hazard Function Ordering.* Two different methods for estimating the RT hazard functions were used: the random smoothing technique of Miller and Singpurwalla (1977) and the quadratic spline estimator of Bloxom (1985). Burbeck and Luce (1982) used the random smoothing method to estimate hazard functions in a simple RT task, but to our knowledge, we are the first, apart from Bloxom himself, to use the spline estimator.

Computation of the random smoothing estimates proceeds as follows. First, let $Z_j$ represent the $j$th smallest of the $n$ observed RTs. Next define the normalized spacings by

$$S_1 = nZ_1, \qquad S_2 = (n-1)(Z_2 - Z_1), \ldots,$$
$$S_i = (n-i+1)(Z_i - Z_{i-1}), \ldots, \qquad S_n = Z_n - Z_{n-1}.$$

Now let $k$ be the number of samples used to estimate the hazard rate within each fixed interval of time. An estimate of the hazard rate during the $i$th interval is given by

$$\hat{h}(i) = \frac{k}{\sum_{j=i-k+1}^{i} S_k},$$

where $k < i < n$. For the first $k-1$ intervals, set $k = i$. This definition is straightforward except in cases where the same RT is recorded on several trials. Specifically, suppose $Z_i = Z_j = \cdots = Z_k$ for some $i$ through $k$. In this case, let the hazard rate for the previous time interval $Z_i - Z_{i-1}$ be the mean of the rates computed for each $i$ through $k$. Note that hazard rates estimated by this procedure will be constant over each interval $(Z_j, Z_{j-1})$ and thus the resulting estimate is a step function.

One of the most attractive features of the random smoothing technique is that it requires no initial assumptions about the general form of the functions to be estimated. In addition, it is an adaptive filter in the sense that the sample sizes used to estimate the hazard rate at each point in time are equalized by adjusting the size of the time window according to the number of samples available locally.

Random smoothing estimates of the hazard functions for each subject and condition were obtained using a smoothing constant of $k = 100$. To aid visual examination, the resulting estimates were then passed through a 100-ms moving window (e.g., Green & Luce, 1971). Examples of the results for Subjects 2 and 4 are given in Fig. 3. Although we know of no statistical test for ordered hazard functions, a visual inspection of Fig. 3 indicates only a few cross-overs, and these are concentrated in the target present conditions. The estimates for Subjects 1 and 3 followed much the same pattern. Overall, then, the data indicate reasonable support for the hypothesis that the hazard functions are ordered by memory set size.

Bloxom's (1979; 1985) method of estimating the hazard function involves fitting a separate quadratic polynomial, called a spline, to each decile of the data (using a penalized maximum likelihood procedure). Each pair of adjacent splines are

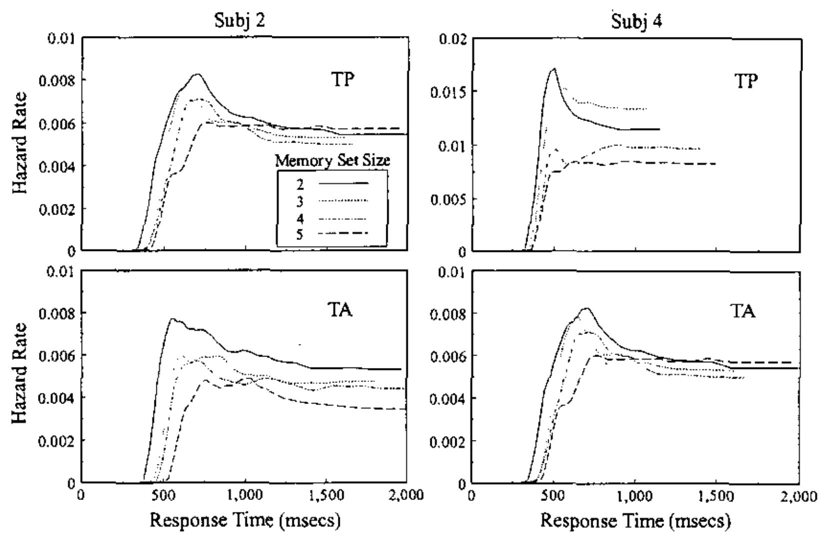


FIG. 3. Random smoothing estimates of the RT hazard functions for Subjects 2 and 4.

480/37/4-4

constrained to have the same first derivative at their point of intersection. In addition, extra side constraints may be placed on the estimation procedure. For example, the hazard function can be estimated under the assumption that it is nondecreasing or that it increases to a peak and then decreases. Goodness-of-fit values are provided for each estimate, thereby allowing a test of the hypothesis that the hazard function is nondecreasing.

For each subject, memory set size, and display type (i.e., target present or target absent), we separately estimated the RT hazard function first under the constraint that it was nondecreasing and then under the constraint that it increased to a peak and thereafter decreased. In roughly half the cases the penalized log likelihood favored the nondecreasing estimates and in roughly half the cases it favored the increasing-then-decreasing estimates.[2] However, a second goodness-of-fit measure overwhelmingly favored the nondecreasing estimates [i.e., the sum of squared deviations between two culmulative RT distribution function estimates; namely, the standard estimate and Eq. (6) with the spline estimate used in place of $h(x)$]. A visual examination revealed that the nondecreasing estimates were approximately ordered by memory set size for all subjects. Only a few small cross-overs occurred. Thus, there is converging evidence from both estimation procedures that the hazard functions are ordered by memory set size.

*Monotonicity of the Likelihood Ratios.* Statistically, likelihood ratios are difficult to estimate. Fortunately, there is a statistically more reliable method for testing whether the likelihood ratio is nondecreasing. A well-known result in signal detection theory states that the likelihood ratio, formed by the signal-plus-noise density function devided by the noise-alone density, is an increasing function of the sensory variable, if and only if the ROC curve is concave[3] (Laming, 1973; Peterson, Birdsall, & Fox, 1954). To take advantage of this result, consider a standard YES–NO signal detection task. Now let $f_{k-1}(t)$ play the role of the noise distribution, $f_k(t)$ play the role of the signal distribution, and $t$ play the role of the sensory variable. Then for any given time $t_i$, the analogue of $P$(hit) is $1-F_k(t_i)$ and the analogue of $P$(false alarm) is $1-F_{k-1}(t_i)$. A latency ROC can be constructed by estimating these probabilities for a number of different values of $t_i$. The RT likelihood ratio $l_k(t)$ is an increasing function of $t$ if and only if the resulting latency ROC is concave. Because the latency ROC involves cumulative distributions

[2] The pattern was unrelated to memory set size. Specifically, it was not generally true that for small memory set sizes increasing-then-decreasing estimates were favored and for large set sizes nondecreasing estimates were favored.

[3] A function $\pi$ is concave over the interval $(a, b)$ if for each $x, y \in (a, b)$ and for each $\alpha$ in the interval $[0, 1]$,

$$\pi[\alpha x + (1-\alpha) y] \geqslant \alpha\pi(x) + (1-\alpha)\pi(y).$$

If the $\geqslant$ is replaced by $\leqslant$, the function is said to be convex (e.g., Royden, 1968). Note that this definition contradicts Laming's (1973) use of these terms.

[actually, survivor functions, $1 - F(t)$] rather than densities, it can be estimated more accurately than the likelihood ratio.

Figure 4 shows the estimated latency ROC curves for Subject 4. Once again the remaining subjects show basically the same results. Note that none of the curves show any significant violations of concavity. The plot of $1 - F_5(t)$ versus $1 - F_4(t)$ in the target present condition falls almost on the main diagonal, which makes it difficult to tell whether concavity holds in this case, but this indicates that $f_4(t)$ and $f_5(t)$ are almost identical and so one does not expect the likelihood ratio to deviate much from a value of $l_5(t) = 1.0$ in this case. Overall, the data support the hypothesis that the likelihood ratio is an increasing function of $t$, in both target present and target absent conditions. Because an increasing likelihood ratio implies

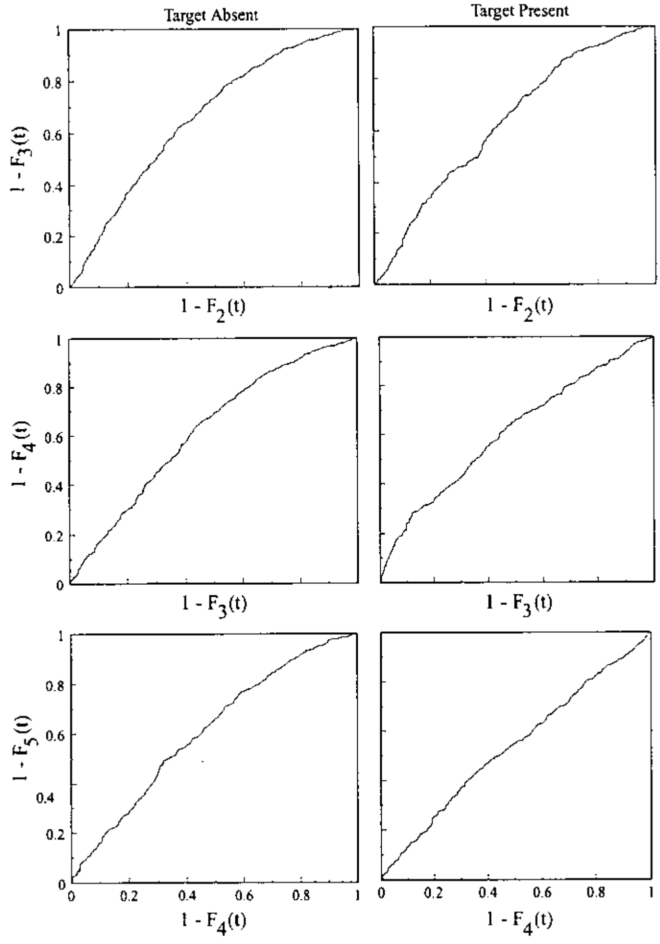


FIG. 4. RT ROC curves for Subject 4. (These curves are concave if and only if the likelihood ratio $f_k(t)/f_{k-1}(t)$ is nondecreasing in $t$.)

ordered hazard functions, these results suggest that the occasional cross-overs found in the hazard function estimates are probably due to statistical artifact rather than to any significant failure of stochastic dominance.

*Summary*

The tests of the stochastic dominance relations indicate that adding an item to the memory set causes an increase in response time that profoundly affects the entire RT distribution. Although this may seem an obvious conclusion, it rules out a number of possible processing models. For example, parallel self-terminating models with unlimited capacity do not predict stochastic dominance at even the mean RT level on target present trials. Other models may predict a low level of stochastic dominance, but not a high level (e.g., ordered hazard functions). For example, a model which assumes that adding an item to the memory set increases RT variance much more rapidly than mean RT predicts a dominance at the level of the mean RTs but not at the level of the cumulative RT distribution functions. A succesful theory of memory scanning must account for the extreme levels of stochastic dominance found in the present data.

## *Serial versus Parallel Search*

*Theoretical Properties*

The shape of the empirical RT hazard function is also of interest because many serial search models predict increasing hazard functions. To see this, consider a serial process with $n$ stages. The hazard rate must initially be low because $n$ successive tasks (i.e., stages) remain to be completed. As each task is completed, fewer remain and so the hazard rate should increase. If $n$ is large and the successive stage durations are statistically independent then, by the Central Limit Theorem, RT will be approximately normally distributed. The hazard function of the normal distribution is strictly increasing. In fact, Barlow, Marshall, and Proschan (1966) showed that if the hazard functions of all stage durations are nondecreasing, then the serial exhaustive RT hazard function is nondecreasing. Serial models therefore, naturally predict nondecreasing hazard functions.

On the other hand, parallel models are not similarly constrained. For example, parallel, exhaustive models predict that RT is the maximum of $n$ stage durations. The limit distribution of the maximum of $n$ independent and identically distributed random variables can be one of three types (e.g., Galambos, 1978). One of these predicts an increasing hazard function but the other two predict the hazard function to rise to a peak and thereafter to decrease (Luce, 1986). Thus, a finding that the RT hazard function is nonmonotonic not only falsifies a very large class of serial models, but in so doing it also provides indirect support for parallelism.

*Results*

In the last section, we saw that the quadratic spline estimators support the hypothesis that the hazard functions are nondecreasing. In contrast, Fig. 3 indicates that the random smoothing estimates may be nondecreasing in $t$ for large memory set sizes but they display a marked peak for small memory set sizes. This pattern was observed in the data of all four subjects. Interestingly, Burbeck and Luce (1982), who also used the random smoothing procedure, found a similar pattern of results in their study of auditory simple RT. Specifically, they obtained nondecreasing hazard estimates for low intensity signals and peaked estimates for high intensity signals, and their estimates were ordered by signal intensity.

Clearly, either the random smoothing technique or the quadratic spline estimators have a consistent bias somewhere in the tails. That is, either the random smoothing technique underestimates the true hazard rate in the tail, which introduces a spurious peak at small memory set sizes, or the quadratic spline estimator overestimates the hazard rate in the tail, which masks a true peak at small set sizes.

In an attempt to resolve this discrepancy, we performed the following additional analysis. First, note that Eq. (6) implies

$$-\log[1-F(t)] = \int_0^t h(x)\,dx$$

and thus

$$\frac{d}{dt}\{-\log[1-F(t)]\} = h(t).$$

The function $1-F(t)$ is known as the survivor function in reliability theory. Thus, the slope of minus the log survivor function at time $t$ equals the hazard function, so if the hazard function is nondecreasing in $t$, then the slope of minus the log survivor function must also be nondecreasing in $t$. In other words, if the hazard function is nondecreasing, then minus the log survivor function is convex.

For each subject, the negative of the log survivor functions was estimated from the cumulative distribution function estimates given in Fig. 2. The resulting estimates showed consistent violations of convexity. Specifically, for small memory set sizes the functions were S-shaped. Thus, this analysis provides converging evidence that memory scanning RT hazard functions are *not* nondecreasing functions of $t$. Instead, like hazard functions in simple RT, they appear to increase to a peak and thereafter to decrease to some nonzero asymptote. Further, the nonmonotonicity decreases with increasing memory set size. As discussed above, these results rule out a large class of serial search models.

## *Self-Terminating versus Exhaustive Search*

### *Theoretical Properties*

Consider those trials of a memory scanning experiment on which the memory set contains the target item. In this case, the subject has enough information to respond YES as soon as the target item is discovered. Search is *self-terminating* if the subject terminates the search as soon as the match is discovered and it is *exhaustive* if search continues through the entire memory set on all trials. Although many other search strategies are possible,[4] these two have received the most attention in the literature. A number of methods for discriminating between self-terminating and exhaustive search have been proposed. In this article we consider six.

First, Sternberg (1966) proposed that exhaustive models predict equal slopes for the mean RT versus memory set size curves for target present and target absent conditions but that self-terminating models predict a shallower slope on target present trials (in fact, simple serial, self-terminating models predict a 2:1 slope ratio). While intuitive, it is well known that this test is imperfect. Exhaustive models can predict unequal target present and target absent slopes and self-terminating models can predict parallel target present and target absent curves (Townsend & Ashby, 1983, pp. 126–128; Townsend & Van Zandt, 1990). In many cases, however, some nonintuitive capacity or processing rate assumptions are required to make these uncharacteristic predictions. This is especially true when exhaustive models try to predict unequal slopes. For example, Townsend and Van Zandt (1990) showed that for serial exhaustive models "sizable slope differences between negative and positive mean reaction time functions are very difficult and usually impossible to predict" (p. 486) and that for parallel exhaustive models "slope differences between positive and negative functions imply super capacity of target processing in many cases" (p. 487). Super capacity requires the processing rate of an individual item to *increase* as the memory set size increases, whereas we expect most biological systems to be of limited capacity (i.e., so the individual item rates decrease as set size increases). The classic empirical result is that the target present and target absent curves are parallel (see Sternberg, 1975, for a review of this literature), although shallower target present slopes are sometimes found (e.g., Briggs & Blaha, 1969; Briggs & Johnsen, 1973; Clifton & Birenbaum, 1970; Schneider & Shiffrin, 1977).

Second, self-terminating models more easily predict serial position effects than exhaustive models. Serial position curves are plots of mean RT versus the serial position of the target item within the memory set display. A serial position effect occurs whenever the serial position curves are not flat. Although exhaustive models can predict serial position effects, the simplest exhaustive models predict flat serial position curves. On the other hand, intuitive self-terminating models can be constructed that can predict any type of serial position curves (including flat). Thus, flat serial position curves reveal little about whether search is exhaustive or self-terminating but pronounced serial position effects strongly indicate self-terminating search, especially when combined with linear target absent and target present mean RT curves (Townsend & Van Zandt, 1990). Strength models, postulating a

[4] For example, after a match is discovered, the subject may try to terminate the search, but because of inertia an additional item or two may be processed before termination is complete. Another possibility is that overall search may be a probability mixture of self-terminating and exhaustive processing.

degenerate form of self-terminating search, also easily predict serial position effects, although whether they predict the correct type of effect is in question (Murdock, 1985). Empirically, serial position effects are common, especially when the delay between the offset of the memory list and the onset of the probe is brief (e.g., Burrows & Okada, 1971; Clifton & Birenbaum, 1970; Forrin & Cunningham, 1973).

Third, it has been noted that most serial, self-terminating models predict that RT variance will increase more sharply with memory set size for target present conditions than for target absent conditions, whereas most exhaustive models (both serial and parallel) predict equal target present and absent slopes (Rossmeissl, Theios, Krunnfusz, 1979; Schneider & Shiffrin, 1977; Townsend & Ashby, 1983, pp. 192–201). On the other hand, most parallel, self-terminating models predict that target present RT variance will increase more slowly than target absent RT variance (Townsend & Ashby, 1983). Thus, equal slopes support an exhaustive search and unequal slopes supports a self-terminating search. The few empirical results that have been obtained are equivocal. Schneider and Shiffrin (1977) found a slight tendency for the target present RT variances to increase faster than the target absent variances, but Rossmeissl *et al.* (1979) found the exact opposite result, namely a slight tendency for the target absent curve to increase faster than the target present curve. The standard error for estimating a variance is $(2\sigma^4/n)^{1/2}$, where $\sigma^2$ is the RT population variance and $n$ is the sample size. Thus, accurate estimation of RT variances requires very large sample sizes (Ratcliff, 1979).

Sternberg (1973) proposed two tests of self-termination that involved examining the RT distribution functions. Specifically, he showed that a large class of serial, self-terminating models make the following two predictions. The first, which Sternberg called the *long RT property*, is simply the cumulative distribution ordering $F_{k-1}(t) \geq F_k(t)$ for all $t > 0$. The second property, called the *short RT property*, states that

$$kF_k(t) \geq (k-1)F_{k-1}(t), \qquad \text{for all} \quad t > 0. \tag{8}$$

Note that this inequality holds trivially for large $t$ since the left side asymptotes at $k$ and the right side at $k-1$. The property is therefore of interest only for small $t$.

Of the serial models that Sternberg (1973) considered, the exhaustive models also predict the long RT property to hold and therefore, the long RT property cannot be used to discriminate between self-terminating and exhaustive search (Townsend & Ashby, 1983, pp. 218–248). On the other hand, the short RT property provides a powerful test between self-terminating and exhaustive search, within the class of serial models and also within the class of independent parallel models (Townsend & Ashby, 1983). For example, most independent parallel, exhaustive models predict the short RT property to fail and most independent parallel, self-terminating models predict it to hold.

Townsend and Ashby (1983) also examined the implications that both properties have on the capacity structure of the processing system. Capacity is said to be *unlimited* if the individual item processing time distributions do not depend on memory set size. If the individual item processing times increase with increases in memory set size, then capacity is said to be *limited* and if they decrease, then capacity is *super*. If the long RT property holds, then the most reasonable conclusion is that capacity is unlimited or limited. If it fails, then super capacity is indicated. If the short RT property holds, then a self-terminating search is strongly supported. If it fails, then search is exhaustive or else capacity is very limited.

Vorberg, Colonius, and Schmidt (1989) derived results similar to the long and short RT properties for parallel, exhaustive models with unlimited capacity. Specifically, they showed that these models predict the long RT property and also the following inequality, which we call the *medium RT property*

$$F_{k-1}(t) \leqslant \tfrac{1}{2}[F_{k-2}(t) + F_k(t)], \qquad \text{for all} \quad t > 0. \tag{9}$$

The medium RT property has not been tested empirically, and little is known about the ability of other models to predict the Eq. (9) inequality.

*Results*

*Target Absent and Target Present Mean RTs.* Figure 1 demonstrates that in all cases, the target present mean RT curves increase more slowly with memory set size than the target absent curves. In general, however, the slope ratio is less than 2:1. Although some exhaustive models (e.g., parallel, super-capacity) can accomodate this slope difference, these results clearly favor self-terminating models (Townsend & Van Zandt, 1990).

*Serial Position Curves.* The serial position curves are shown in Fig. 5. Note first that all subjects show substantial serial position effects. Second, note that Subjects 1, 2, and 4 all show a tendency to respond more slowly when the target is in either end position of the display. They respond fastest when the target is in a middle display position. On the other hand, Subject 3 shows a very different pattern of responding. For this subject, RT tends to increase as the target moves rightward through the display. These results, together with the shallower target present mean RT curves, are extremely difficult for an exhaustive search model to predict (Townsend & Van Zandt, 1990).

One possibility is that Subjects 1, 2, and 4 stored representations of the memory set items in some visual short-term memory system, but Subject 3 stored the items in an acoustic memory system. Under this interpretation, the V-shaped serial position curves of Subjects 1, 2, and 4 are the result of laterality effects, since the shorter RTs were associated with the more foveal items. If the stored representations are of an acoustic nature, then RT should be unaffected by display laterality. The increasing serial position curves of Subject 3 could indicate a left-to-right search through such an acoustic store.

Two other factors support this hypothesis. First, an examination of Fig. 1 indicates that Subject 3 responded more slowly than the other subjects. Second, athough the memory scanning task is usually thought to involve a short-term acoustic store (e.g., Townsend & Roos, 1973), the method of displaying the memory

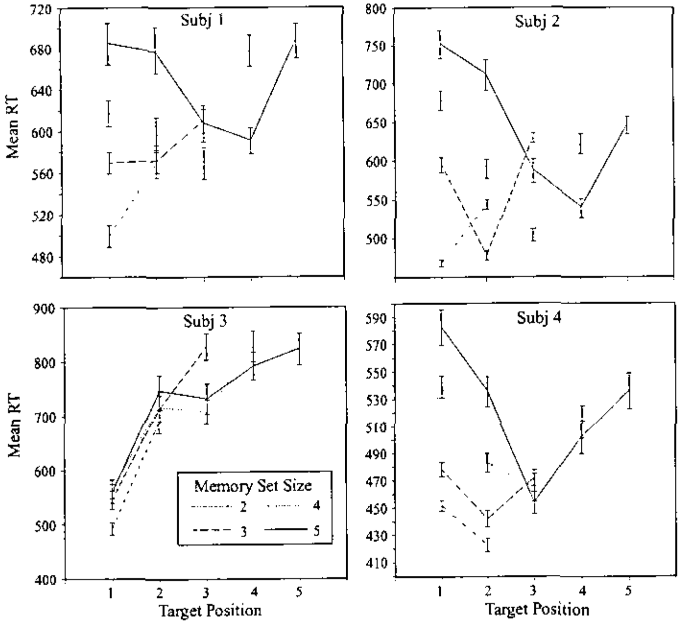


FIG. 5. Serial position curves for each of the four subjects.

set items in the present experiment facilitated storage in a visual form. Sternberg (1966) presented the memory set items sequentially, whereas we presented them simultaneously. Sequential presentation facilitates rehearsal, which presumably facilitates storage in an acoustic form. With sequential presentation, serial position curves often display primacy or recency effects (or both, see, e.g., Sternberg, 1975). Thus, the primacy effects displayed by Subject 3 are consistent with the sequential presentation data. On the other hand, the use of simultaneous presentations makes the memory scanning task more similar to visual search (Atkinson, Holmgren, & Juola, 1969) and in visual search it is thought that subjects search through a visual short-term store (Townsend & Roos, 1973).

The pronounced serial position effects support the hypothesis that the target present RT distributions are a probably mixture of the distributions associated with the different serial positions. Fortunately, this complication has little effect on the present analyses. The tests of stochastic dominance are model free. They merely provide a description of the degree to which one set of RTs dominates another. They make no assumptions about processing. On the other hand, the tests of pure insertion considered below were derived from a specific processing model. However, the model assumes exhaustive processing and so these analyses will be restricted to target absent data.

*Target Present and Target Absent RT Variances.* The target present and target absent RT variance estimates are shown in Fig. 6. For Subjects 2 and 3 the target

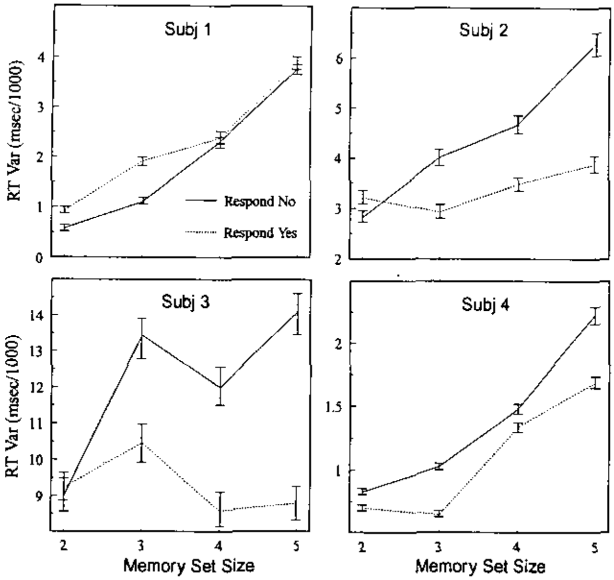


FIG. 6. Target present and target absent RT variance estimates as a function of memory set size for each of the four subjects.

present curves have a shallower slope whereas for Subjects 1 and 4 the curves are roughly parallel. Because of the large sample size, virtually all differences in Fig. 6 are statistically significant.

The shallow target present curves of Subjects 2 and 3 are consistent with the predictions of parallel, self-terminating models. For these models, target present RT is determined completely by the processing time of the target item (and residual processes such as encoding, response selection, and response execution). If capacity is limited, then target item processing time increases with memory set size. In many candidate processing time distributions (e.g., exponential), the variance increases with the mean. Thus, limited capacity, parallel, self-terminating models predict that target present RT variance will increase with memory set size. If, in addition, the processing rate on targets and nontargets is different and capacity is very limited then the increase in RT variance on target present trials may be as great as on target absent trials.

*The Short RT Property.* Although Sternberg's long RT property is mute with respect to the self-terminating versus exhaustive issue, his short RT property, given by Eq. (8), is a good test of self-termination. Figure 7 contains estimates of $(k-1)\,F_{k-1}(t)$ and $kF_k(t)$ on target present trials for values of $k$ from 3 to 5, and for Subjects 2 and 4. Because the short RT property holds trivially for large values of $t$, Fig. 7 is limited to small $t$.

Note that in 4 of the 6 cases, the short RT property is violated for at least some value of $t$. For Subjects 1 and 3 the property is violated in every case. Although

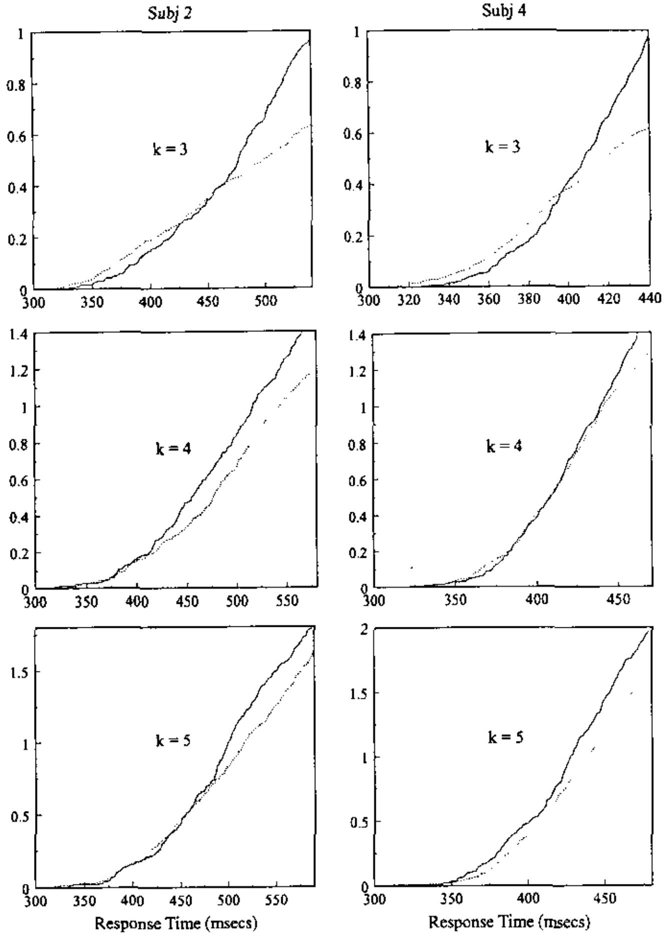


FIG. 7. Plots of $kF_k(t)$ (solid line) and $(k-1)\,F_{k-1}(t)$ (dashed line) for target present conditions for Subjects 2 and 4.

there is no accepted statistical test of the hypothesis that $kF_k(t) \geqslant (k-1)\,F_{k-1}(t)$ for all values of $t$, the fact that violations were observed in 10 of 12 cases is strong evidence against the short RT property. These results, therefore, rule out a large class of self-terminating models. Specifically, we can conclude either that search is not self-terminating or that capacity is very limited (Townsend & Ashby, 1983, pp. 218–248).

*The Medium RT Property.* Figure 8 shows estimates of $F_{k-1}(t)$ and $[F_{k-2}(t)+F_k(t)]/2$ obtained from the target present conditions for all subjects and all relevant values of $k$ (i.e., for $k=4$ and 5). Like many of the other properties examined in this article, we know of no accepted statistical test of the medium RT property. Even so, in 6 of the 8 tests, the ordering is violated for at least some value

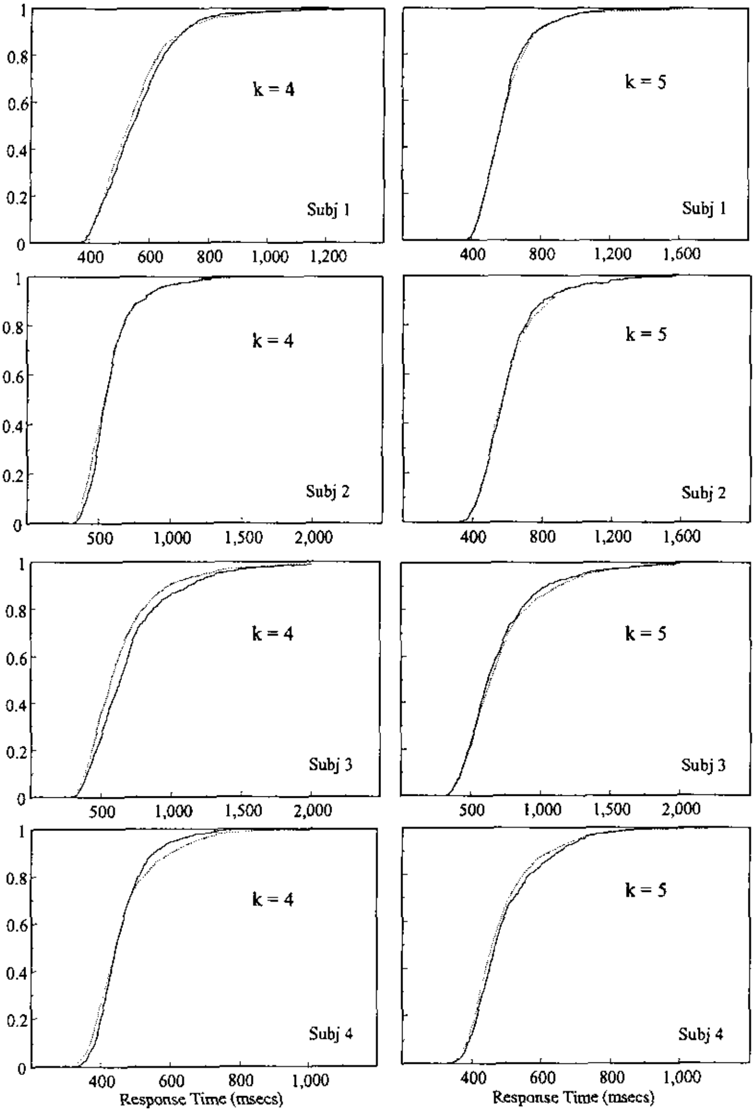


FIG. 8. Plots of $F_{k-1}(t)$ (solid line) and $\frac{1}{2}[F_{k-2}(t)+F_k(t)]$ (dashed line) for target present conditions for each of the four subjects.

of $t$. In one instance (Subject 3, $k=4$), the violations are large. In addition to an exhaustive search, the medium RT property assumes unlimited capacity and parallel processing. It is unknown how robust the property is with respect to violations of either of these latter two assumptions. Even so, these results seem to rule out a large class of unlimited capacity, parallel exhaustive models.

*Summary*

Of the five tests of self-terminating versus exhaustive search considered above, three yielded strong evidence against an exhaustive search for every subject. Of the remaining two tests, the RT variances were inconsistent with exhaustive models for two subjects and inconsistent with self-terminating models for two subjects. The short RT property produced evidence against self-termination for all four subjects. Although these conclusions seem mixed, they are much more consistent with the predictions of self-terminating than exhaustive search strategies. In fact, all of these results could be predicted (at least qualitatively) by a parallel, self-terminating model with very limited capacity. One must also remember that many search strategies exist that are neither self-terminating nor exhaustive. Therefore, evidence against an exhaustive search is not necessarily evidence for self-termination. Perhaps the correct model will postulate some hybrid combination of the two strategies.

## *Discrete-Stage versus Continuous-Flow Processing.*

*Theoretical Properties*

Discrete-stage models assume that later processing stages do not begin until earlier stages have completed processing. In contrast, continuous-flow models assume that information flows continuously from one stage to the next. Although these notions seem quite different, it has been surprisingly difficult to test between them. In an extended review of the relevant literature, Miller (1988) could find "no decisive evidence of continuity" (p. 249). He concluded that "if anything, the available evidence supports discrete transmission" (p. 250). Perhaps the most sophisticated tests, however, were developed by Schweickert (1989). He showed that a large class of continuous-flow models predicts certain observable relations between RT and any number of performance measures (e.g., $d'$), and that these relations are unlikely to be predicted by discrete-stage models. In a preliminary application of these tests, Schweickert (1989) found evidence against the continuous-flow models in two separate experiments. Unfortunately, however, Schweickert's tests require data from a number of speed–accuracy conditions and so they are untestable with the data reported here.

Fortunately, it is somewhat easier to derive predictions that characterize discrete-stage models. For example, when discrete-stage models are applied to memory scanning experiments they almost always assume that adding an item to the memory set causes a separate discrete stage to be inserted into the processing chain. The special case in which the inserted stage has no effect on the duration of any other stage is known as pure insertion (Donders, 1969; Sternberg, 1969). If we let $\mathbf{T}_k$ denote the duration of the inserted stage when there are a total of $k$ items in the memory set, then the assumption of pure insertion can be written as

$$\mathbf{RT}_k = \mathbf{RT}_{k-1} + \mathbf{T}_k. \tag{10}$$

The strongest possible form of pure insertion assumes $\mathbf{T}_k$ and $\mathbf{RT}_{k-1}$ are statistically independent. Although Eq. (10) has a natural serial interpretation, it could also describe a parallel model (Ashby & Townsend, 1980). In this case, rather than an actual processing time, $\mathbf{T}_k$ represents an intercompletion time; that is, the time between the successive completions of separate stages.

Pure insertion is difficult to test rigorously. It is easy to see that Eq. (10) implies an ordering of the mean RTs, but an ordering of means is such a weak form of dominance that, in this case, its empirical verification is not convincing. Stronger tests of pure insertion are possible, however. For example, Eq. (10) implies an ordering of the RT cumulative distribution functions (Ashby, 1982). In fact, Townsend and Schweickert (1989) showed that if the cumulative distribution functions are ordered, then there exists some discrete-stage model that is compatible with the data (although there may not be one in which statistical independence holds). In addition, Ashby (1982) showed that if pure insertion holds and if $h_{k-1}(t)$ is nondecreasing in $t$, then the hazard functions must be ordered. These results emphasize the importance of empirically testing the stochastic dominance orderings, but note that the latter result also makes it important to test whether the hazard functions are nondecreasing in $t$.

A second assumption made by many discrete-stage models is that the duration of one or more of the processing stages is exponentially distributed (e.g., Christie & Luce, 1956; McGill, 1963; Townsend, 1976). For example, let $g_k(t)$ denote the processing time density function for the $k$th processing stage. According to this assumption

$$g_k(t) = V_k \exp(-V_k t). \tag{11}$$

The parameter $V_k$ is known is the processing rate because $E(\mathbf{T}_k) = 1/V_k$. Although this is a difficult assumption to test in isolation, it is known that if the duration of every discrete stage has a nondecreasing hazard function, and if one or more of the stage durations is exponentially distributed, then the tail of the RT hazard function will be flat (Ashby, 1982). Note that the reverse implication is not true. A flat tail on the RT hazard function does not guarantee an inserted stage with an exponentially distributed duration.

Combining the assumption of pure insertion with the assumption that the duration of the inserted stage is exponentially distributed permits much stronger tests. First, Ashby and Townsend (1980) showed that *if and only if* Eqs. (10) and (11) both hold, then

$$f_k(t) = V_k[F_{k-1}(t) - F_k(t)], \qquad \text{for all} \quad t > 0. \tag{12}$$

Equation (12) permits a strong test of both assumptions, since it states that a plot of $f_k(t)$ against $[F_{k-1}(t) - F_k(t)]$ should be linear with positive slope and zero intercept. Further, the slope of the linear regression should predict the mean RT increase that results from adding a k*th* item to the memory set, because if Eqs. (10) and (11) hold, then

$$E(\mathbf{RT}_k) - E(\mathbf{RT}_{k-1}) = 1/V_k. \tag{13}$$

By taking the derivative of both sides of Eq. (12) with respect to time, one can also show that for each $k$, $f_{k-1}(t)$ and $f_k(t)$ should intersect at the mode of $f_k(t)$ (Ashby, 1982). This second test is simpler but not as strong as the Eq. (12) test, since it is not stated in an "if and only if" fashion. Checking the RT density functions for their points of intersection should therefore be a preliminary test of the two assumptions (i.e., Eqs. (10) and (11)). Both tests have yielded tentative support for Eqs. (10) and (11) (Ashby, 1982; Ashby & Townsend, 1980), but more empirical attention is needed.

When interpreting these results, one must also consider the possibility that other models, which do not make the exponential pure insertion assumption, may *mimic* the predicted effects statistically. For example, Ratcliff (1988) showed that his diffusion model predicts results that are similar enough to those of the discrete-stage model that they pass the statistical criteria suggested by Ashby and Townsend (1980).

*Results*

*Pure Insertion.* Pure insertion predicts that the cumulative RT distribution functions are ordered by memory set size and that, if $h_{k-1}(t)$ is nondecreasing in $t$, then the hazard functions are also ordered by memory set size. Although the data satisfy both of these ordering relations, for small memory set sizes the RT hazard functions are nonmonotonic. Thus, we cannot interpret the fact that they are ordered as evidence for pure insertion. On the other hand, because the cumulative distribution functions are ordered, we know at least that some (possibly dependent) discrete-stage model is compatible with the data (Townsend & Schweickert, 1989).

*Exponentially Distributed Processing Durations.* If the duration of all processing stages have nondecreasing hazard functions, and if one or more of these is exponentially distributed, then the tail of the RT hazard function will be flat (Ashby, 1982). Both the random smoothing and the quadratic spline estimates have flat tails in almost every case, and thus the data are consistent with the exponential assumption.

*Inserting a Stage with an Exponentially Distributed Duration.* If pure insertion holds and if the duration of the inserted stage is exponentially distributed (i.e., if Eqs. (10) and (11) hold), then $f_{k-1}(t)$ and $f_k(t)$ should intersect at the mode of $f_k(t)$ (Ashby, 1982) and a plot of $f_k(t)$ versus $F_{k-1}(t) - F_k(t)$ should be linear with zero intercept and slope equal to the rate of the inserted stage (Ashby & Townsend, 1980).

To estimate the RT density functions, we used a Parzen (1962) estimate with a Gaussian kernel,

$$\hat{f}_k(t; n) = \left(\frac{1}{nh\sqrt{2\pi}}\right) \sum_{i=1}^{n} \exp\left[-\left(\frac{t-T_i}{h}\right)^2\right], \tag{14}$$

where $T_i$ is the $i$th observed RT and $n$ is the number of trials. The parameter $h$ is

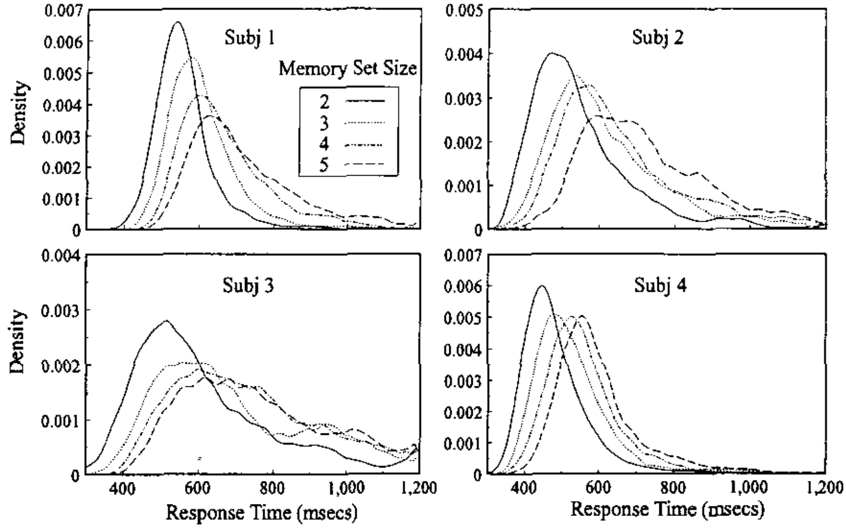


FIG. 9. Parzen estimates (Gaussian kernel) of the RT density functions on target absent trials for each of the four subjects.

a measure of kernel width. If some care is taken in the selection of $h$, the estimate can be shown to be uniformly consistent, provided $f_k(t)$ is uniformly continuous (Parzen, 1962). The asymptotic variance of the Eq. (14) estimator is about half as large as the more widely known histogram estimator (Parzen, 1962). In the present applications, kernel width was set to $h = 10$ ms.

The density estimates are presented in Fig. 9 for target absent trials. To aid visual examination, some of the higher frequency components were attenuated by passing the density estimates through a moving window of width 100 ms (Green & Luce, 1971). This procedure is essentially the same as convolving the estimates with a uniform distribution of zero mean. It does not change the mean of the density estimates, but it could change the mode slightly. By the same token, however, the high frequency components that the smoothing eliminates can make determination of the mode a difficult task.

An examination of Fig. 9 indicates surprisingly good agreement with the modal intersection hypothesis. Some violations are evident, particularly for Subject 3, but an appeal to the stronger Eq. (12) "if and only if" test is clearly warranted.

Figure 10 shows scatter-plots[5] of $f_k(t)$ versus $F_{k-1}(t) - F_k(t)$ for Subjects 2 and 4. Although, there is some variability, overall the results appear reasonably linear, and so a more careful examination is required. Table 2 shows the results of a regression analysis of the Eq. (12) prediction. First, note the values of the correlation coefficient $R$. Except for Subject 3, these are all greater than 0.950, suggesting

[5] Less then 2% of the data points were omitted from these analyses because they were judged to be too close to the origin. We reasoned that these points would unfairly bias the intercept toward zero.

strong evidence of linearity. Note also that all $y$-intercepts are within 0.0001 of the predicted value of zero. Despite their small size, however, in about half the cases these deviations are statistically significant.

Although it may not be evident in Fig. 10, in some cases there is a tendency for the data to be separated into two distinct regions of the figure, one above and the other below the diagonal [e.g., see the $f_4(t)$ versus $F_3(t) - F_4(t)$ plot for Subject 4]. This separation reflects a violation of the model's assumptions. Also worth noting is that the size of this effect appears to increase with the magnitude of the violations in the model intersection test. If the pure insertion and exponential assumptions are correct, the estimates of $1/V_k$ computed from the regression slopes should equal those computed from the mean RT differences (i.e., see Eq. (13)). An examination of Table 2 indicates that, in general, the agreement is good. Except for Subject 3, the largest deviation is 13 ms. Note, however, that the deviations are consistently in

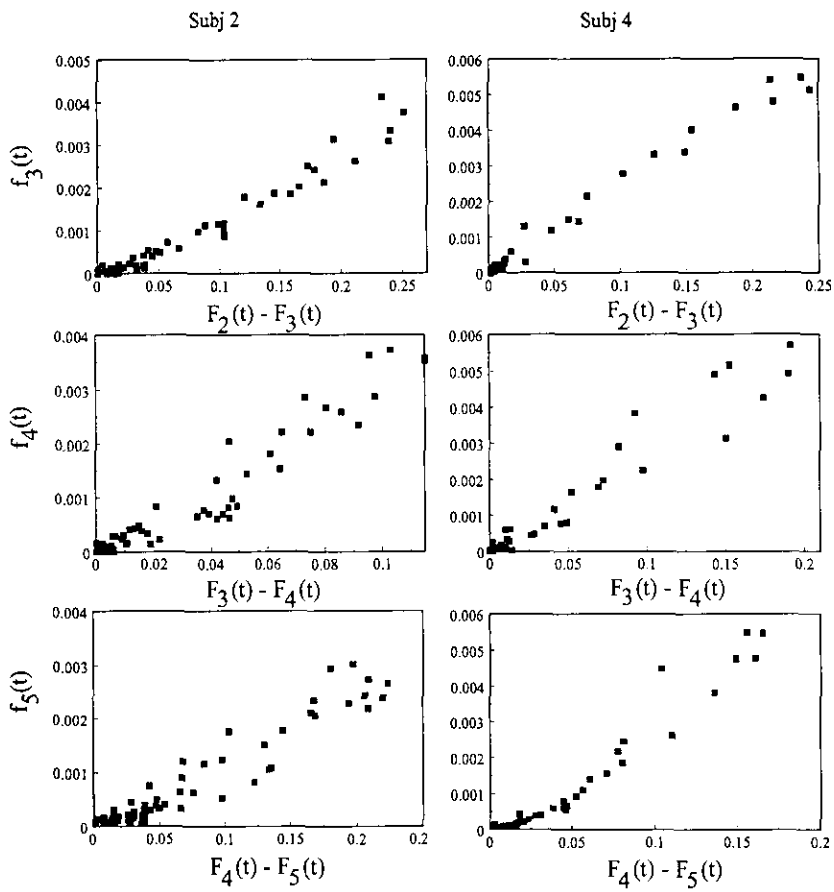


FIG. 10. Plots of $f_k(t)$ versus $[F_{k-1}(t) - F_k(t)]$ for target absent conditions for Subjects 2 and 4.

TABLE 2

Linear Regression Analysis of Eq. (12)

| | $(\text{Slope})^{-1}$ | $\overline{RT}_k - \overline{RT}_{k-1}$ | Intercept | $p_{y\text{-int}}$ | $R$ |
|---|---|---|---|---|---|
| Subject 1 | | | | | |
| 2–3 | 48 | 52 | −0.0001 | 0.023 | 0.983 |
| 3–4 | 59 | 65 | −0.00007 | 0.084 | 0.988 |
| 4–5 | 55 | 68 | −0.00016 | 0.025 | 0.952 |
| Subject 2 | | | | | |
| 2–3 | 72 | 81 | −0.00008 | 0.004 | 0.973 |
| 3–4 | 33 | 40 | −0.00006 | 0.035 | 0.965 |
| 4–5 | 82 | 88 | −0.00006 | 0.023 | 0.951 |
| Subject 3 | | | | | |
| 2–3 | 113 | 133 | −0.00006 | 0.100 | 0.913 |
| 3–4 | 57 | 32 | 0.00019 | 0.001 | 0.765 |
| 4–5 | 70 | 67 | 0.00001 | 0.802 | 0.858 |
| Subject 4 | | | | | |
| 2–3 | 42 | 38 | 0.00001 | 0.579 | 0.980 |
| 3–4 | 35 | 43 | −0.00004 | 0.240 | 0.960 |
| 4–5 | 33 | 37 | −0.00012 | 0.001 | 0.967 |

one direction. Specifically, the slope estimates are consistently smaller than the mean RT differences.

While these results provide reasonable support for the pure insertion and exponential assumptions, they also suggest the possibility that the assumptions do not hold exactly. One possibility is that the assumptions do hold but that in addition, adding an item to the memory set causes some irreducible minimum decision time to be inserted into the processing chain. This model can be expressed as

$$\mathbf{RT}_k = \mathbf{RT}_{k-1} + t_0 + \mathbf{T}_k,$$

where $t_0$ is a constant and $\mathbf{T}_k$ is exponentially distributed. As a test of this hypothesis, we estimated $t_0$ by the difference between the $1/V_k$ estimates obtained from the Table 2 slopes and from the mean RT differences. We then subtracted this value from all observed values of $RT_k$ and retested the Eq. (12) prediction. In virtually all cases, $R$ increased, the bowing effect decreased, and the correspondence increased between the $1/V_k$ estimates obtained from the regression slopes and from the mean RT differences (except for Subject 3). The only negative result that did not disappear was the significant $y$-intercepts.

In summary, although the results do not fully support the pure insertion and exponential assumptions, the results do suggest that this model, or perhaps a model that is able to closely mimic these assumptions, such as the diffusion model (Ratcliff, 1988), provides a good first approximation to the data.

## CONCLUSIONS

The goal of this article was to develop and characterize an empirical data base against which current and future theories of memory scanning can be tested. Toward this end, results were reported from a standard varied-set memory scanning task in which each subject participated in about 1500 trials per memory set size. This large sample size made it possible to examine a number of nonparametric distributional properties of the RT data. On the basis of these analyses, a number of important conclusions stand out.

First, increasing the size of the memory set induces the strongest possible form of stochastic dominance on both target present and target absent trials. This result rules out a number of possible processing models, including parallel, self-terminating models with unlimited capacity.

Second, evidence was found that for the smaller memory set sizes, the RT hazard functions increased to a peak and then decreased to some asymptotic value. For larger memory set sizes, the hazard functions appeared to increase monotonically to an asymptote (although the possibility of a small decrease cannot be ruled out). Because most serial models predict nondecreasing hazard functions for all memory set sizes, this result falsifies a large class of serial search models.

Third, the evidence strongly disconfirmed exhaustive search. Exhaustive models would find it extremely difficult, if not impossible, to predict the combination of shallower target present RT mean and variance curves and pronounced serial position effects that we observed.

Fourth, some evidence was found in support of the assumption that adding an item to the memory set inserts a discrete stage with exponentially distributed duration into the processing chain, at least on target absent trials. Apparently, however, this assumption is not exactly correct. One possibility that appears to account for the major discrepancies is that adding an item to the memory set also adds an irreducible minimum delay, and thus instead of an exponential distribution, the duration of the inserted stage has the distribution of $t_0 + \mathbf{T}_k$, where $t_0$ is a constant an $\mathbf{T}_k$ is exponentially distributed.

Finally, we found some evidence that three of the subjects stored the representations of the memory set items in a visual short-term memory system and the fourth subject used an acoustic short-term system.

To our knowledge, the only extant model able to predict all qualitative results reported in this article assumes that search is parallel, self-terminating, and of very limited capacity. Because Ratcliff's (1978) diffusion model is parallel and self-terminating and also makes predictions about response accuracy, these results make it an excellent candidate for more detailed study.

## REFERENCES

ASHBY, F. G. (1982). Testing the assumption of exponential, additive reaction time models. *Memory and Cognition*, **10**, 125–134.

ASHBY, F. G., & TOWNSEND, J. T. (1980). Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology*, **21**, 93–123.

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

BADDELEY, A. D., & ECOB, J. R. (1973). Reaction time and short-term memory: Implications of repetition effects for the high speed exhaustive scan hypothesis. *Quarterly Journal of Experimental Psychology*, **25**, 229–240.

BALAKRISHNAN, J. D., & ASHBY, F. G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research*, **54**, 80–90.

BARLOW, H. B., MARSHALL, A. W., & PROSCHAN, F. (1966). Properties of probability distributions with monotone hazard rate. *Annals of Mathematical Statistics*, **37**, 1574–1592.

BARLOW, R. E., & PROSCHAN, F. (1965). *Mathematical theory of reliability*. New York: Wiley.

BLOXOM, B. A. (1979). Estimating an unobserved component of a serial response-time model. *Psychometrika*, **44**, 473–484.

BLOXOM, B. A. (1984). Estimating response time hazard functions: An exposition and extension. *Journal of Mathematical Psychology*, **28**, 401–420.

BLOXOM, B. A. (1985). A constrained spline estimator of a hazard fuction. *Psychometrika*, **50**, 301–321.

BRIGGS, G. E., & BLAHA, J. (1969). Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology*, **79**, 395–402.

BRIGGS, G. E., & JOHNSEN, A. M. (1973). On the nature of central processing in choice reactions. *Memory & Cognition*, **1**, 91–100.

BURBECK, S. L., & LUCE, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics*, **32**, 117–133.

BURROWS, D., & OKADA, R. (1971). Serial position effects in high-speed memory search. *Perception & Psychophysics*, **10**, 305–308.

CAVANAGH, J. P. (1976). Holographic and trace-strength models of rehearsal effects in the item recognition task. *Memory and Cognition*, **4**, 186–199.

CHRISTIE, L. S., & LUCE, R. D. (1956). Decision structures and time relations in simple choice behavior. *Bulletin of Mathematical Biophysics*, **18**, 89–112.

CLIFTON, C., & BIRENBAUM, S. (1970). Effects of serial position and delay of probe in a memory scan task. *Journal of Experimental Psychology*, **86**, 69–76.

DONDERS, F. C. (1969). Over de snelheid van psychische processen. (On the speed of mental processes.) *Onderzoekingen gedaan in het physiologisch Laboratorium der Ugtrechtsche Hoogeschool*, 1868–1869, Tweede reeks, **11**, 92–130. Translated by W. F. Koster, in W. G. Koster (Ed.), Attention and performance II, *Acta Psychologica*, **30**, 412–431.

FORRIN, B., & CUNNINGHAM, K. (1973). Recognition time and serial position of probed item in short-term memory. *Journal of Experimental Psychology*, **99**, 272–279.

GALAMBOS, J. (1978). *The asymptotic theory of extreme order statistics*. New York: Wiley.

GREEN, D. M, & LUCE, R. D. (1971). Detection of auditory signals presented at random times: III. *Perception & Psychophysics*, **9**, 257–268.

KOHFELD, D. L., SANTEE, J. L., & WALLACE, N. D. (1981). Loudness and reaction time: I. *Perception & Psychophysics*, **29**, 550–562.

LAMING, D. (1973). *Mathematical psychology*. New York: Academic Press.

LUCE, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford Univ. Press.

MCGILL, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley.

MILLER, D. R., & SINGPURWALLA, N. D. (1977). Failure rate estimation using random smoothing. National Technical Information Service, No. AD-A040999/5ST.

MILLER, J. (1988). Discrete and continuous models of human information processing: Theoretical distinctions and empirical results. *Acta Psychologica*, **67**, 191–257.

MONSELL, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, **10**, 465–501

MURDOCK, B. B., JR. (1985). An analysis of the strength–latency relationship. *Memory & Cognition*, **13**, 511–521.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **33**, 1065–1076.

PETERSON, W. W., BIRDSALL, T. G., & FOX, W. C. (1954). The theory of signal detectability. *Trans. IRE Professional Group on Information Theory*, PGIT-4, 171–212.

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

RATCLIFF, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446–461.

RATCLIFF, R. (1988). A note on mimicking additive reaction time models. *Journal of Mathematical Psychology*, **32**, 192–204.

ROSS, S. M. (1983). *Stochastic processes*. New York: Wiley.

ROSSMEISSL, P. G., THEIOS, J., & KRUNNFUSZ, D. (1979). *Parallel processing models for joint visual and memory scanning*. Paper presented at the Twelfth Annual Mathematical Psychology Meetings, Brown University.

ROYDEN, H. L. (1968). *Real analysis* (2nd ed.). New York: Macmillan.

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

SCHWEICKERT, R. (1985). Separable effects of factors on speed and accuracy: Memory scanning, lexical decision, and choice tasks. *Psychological Bulletin*, **97**, 530–546.

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

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

STERNBERG, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance* (Vol. 2). Amsterdam: North-Holland.

STERNBERG, S. (1973). *Evidence against self-terminating memory search from properties of RT distributions*. Paper presented at the Meeting of the Psychonomic Society, St. Louis.

STERNBERG, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, **17**, 1–32.

TOWNSEND J. T. (1976). A stochastic theory of matching processes. *Journal of Mathematical Psychology*, **14**, 1–52.

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

TOWNSEND, J. T., & ASHBY, F. G. (1978). Methods of modeling capacity in simple processing systems. In N. J. Castellan, Jr., and F. Restle (Eds.), *Cognitive theory*, **3** (pp. 199–239). Hillsdale, NJ: Erlbaum.

TOWNSEND, J. T., & ASHBY, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge Univ. Press.

TOWNSEND, J. T., & ROOS, R. N. (1973). Search reaction times for single targets in multiletter stimuli with brief visual displays. *Memory and Cognition*, **1**, 319–332.

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

TOWNSEND, J. T., & VAN ZANDT, T. (1990). New theoretical results on testing self-terminating vs. exhaustive processing in rapid search experiments. In H. G. Geissler (Ed.), *Psychological explorations of mental structures* (pp. 469–489). Toronto: Hogrefe & Huber.

VORBERG, D., COLONIUS, H., & SCHMIDT, R. (1989). *Distribution inequalities for parallel models with unlimited capacity*. Perdue Mathematical Psychology Program Technical Report No. 89-5.

WALSH, J. E. (1965). *Handbook of nonparametric statistics, II*. New York: Van Nostrand.

WELFORD, A. T. (Ed.) (1980). *Reaction times*. London: Academic Press.

WICKELGREN, W. A., & NORMAN, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, **3**, 316–347.

RECEIVED: March 5, 1990