JOURNAL OF MATHEMATICAL PSYCHOLOGY **38**, 35–58 (1994)

Distribution Inequalities for Parallel Models with Unlimited Capacity

HANS COLONIUS

Universität Oldenburg

AND

DIRK VORBERG

Technische Universität Braunschweig

Inequalities on reaction time distribution functions for n-channel parallel models with an unlimited capacity assumption are presented, extending previous work on serial models. First, the first-terminating and the exhaustive cases are considered without any assumption on the distribution functions. Later, certain general stochastic dependence properties are introduced, and appropriate distribution inequalities are derived and illustrated by specific parametric examples. Moreover, the inequalities are extended to finite mixture distributions and to censoring.

1. INTRODUCTION

It is generally assumed that the processing of information in a reaction time task can be decomposed into several subprocesses. A number of authors have shown that, in general, it is not possible to infer from reaction time measurements whether these subprocesses operate in parallel or serially (Townsend, 1972, 1974; Vorberg, 1977; Townsend & Ashby, 1983; Vorberg & Ulrich, 1987; for a recent review paper, see Townsend, 1990). However, if additional assumptions are introduced constraining the class of possible models, testable conditions on their reaction time distributions can eventually be derived. In particular, Sternberg (1973) derived and empirically tested distributional properties of self-terminating search models (for a generalization, see Vorberg, 1981). Moreover, Townsend & Ashby (1983, Chap. 8) derived a host of results on both parallel and serial self-terminating and exhaustive systems under various conditions of independence and capacity. The aim of this paper is to extend this approach to models of parallel processes only constrained by an assumption of "limited capacity" made precise later. Systems of reaction time

Reprint requests should be addressed to: Hans Colonius, Institut für Kognitionsforschung, Universität Oldenburg, Fachbereich 5-VG, P.O. Box 2503, D-26111 Oldenburg, Germany. E-mail: colonius@psychologie.uni-oldenburg.de

0022-2496/94 $6.00

distribution function inequalities are derived that must hold for any model within that class.

The basic idea underlying the proposed tests can be illustrated by considering a typical visual search task where the subject is asked to look for one and the same target trial after trial (e.g., Egeth *et al.*, 1972). The target appears in a background of some variable number of distractors, for example, a digit among a set of letters. The subject must indicate whether the target is present or absent. Based on the finding that mean latency and errors show only very small (or no) increases with increases in the number of characters in the presentation frame, a typical interpretation is that of an essentially unlimited capacity, parallel detection process (e.g., Schneider & Shiffrin, 1977). Using only the target absent response times and varying the number of distractors it can be shown that this model predicts specific upper and lower bounds for the response time distribution in the condition with n distractors in terms of the response time distributions from conditions with less than n distractors. Without introducing any distributional assumptions, these inequalities constitute a stronger test of the underlying processing assumptions than those based on the comparison of means alone. Furthermore, results which are assumed to support a parallel limited capacity model (e.g., Fisher, 1982; Rumelhart, 1970) could be tested for possible consistency with a parallel unlimited capacity interpretation. Unfortunately, since typically only mean latencies and errors are reported in the literature, a reanalysis of the original data sets would be necessary.

Another application arises from a related task, the same–different experiment (e.g., Bamber, 1969). On any trial, two letter strings are presented to the subject, e.g.,

b f w s

b f r s.

The subject's task is to decide as fast as possible whether the strings are the same or different. A simple model for this situation would assume channels corresponding to the columns working in parallel and trying to match the letters. A "different" response is initiated as soon as one of the channels has registered a mismatch, whereas the "same" response cannot be initiated unless all channels have signaled a match. No claim is made here that this type of model gives the most plausible account of all empirical results known so far. In fact, a number of different models have been discussed in the literature (see, e.g., Townsend & Ashby, 1983, p. 153). The following remarks are meant to serve for illustrative purposes only, and our method is restricted to the analysis of (i) correct "same" responses, and (ii) correct "different" responses with all columns containing mismatching letters, respectively.

How should reaction time change as the number of letters per string, n, say, is varied? Obviously, correct "same" decisions should take longer the longer the strings to be compared; similarly, correct "different" decisions should become faster with increasing string length. However, it is of interest here that these changes in the reaction time (RT) distributions are limited even without making any specific distributional assumptions. Specifically, for any parallel model with unlimited capacity there exist upper and lower bounds for the distributions depending on the experimental conditions.

Our approach can also be conceived of as generalizing certain tests used in a related research area, the detection of multimodal stimuli. When human observers monitor two or more sources of information for a target, performance is facilitated if the target is present in two or more sources rather than in only one. For example, if in a simple reaction time task with visual and auditory stimuli the subject is instructed to respond to whichever stimulus is perceived first, reaction time to bimodal stimuli is typically faster than reaction time to a unimodal stimulus (Hershenson, 1962). This facilitation of RT effect is reflected not only in the order of the means but also in distribution inequalities of the type suggested in this paper yielding a nonparametric test of the underlying parallel processing assumptions.

As suggested by our initial examples, it is not uncommon to have an empirical situation calling for a model with more than two channels working in parallel. The goal of this paper is to derive inequalities for more than two channels and to demonstrate the usefulness of the ensuing systems of inequalities in testing parallel models of unlimited capacity. The next section defines two basic types of parallel models and specifies the notion of unlimited capacity. Section 3 illustrates these notions by a number of specific examples. Section 4 presents an elementary probability inequality from which distributional inequalities for parallel models are then derived in Section 5. In Section 6, bounds for RT models with more specific dependence assumptions are presented. In Section 7, our approach is extended in two respects. First, reaction times with finite mixture distributions are considered. Second, we allow reaction times to be censored from the right and/or from the left. A final section discusses further extensions and some open problems.

2. Basic Definitions

We describe a stochastic model of parallel system with n processing channels by nonnegative random variables $T_1, T_2, ..., T_n$ which refer to the processing durations of the channels. Each stimulus starts processing as soon as the stimulus is presented. Channels may interact with each other; thus, processing times T_k, $k = 1, ..., n$, are allowed to be stochastically dependent. A parallel system is called *first-terminating* if it initiates a response as soon as the first channel finishes processing. Thus, the contribution of the channels' processing to overall reaction time is the time needed by the fastest channel to finish processing, i.e., the minimum of the T_k. Alternatively, a parallel system is called *exhaustive* if it cannot initiate a response unless all its channels have finished. In this case, the contribution to overall reaction time is given by the time of the slowest channel, i.e., the maximum of the T_k. We assume that overall reaction time comprises components like stimulus encoding time, response selection time, and motor execution time; all these additional components are summarized by a random variable, M.

DEFINITION 1. A reaction time model is called *parallel first-terminating* if

$$\mathrm{RT} = \min_{k=1,\dots,n} T_k + M;$$

it is called *parallel exhaustive* if

$$\mathrm{RT} = \max_{k=1,\dots,n} T_k + M.$$

It should be noted that our definition of a first-terminating process differs from the notion of self-terminating processing familiar from the visual and memory search literature. In this literature, a parallel system is called *self-terminating* if the analysis stops as soon as a prespecified target is identified against a background of, say, n, distractors. Thus, a target-present response can be executed after any number 0, 1, ..., n distractors have been analyzed, not after the first channel terminates processing. Consistency would suggest introducing the term "last-terminating" instead of "exhaustive." However, we decided to keep with the latter term, which common in the experimental literature.

We confine attention to parallel models with unlimited processing capacity. By this we mean that the system allots the same amount of capacity to a given channel no matter how many additional channels operate at the same time. This notion of unlimited capacity can be made precise by considering the joint probability distribution functions of the times T_k over experimental conditions with different sets of active channels. Suppose we have a subset of m out of the n channels of the system. Unlimited capacity stipulates that their (marginal) distributions are the same no matter how many and which of the remaining $n-m$ channels are active.

DEFINITION 2. Let $B = \{i_1, i_2, \dots, i_m\}$ $(m \leqslant n)$ be the set of channels active under a given experimental condition and let P_B denote the corresponding probability measure implied by the m-channel model; an n-channel parallel model is said to have *unlimited capacity* if

$$\begin{aligned} &P_B(T_{i_1} \leqslant t_1, T_{i_2} \leqslant t_2, \dots, T_{i_m} \leqslant t_m) \\ &\quad = P(T_{i_1} \leqslant t_1, T_{i_2} \leqslant t_2, \dots, T_{i_m} \leqslant t_m) \\ &\qquad \text{for all} \quad B \subset \{1, 2, \dots, n\} \text{ and all } t_1, t_2, \dots, t_n \in \Re^+, \end{aligned} \tag{1}$$

where P denotes the probability measure corresponding to the model where all n channels are active.

The reader should note that this definition of unlimited capacity neither implies nor is implied by stochastic independence of the processing times. An analogous definition was proposed by Vorberg (1981) for serial models.

3. Examples

To illustrate these definitions, we consider some first-terminating parallel models with dependent processing times. In each case, we start from a model with independent exponentially distributed durations. By randomizing their rate parameters, positive dependence is generated. In the first example, this is done by assuming rate fluctuation across trials. This examples illustrates the notion of unlimited capacity. In the second example, the system is assumed to adjust its rates by monitoring a deadline. By this example, we further show that unlimited capacity in the sense of Definition 2 is sufficient but not necessary for the inequalities to be presented later to hold.

Example 1 (Conditionally Independent Exponentials). On any trial, the channel processing times are conditionally independent exponential random variables with rate α. The rates are not constant but fluctuate across trials in the following way. We assume that they are determined as $\alpha = A\beta$, where the β reflect channel-specific properties whereas A represents the system's momentary level of alertness which is itsel exponentially distributed with parameter τ. We characterize the processing time distributions by its joint survivor-function, S_N, defined as

$$S_N(t_1, ..., t_n) = P_N(T_1 > t_1, ..., T_n > t_n).$$

Note that the distribution is indexed by $N = \{1, 2, ..., n\}$ to make explicit that it refers to the situation where processing is initiated on all n channels. (For simplicity, we also write $N-1$ for $\{1, 2, ..., n-1\}$). By randomizing A, S_N is obtained as

$$\begin{aligned} S_N(t_1, ..., t_n) &= \int_0^\infty S_N(t_1, ..., t_n \mid A = a)\, \tau \exp(-\tau a)\, da \\ &= \int_0^\infty \prod_{i=1}^n \exp(-a\beta t_i) \tau \exp(-\tau a)\, da \\ &= \int_0^\infty \exp\left(-a\beta \sum_{i=1}^n t_i\right) \tau \exp(-\tau a)\, da \\ &= \frac{\tau}{\tau + \beta \sum_{i=1}^n t_i}. \end{aligned} \tag{2}$$

It is straightforward to derive from (2) the joint survivor function for any subset of channels by using the fact that, e.g.,

$$\begin{aligned} &P_N(T_1 > t_1, ..., T_{i-1} > t_{i-1}, T_{i+1} > t_{i+1}, ..., T_n > t_n) \\ &\quad = S_N(t_1, ..., t_{i-1}, 0, t_{i+1}, ..., t_n). \end{aligned}$$

Therefore, the marginal distribution of the subset of channels 1 through $n-1$ in the n-channel situation equals the distribution obtaining for the corresponding $n-1$-channel situation; i.e.,

$$S_N(t_1, ..., t_{n-1}, 0) = S_{N-1}(t_1, ..., t_{n-1}).$$

By analogous reasoning, the marginal processing time distribution for any subset of channels obtained under n-channel conditions equals that for the corresponding situation with just those channels operative. This proves that the model obeys our definition of unlimited capacity.

EXAMPLE 2 (Deadline Controlled Processing Rates). As before, we assume that the processing durations per channel are exponential with rate α. At processing onset, the system starts a deadline timer that is used for monitoring the channels' rates: When the deadline expires, they may be adjusted instantaneously to some new value, β. We consider the consequences of two alternative adjustment schemes.

Scheme 1. When the deadline expires, the rate of each channel still active is changed from α to β.

Scheme 2. If none of the channels has finished processing when the deadline expires, all rates are changed from α to β; otherwise, they remain at their original value, α.

As far as the minimum completion time is concerned, the two schemes lead to identical predictions. It may come as a surprise, however, to learn that Scheme 2 violates our definition of unlimited capacity, whereas Scheme 1 does not.

For both versions of this model, we assume a deadline distribution that is exponential with rate τ; moreover, the processing durations are assumed to be conditionally independent of each other given the deadline D.

Minimum Distribution

Let g_N denote the density of the minimum processing time; i.e.,

$$g_N(t) = dP_N[\min_{i=1,\ldots,n} T_i \leqslant t]/dt.$$

To derive this density, we randomize the deadline, D. For the minimum to attain value t, either one of two events must happen; (i) the deadline expires before any of the n channels, leading to a rate change from α to β, or (ii) without a rate change, the fastest channel finishes at t whereas the deadline expires after t. Noting that the minimum of n independent exponential random variables with rate μ is exponential with rate $n\mu$, we obtain

$$\begin{aligned} g_N(t) &= \int_0^t [\tau \exp(-\tau v) \exp(-n\alpha v)](n\beta \exp(-n\beta(t-v))\, dv \\ &\quad + \int_0^t \tau \exp(-\tau v)\, n\alpha \exp(-n\alpha t)\, dv \\ &= \frac{\tau n\beta}{\tau + n(\alpha-\beta)} \{\exp(-n\beta t) - \exp[-(\tau+n\alpha)t]\} \\ &\quad + n\alpha \exp[-(\tau+n\alpha)t], \qquad \alpha \neq \beta. \end{aligned} \tag{3}$$

Obviously, the reasoning above holds for both schemes, implying that the time for the first of the n channels to finish follows (3) in both cases.

Marginals

Let f_N denote the marginal distribution of a particular channel, say i, when processing is initiated on n channels, i.e., $f_N(t) = dP_N(T_i \leqslant t)/dt$. Under Scheme 1, the processing rate is changed if $D \leqslant T_i$, otherwise it remains unchanged. Thus,

$$\begin{aligned} f_N(t) &= \int_0^t \tau \exp(-\tau v) \exp(-\alpha v) \exp[-\beta(t-v)]\, dv \\ &\quad + \exp(-\tau t)\, \alpha \exp(-\alpha t) \\ &= \tau\beta/(\tau+\alpha-\beta)\{\exp(-\beta t) - \exp[-(\tau+\alpha)t]\} \\ &\quad + \alpha \exp[-(\tau+\alpha)t], \qquad \alpha \neq \beta. \end{aligned}$$

Note that the marginal does not depend on the total number, n, of channels operative.

This is not so under Scheme 2. Here, no rate change occurs and T_i follows the exponential density $\alpha \exp(-\alpha t)$ unless the deadline is faster than the remaining $n-1$ channels and expires before t. A rate change from α to β occurs and T_i is governed by the density $\exp(-\alpha v)\, \beta \exp(t-v))$ if the deadline expires at $v \leqslant t$, before any of the n channels. Therefore,

$$\begin{aligned} f_N(t) &= \{1 - \tau/(\tau+(n-1)\alpha)[1-\exp(-(\tau+(n-1)\alpha)t]\}\, \alpha \exp(-\alpha t) \\ &\quad + \int_0^t \tau \exp(-\tau v) \exp(-n\alpha v)\, \beta \exp(-\beta(t-v))\, dv \\ &= \{1 - \tau/(\tau+(n-1)\alpha)[1-\exp(-(\tau+(n-1)\alpha)t]\}\, \alpha \exp(-\alpha t) \\ &\quad + \tau\beta/(\tau+n\alpha-\beta)\{\exp(-\beta t) - \exp[-(\tau+n\alpha)t]\}. \end{aligned}$$

Under Scheme 2, the marginal does depend on the total number of channels initially activated, and thus violates the unlimited capacity condition: If $\alpha \leqslant \beta$, i.e., if the systems speeds up processing whenever the deadline is over before any of the channels has finished, (3) implies that the mean processing time of a channel increases with the total number of channels, n.

At first sight, this seems counterintuitive since the deadline parameter as well as the rates before or after the change are invariant with n. This shows that parameter invariance cannot necessarily be equated with unlimited capacity in the sense of Definition 2. Intuition may be helped by noting that the model's behavior under Scheme 2 is due to a subtle effect of "statistical facilitation:" The larger the number of channels to compete with, the smaller the probability that the deadline will win the race and, consequently, lead to a processing rate increase. Thus, the more active channels there are, the less likely a rate change will be effected from which a particular channel might profit.

The comparison of the two deadline model versions shows that the proposed unlimited capacity condition is sufficient but not necessary for the validity of the distribution inequalities: Under Scheme 1, unlimited capacity holds; therefore, the minimum processing time distributions generated by this model clearly obey the inequalities. Under Scheme 2, unlimited capacity is violated; the same minimum distributions are generated as under Scheme 1, however, showing that it is possible for a parallel model to obey the distribution inequalities without conforming to the unlimited capacity condition.

4. An Elementary Probability Inequality

Let $B_1, B_2, ..., B_n$ $(n > 2)$ be events in some given probability space. In this section, an elementary probability inequality on unions of these events is presented. It is basic for the derivation of the distribution inequality tests for parallel models in the next section. Let $P(\bigcup_{k \neq i} B_k)$ denote $P(\bigcup_{k=1, k \neq i}^{n} B_k)$ and so on. Then the lemma is

Lemma 1.

$$P\left(\bigcup_{k \neq i} B_k\right) \leqslant P\left(\bigcup_{k} B_k\right) \leqslant P\left(\bigcup_{k \neq i} B_k\right) + P\left(\bigcup_{k \neq j} B_k\right) - P\left(\bigcup_{k \neq i, j} B_k\right)$$

for any $i, j \leqslant n$, $i \neq j$. (4)

Proof. Let A, B, C be arbitrary events in the probability space: then

$$\begin{aligned}
&P(A \cup B \cup C) \\
&\quad = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \\
&\quad = [P(A) + P(B) - P(AB)] + [P(A) + P(C) - P(AC)] \\
&\qquad - P(A) - [P(BC) - P(ABC)] \\
&\quad \leqslant P(A \cup B) + P(A \cup C) - P(A),
\end{aligned}$$

where the inequality follows from $P(BC) - P(ABC)$ being nonnegative by the monotonicity of any probability measure. The upper bound in the lemma follows from putting $A = \bigcup_{k \neq i, j} B_k$, $B = B_i$, and $C = B_j$. The lower bound is trivial. Q.E.D.

Inequality (4) gives an upper and a lower bound for the probability that at least one of the events $B_1, ..., B_n$ occurs. Further bounds can be derived by considering unions of events with more than two events being excluded from the union at the same time. This is discussed in the final section.

5. Distribution Inequality Tests

In a parallel first-terminating system with n channels a response is initiated as soon as any of the channels finishes processing. Intuitively, adding another channel to the race cannot but increase—or, at least keep constant—the probability that the first channel finishes before some time t if unlimited capacity is assumed to hold because it is the additionally channel which may be the fastest. This intuition is corroborated by the first inequality (lower bound) in the lemma to follow. On the other hand, the probability of finishing processing with n channels before time t is bounded from above by a linear combination of the corresponding probabilities with $n-1$ and $n-2$ channels operating. This bound is specified by the right hand side of the first inequality in the lemma to follow. Analogously, adding another channel in a parallel exhaustive system cannot but decrease the probability that processing will be finished before some time t, since this added channel may take longer to finish than any of the others. Again, however, this probability is bounded from below as specified in the second inequality in the lemma.

To keep notation simple, distribution functions for first-terminating and exhaustive parallel models with n channels are denoted by F_n and G_n, respectively; thus,

$$F_n(t) = P\,(\mathrm{RT} \leqslant t) \qquad \text{(first-terminating } n\text{-channel process)}$$

and

$$G_n(t) = P\,(\mathrm{RT} \leqslant t) \qquad \text{(exhaustive } n\text{-channel process).}$$

Similarly, we write $F^{(i)}_{n-1}(t)$, $F^{(ij)}_{n-2}(t)$, etc., for the distribution function if all channels but channel i, all channels but channels i, j, etc., are active, respectively. Thus, using the notation of Definition 2 with $C = \{1, 2, ..., n\}$ and neglecting the random component M we have, for example,

$$G^{(i)}_{n-1}(t) = P_{C\setminus\{i\}}(T_1 \leqslant t, T_2 \leqslant t, ..., T_{i-1} \leqslant t, T_{i+1} \leqslant t, ..., \; T_n \leqslant t).$$

By the unlimited capacity assumption, this equals

$$P(T_1 \leqslant t, T_2 \leqslant t, ..., T_{i-1} \leqslant t, T_{i+1} \leqslant t, ..., T_n \leqslant t).$$

Lemma 2. (a) *For an n-channel parallel model ($n > 2$) with unlimited capacity with $1 \leqslant i, j \leqslant n$, $i \neq j$, the following holds for all real t ($t > 0$):*

$$F^{(i)}_{n-1}(t) \leqslant F_n(t) \leqslant F^{(i)}_{n-1}(t) + F^{(j)}_{n-1}(t) - F^{(ij)}_{n-2}(t) \tag{5}$$

(first-terminating process) and

$$G^{(i)}_{n-1}(t) + G^{(j)}_{n-1}(t) - G^{(ij)}_{n-2}(t) \leqslant G_n(t) \leqslant G^{(i)}_{n-1}(t) \tag{6}$$

(last-terminating process).

(b) *For a two-channel parallel model* ($n=2$) *with unlimited capacity the above inequalities simplify to*:

$$F_1^{(i)}(t) \leqslant F_2(t) \leqslant F_1^{(i)}(t) + F_1^{(j)}(t) \qquad i, j = 1, 2, i \neq j \tag{7}$$

and

$$G_1^{(i)}(t) + G_1^{(j)}(t) - 1 \leqslant G_2(t) \leqslant G_1^{(i)}(t) \qquad i, j = 1, 2, i \neq j. \tag{8}$$

According to Definition 1, reaction time (RT) is assumed to be an additive composition of processing time, $\min(T_k)$ or $\max(T_k)$ for first-terminating or exhaustive models, respectively, and another random component M, encompassing motor execution time, etc. In proving Lemma 2 we need not make explicit component M since obviously

$$\text{RT} = \min_k T_k + M = \min_k (T_k + M) = \min T_k^*$$

and

$$\text{RT} = \max_k T_k + M = \max_k (T_k + M) = \max T_k^*,$$

for the first-terminating and exhaustive case, respectively. Since no distributional assumptions about the T_k are being made, we may as well go over to the T_k^*. It is clear, however, that component M does induce a certain amount of positive dependence among the T_k^* the impact of which depends on the size of the variance of M relative to the variances of the T_k (see Section 6 for a further discussion of dependence).

Before presenting a simple proof of the above lemma, let us note two consequences. The upper (lower) bounds on F_n and G_n can be sharpened by minimizing (maximizing) over the indices i, j.

THEOREM 1. *Under Lemma* 2, *we have for all* t

$$\max_i F_{n-1}^{(i)}(t) \leqslant F_n(t) \leqslant \min_{i,j} \left[F_{n-1}^{(i)}(t) + F_{n-1}^{(j)}(t) - F_{n-2}^{(ij)}(t)\right]$$

and

$$\max_{i,j} \left[G_{n-1}^{(i)}(t) + G_{n-1}^{(j)}(t) - G_{n-2}^{(ij)}(t)\right] \leqslant G_n(t) \leqslant \min_i G_{n-1}^{(i)}(t). \tag{9}$$

If all marginal distributions $F_{n-1}^{(i)}$, $F_{n-2}^{(ij)}$, etc. are independent of the choice of i and j, the inequalities take a specially simple form.

COROLLARY 1. *Under Theorem* 1, *if the marginal distributions* $F^{(i)}_{n-1}$, $F^{(ij)}_{n-2}$, $G^{(i)}_{n-1}$, *and* $G^{(ij)}_{n-2}$ *are the same for any choice of* $i, j \in \{1, ..., n\}$, *then*

$$F_{n-1}(t) \leqslant F_n(t) \leqslant 2F_{n-1}(t) - F_{n-2}(t) \tag{10}$$

and

$$2G_{n-1}(t) - G_{n-2}(t) \leqslant G_n(t) \leqslant G_{n-1}(t). \tag{11}$$

Proof of Lemma 2. For Inequality (5) (the first-terminating case), we put, for any t,

$$B_k = \{T_k \leqslant t\};$$

thus,

$$F_n(t) = P\left(\bigcup_k B_k\right), \qquad F^{(i)}_{n-1}(t) = P\left(\bigcup_{k \neq i} B_k\right),$$

etc. Hence (5) is an immediate consequence of Lemma 1. For Inequality (6) (the exhaustive case), we put

$$B_k = \{T_k > t\}$$

for any t; thus, again by Lemma 1

$$\begin{aligned} G_n(t) &= 1 - P(\max(T_k) > t) \\ &= 1 - P\left(\bigcup_k B_k\right) \\ &\leqslant 1 - P\left(\bigcup_{k \neq i} B_k\right) \\ &= G^{(i)}_{n-1}(t), \end{aligned}$$

settling the upper bound. The lower bound and the binary cases can be shown analogously. Q.E.D.

A simple consequence of the above theorem in the $n = 2$ case is the following inequality predicted by any parallel model with unlimited capacity in the bimodal detection paradigm. Let s_V and s_A denote the event of a visual or auditory stimulus being presented, respectively. Then, as Miller (1982) showed (compare Lemma 2(b)),

$$\begin{aligned} \max[P(T \leqslant t \mid s_V), P(T \leqslant t \mid s_A)] &\leqslant P(T \leqslant t \mid s_V \,\&\, s_A) \\ &\leqslant P(T \leqslant t \mid s_V) + P(T \leqslant t \mid s_A) \end{aligned} \tag{12}$$

for any nonnegative t, with T denoting the response time random variable. Thus, according to (12) the bimodal RT distribution, while not being smaller than either of the unimodal ones, is bounded above by the sum of the unimodal RT distributions. The inequality has been taken to test the *statistical facilitation hypothesis* (Raab, 1962) as an explanation for the speed-up of responses in the bimodal condition. Although it is not very restrictive, it tends to be violated as demonstrated in a number of studies (e.g., Miller, 1982, 1986; Diederich & Colonius, 1987; Diederich, 1992a, 1992b) indicating that the speed-up of RT in the bimodal condition is larger than predicted by the statistical fact that the mean of the minimum of two random variables is at most as large as the smaller of the individual means. The above inequality has also been tested in the context of the *redundant targets effect* in letter identification (Grice *et al.*, 1984; van der Heijden *et al.*, 1984) where violations have been less consistent. Furthermore, it has been the subject of several theoretical and simulations studies (Ulrich & Giray, 1986; Colonius, 1986, 1988, 1990; Ashby & Townsend, 1986; Eriksen, 1988). A probability inequality corresponding to (12) has also been used in artificial intelligence to describe the propagation of probability bounds through opinion nets (cf. Winston, 1992, p. 234).

Turning back to the general ($n \geqslant 2$) case, in an extensive empirical study applying nonparametric properties of response time distributions to memory scanning data, Ashby *et al.* (in press) tested the inequality for the exhaustive case in Theorem 1 (renaming it the *medium RT property*). As a result, they could rule out a large class

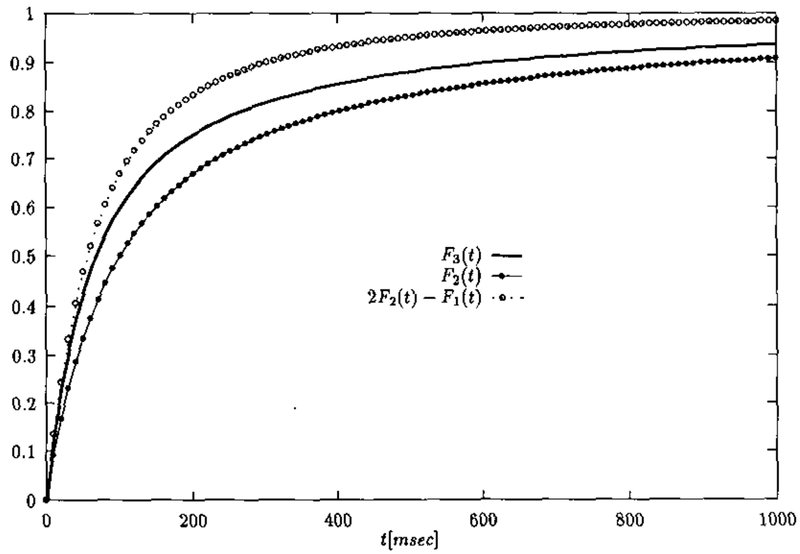

FIG. 1. The first-terminating process of Example 1 (conditionally independent exponentials with fluactuating rate parameter; cf. Eq. (2)) illustrating distribution inequality (10) with $n=3$ parallel channels. The distribution function $F_3(t)$ (dashed line) is bounded from below by $F_2(t)$ and from above by $2F_2(t)-F_1(t)$. Parameter values are $\tau=1000$, $\beta=5$.

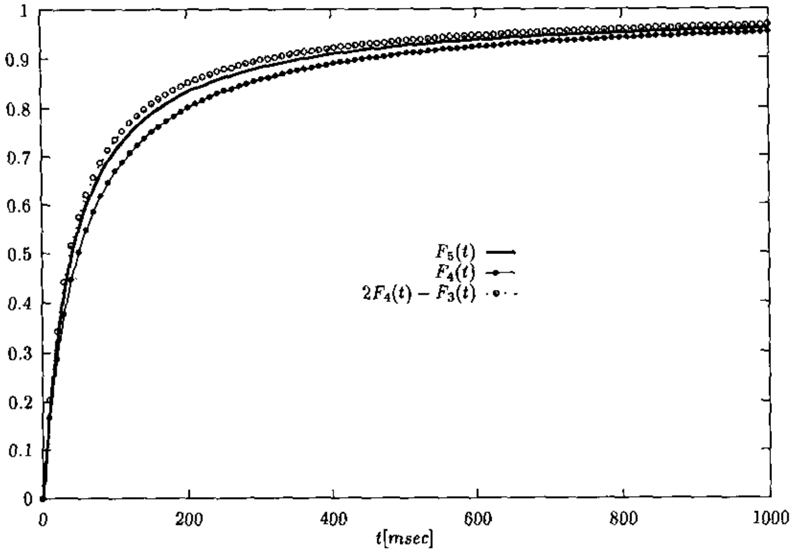

FIG. 2. Same as Fig. 1 with $n = 5$ parallel channels. It can be seen that increasing the number of channels leads to stricter bounds.

of unlimited capacity, parallel exhaustive scanning models. Here, we graphically illustrate the above inequalities using the multivariate distribution from Example 1 in the last section. Figure 1 shows inequality (10) for a three-channel system of exponentials with fluctuating rates. When the number of channels is increased to five, the inequalities become sharper, as Fig. 2 demonstrates (for details on the parameter values, consult the figures).

6. TESTS FOR POSSIBLY DEPENDENT PARALLEL MODELS

The inequalites in Theorem 1 testing first-terminating and exhaustive models do not require any assumptions about stochastic dependence or independence among the processing channels, thus allowing tests of the entire class of unlimited capacity models. In this section, we concentrate on the model with an explicit residual component M, that is,

$$\mathrm{RT} = \min_k T_k + M.$$

It is shown that adding certain assumptions of dependence among the random variables may yield inequalities sharper than those of Theorem 1. As a starting point, let us assume $T_1, ..., T_n$ and M are jointly independent. As pointed out above, a non-constant M component generates some dependence among the $T_k + M$ variables.

PROPOSITION 1. *For an n-channel parallel model* ($n \geqslant 2$) *with unlimited capacity assume* $T_1, ..., T_n$, M *to be jointly independent; the following holds for all real* t ($t \geqslant 0$).

$$\begin{aligned} F_n(t) &= P[\min(T_k) + M \leqslant t] \\ &\leqslant 1 - \prod_k [1 - P(T_k + M \leqslant t)] \\ &= 1 - \prod_k [1 - F_{1,k}(t)] \end{aligned} \tag{13}$$

(*first-terminating case*);

$$\begin{aligned} G_n(t) &= P[\max(T_k) + M \leqslant t] \\ &\leqslant \prod_k P(T_k + M \leqslant t) \\ &= \prod_k G_{1,k}(t) \end{aligned} \tag{14}$$

(*exhaustive case*), *where* $F_{1,k}$, $G_{1,k}$ *denote the respective univariate marginals.*

We do not give a proof of Proposition 1 here since it follows as a special case of a theorem below showing that assumptions weaker than independence do suffice. It should be pointed out, moreover, that the bounds (13), (14) (as well as those that follow in this section) need not be sharper uniformly in t than those of Theorem 1. Whether or not they are sharper, and for which values of t, depends on the specific marginals. Finally, if the residual component M is set equal to zero, the inequalities in the above proposition become identities.

Once the assumption of stochastic independence has been dropped, the question is what notion of dependence it should be replaced by. On the one hand, that notion should be as weak as possible in order to be applicable to a sufficiently large range of data; on the other hand, it needs to be strong enough to allow the derivation of upper and/or lower bounds on the distribution functions.

A simple concept of positive dependence was introduced by Lehmann (1966): a pair of random variables X, Y is called *positive quadrant dependent* (PQD) if for all x, y

$$P(X \leqslant x, Y \leqslant y) \geqslant P(X \leqslant x)\, P(Y \leqslant y) \tag{15}$$

or, equivalently,

$$P(X > x, Y > y) \geqslant P(X > x)\, P(Y > y). \tag{16}$$

The following definition generalizes this concept to the $n > 2$ case (cf. Block & Sampson, 1983).

DEFINITION 3. A real-valued random vector $\mathbf{X} = (X_1, ..., X_p)$ is called *positive upper orthant dependent* (PUOD) if for all $x_1, ..., x_p$

$$P(X_1 > x_1, ..., X_p > x_p) \geqslant \prod_i P(X_i > x_i); \tag{17}$$

$\mathbf{X}$ is called *positive lower orthant dependent* (PLOD) if for all $x_1, ..., x_p$

$$P(X_1 \leqslant x_1, ..., X_p \leqslant x_p) \geqslant \prod_i P(X_i \leqslant x_i). \tag{18}$$

It has been shown (e.g., Block & Sampson, 1983) that PUOD is equivalent to PLOD only if $p = 2$. The proposition below is a simple consequence of these various dependence properties.

PROPOSITION 2. *For an n-channel parallel model* ($n \geqslant 2$) *with unlimited capacity, with* $\mathbf{T}^* = (T_1^*, ..., T_n^*)$ $(k = 1, ..., n)$, $F_n(t) = P[\min(T_k^*) \leqslant t]$, *and* $G_n(t) = P[\max(T_k^*) \leqslant t]$, *the following holds* (*for all t, respectively*):

$$F_n(t) \leqslant 1 - \prod_k [1 - P(T_k^* \leqslant t)] \tag{19}$$

if $\mathbf{T}^*$ *is PUOD;*

$$G_n(t) \geqslant \prod_k P(T_k^* \leqslant t) \tag{20}$$

if $\mathbf{T}^*$ *is PLOD.*

EXAMPLE 1 (Continued: Dependence Generated by Rate Fluctuations). The derivation of the joint survivor distribution presented earlier also implies that the marginal distribution for any individual channel follows

$$P_N(T_i > t_i) = \tau/(\tau + \beta t_i).$$

It is not hard to show that

$$\frac{\tau}{\tau + \beta \sum_{i=1}^n t_i} \geqslant \prod_{i=1}^n \frac{\tau}{\tau + \beta t_i}.$$

Therefore,

$$P_N\left(\bigcap_{i=1}^n T_i > t_i\right) \geqslant \prod_{i=1}^n P_N(T_i > t_i),$$

all $t_i \geqslant 0$.

This proves that the model exhibits positive upper orthant dependence (PUOD). The positive dependence among the processing durations is due to the fluctuation of the system's rate A affecting all the channels in the same way.

While the bounds in Proposition 2 are derived under relatively weak dependence assumptions, the result is somewhat unsatisfactory because the dependence assumptions are formulated in terms of the T_k^* rather than the T_k. An obvious suggestion would be to impose the dependence properties of Definition 3 on $\mathbf{T} = (T_1, ..., T_n)$ and to derive appropriate bounds by assuming M to be independent of $\mathbf{T}$. Unfortunately, this seems to be feasible only for the binary ($n = 2$) case (see, e.g., Tong, 1980, Theorem 5.1.4.).

For the general case, a concept of positive dependence stronger than PUOD and PLOD has to be imposed on $\mathbf{T}$. The following concept of *association* was introduced by Esary *et al.* (1967) to obtain bounds related to functions in the theory of reliability and has found much interest in the statistical literature (see, e.g., Tong, 1980, Chap. 5; for an application in item response theory, the reader is referred to Rosenbaum, 1984, and Holland & Rosenbaum, 1986; in random utility theory, to Colonius, 1983; association has also been used in a nearest neighbor analysis of a multidimensional scaling problem, see Tversky *et al.* 1983; Newman *et al.*, 1983).

DEFINITION 4. A real-valued random vector $\mathbf{X} = (X_1, ..., X_p)$ (or, equivalently, the set of random variables $X_1, ..., X_p$, or its distribution) is said to be *associated* if for every pair of increasing[1] real functions f, g defined on $\mathfrak{R}^p$,

$$\operatorname{Cov}[f(\mathbf{X}), g(\mathbf{X})] \geqslant 0. \tag{21}$$

Association has the following useful properties which can be verified from the definition. Suppose $X_1, ..., X_p$ are associated; then

(P1) $\operatorname{Cov}(X_i, X_j) \geqslant 0$, $i, j = 1, ..., p$.

(P2) Any subset of $X_1, ..., X_p$ is associated.

(P3) The union of two independent associated sets is associated.

(P4) A set consisting of a single random variable is associated.

(P5) Increasing (decreasing) functions of associated random variables are associated.

It follows immediately from P3 and P4 that independent random variables are associated. Conversely, it has been shown (see, e.g., Joag-Dev, 1983; Newman, 1984) that associated random variables that are pairwise uncorrelated are jointly independent. Thus, for associated random variables, the independence structure is largely determined by the covariance structure. While the defining condition for

[1] Throughout this paper, we use "increasing" in place of "nondecreasing" and "decreasing" in place of "nonincreasing."

association (Eq. (13)) may not be checked easily in general, associated random variables can be constructed from independent ones via Properties P3, P4, and P5 (see below). Moreover, large classes of multivariate distributions have been shown to possess the associatedness property.

The following well-known lemma is needed for our next theorem. Its proof (see Appendix) is included here for illustrative purposes.

LEMMA 3. *Association implies PUOD and PLOD.*

The next theorem presents conditions on the dependence structure weaker than those of Proposition 1 while retaining the same upper and lower bounds for the first-terminating and the exhaustive cases, respectively.

THEOREM 2. *For an n-channel parallel model* ($n \geqslant 2$) *with unlimited capacity assume* (i) $T_1, ..., T_n$ *to be associated, and* (ii) M *to be stochastically independent of* $T_1, ..., T_n$; *then the bounds of Proposition* 1 *hold for the first-terminating and exhaustive case, respectively.*

Proof. By P4 and P3, the random variables $T_1, ..., T_n$, M are associated; by P5, then $T_1 + M, ..., T_n + M$ are associated, too. The bounds then are immediate consequences of Lemma 2 and Proposition 2. Q.E.D.

Since independence implies association, Proposition 1 is a direct consequence of the above theorem. It is obvious from its proof that the assumptions of Theorem 2 can be weakened further. All that is really needed is that $T_1, ..., T_n$, M are (jointly) associated. Interestingly, there is another condition the latter can be derived from, which also represents a weakening of the assumptions of Theorem 2. We need the following definition.

DEFINITION 5. *Let* $\mathbf{X} = (X_1, ..., X_p)$ be a random vector and W be a random variable on the same probability space; the distribution of $\mathbf{X}$ is called a *monotone mixture with* W if for every increasing function k

$$h(W) = E[k(\mathbf{X}) \mid W] \tag{22}$$

is increasing in W.

The following lemma shows that the monotone mixture condition allows conditionally associated random variables to remain associated (for a generalization, see Jogdeo, 1978).

LEMMA 4. *Assume the conditional distribution of* $\mathbf{X} = (X_1, ..., X_p)$ *given* W *is associated* (*with probability one*). *If the distribution of* $\mathbf{X}$ *is a monotone mixture with* W, *then* $X_1, ..., X_p$, W *are associated.*

Proof. Let f, g be increasing; then

$$\begin{aligned}\operatorname{Cov}&[f(X_1, ..., X_p, W), g(X_1, ..., X_p, W)]\\ &= \operatorname{Cov}[E(f \mid W), E(g \mid W)] + E[\operatorname{Cov}(f, g \mid W)]. \qquad (23)\end{aligned}$$

By hypothesis, $E(f \mid W)$ and $E(g \mid W)$ are increasing in W; thus, by property P4 of association, the first term above is nonnegative. Since $\mathbf{X}$ is associated given W, the second term is nonnegative also, proving the lemma. Q.E.D.

THEOREM 3. *For an n-channel model* ($n \geq 2$) *with unlimited capacity assume* (i) $T_1, ..., T_n$ *are conditionally associated given* M (*with probability* 1), *and* (ii) *the distribution of* $T_1, ..., T_n$ *is a monotone mixture with* M; *then the bounds of Proposition* 1 *hold for the first-terminating and the exhaustive case, respectively.*

Since by Lemma 4, the assumptions of Theorem 3 imply that $T_1, ..., T_n, M$ are associated, the proof of the above is obvious. Moreover, it is not difficult to see that condition (i) and (ii) of Theorem 2 imply conditions (i) and (ii) of Theorem 3.

Concluding this section, a comment on the case of negative dependence seems appropriate. In analogy to our treatment of positive dependence, one might try to impose a multivariate negative dependence property directly on $T_1, ..., T_n$ and then—assuming independence between $T_1, ..., T_n$ and M—derive bounds from there. The most versatile concept of multivariate negative dependence seems to be *negative association* introduced by Joag-Dev & Proschan (1983). Unfortunately, negative association of $\mathbf{T}$ does not carry over to $\mathbf{T}^*$. This should not come as a surprise. As mentioned in Section 5 above, the addition of a component M to each T_k generates a tendency toward positive dependence among the T_k^*. Thus, any negative dependence among the T_k may eventually be offset by M having a sufficiently large variance. This can be illustrated most easily in the binary case. Assume T_1, T_2 to be negatively correlated and independent of M. Then

$$\operatorname{Cov}(T_1 + M, T_2 + M) = \operatorname{Cov}(T_1, T_2) + \operatorname{Var}(M).$$

Obviously, any negative $\operatorname{Cov}(T_1, T_2)$ can be turned into a positive $\operatorname{Cov}(T_1 + M, T_2 + M)$ by some sufficiently large $\operatorname{Var}(M)$. On the other hand, in most empirical situations the variance of the base component M is not likely to be practically unbounded. Thus, conditions on the distributions that allow negative dependence to carry over from $\mathbf{T}$ to $\mathbf{T}^*$ remain to be studied more thoroughly.

7. EXTENSIONS: MIXTURE AND CENSORING

This section extends our analysis in two respects. First, we consider random variables with a finite mixture distribution and, second, we allow random variables to be censored from the right or from the left. The derivations are straightforward, mainly exploiting the fact that the upper and lower bounds in Theorem 1 are linear functions of the marginals. Nonetheless, the results in this section are useful in enlarging the range of applicability of our previous results.

(a) *Finite Mixture Distributions*

Let $A_1, ..., A_s$ be a partition of the sample space, i.e., $P(A_r \cap A_{r'}) = 0$ for $r, r' = 1, ..., s$ $(r \neq r')$ and $\sum_{r=1}^{s} P(A_r) = 1$. Typically, A_r represents a particular attentional state of the subject. The existence of conditional processing time distributions

$$P(T_1 \leqslant t_1, ..., T_n \leqslant t_n \mid A_r), \qquad r = 1, ..., s,$$

each satisfying the unlimited capacity condition is assumed here. Since Lemma 1 obviously applies to conditional probability measures, too, all bounds derived in Section 4 hold for conditional probability distributions separately for each of the conditioning events A_r $(r = 1, ..., s)$. For example, the lower bound in Lemma 2 becomes

$$F_{n-1}^{(i)}(t \mid A_r) \leqslant F_n(t \mid A_r), \qquad r = 1, ..., s.$$

By multiplying both sides of the above by $P(A_r)$ and summing over r one recovers the unconditional probability distribution inequalities of Section 4. Note, however, that the lower (upper) bounds of Theorem 1 can be sharpened somewhat by maximizing (respectively, minimizing) over the indices i, j first. The following corollary illustrates this for the first-terminating case, but the exhaustive case follows analogously.

COROLLARY 3. *Under the above conditions, we have for all t*

$$\begin{aligned}
\max_i F_{n-1}^{(i)}(t) &\leqslant \sum_{r=1}^{s} \max_i \left[F_{n-1}^{(i)}(t \mid A_r)\right] P(A_r) \\
&\leqslant F_n(t) \\
&\leqslant \sum_{r=1}^{s} \min_{i,j} \left[F_{n-1}^{(i)}(t \mid A_r) + F_{n-1}^{(j)}(t \mid A_r) \right. \\
&\qquad \left. - F_{n-2}^{(i,j)}(t \mid A_r)\right] P(A_r) \\
&\leqslant \min_{i,j} \left[F_{n-1}^{(i)}(t) + F_{n-1}^{(j)}(t) - F_{n-2}^{(i,j)}(t)\right].
\end{aligned}$$

Proof. The first and the last inequality above follow from the subadditivity of the max function and the superadditivity of the min function, respectively. For the two inequalities in the middle, note that by hypothesis for $r = 1, ..., s$,

$$\begin{aligned}
\max_i F_{n-1}^{(i)}(t \mid A_r) &\leqslant F_n(t \mid A_r) \\
&\leqslant \min_{i,j} \left[F_{n-1}^{(i)}(t \mid A_r) + F_{n-1}^{(j)}(t \mid A_r) - F_{n-2}^{(i,j)}(t \mid A_r)\right].
\end{aligned}$$

Multiplying by $P(A_r)$ and summing over r then establishes the results. Q.E.D.

Another, potentially important application of this extension to mixture distributions could be directed toward aggregating individual data to group data, whereby the events A_r would represent individual subjects.

(b) *Censoring*

It is common practice in the empirical analysis of reaction times to exclude certain extreme values from further consideration. These extreme values, the argument goes, most likely are not a result of the processes the researcher's interest is focused on. For example, reaction times shorter than 70 ms are often discarded from the analysis since they are likely to result from anticipatory actions by the subject. Moreover, censoring may also be imposed on the observed data by the experimenter using a deadline procedure forcing the subject to respond within a certain time interval.

Also, in memory research there may be reasons to test stochastic inequalities on the latency data provided by correct responses only. The assumption here would be that response omissions turn into correct responses if the subjects are provided with unlimited time to respond.

Below it is shown that modifications of the bounds of Section 4 hold under censoring. Let $I = [t_0, t_1]$ be the observation interval, i.e., all reaction times falling outside of I are either not observed or discarded from the analysis. It is assumed here that the probability of a reaction time occurring within I is known or, in practice, estimable from the data. Since the exhaustive case is analogous, we restrict attention to the first-terminating case. Obviously, for all $t \in I$,

$$\begin{aligned} F_n(t) &= P(\min(T_k) \leqslant t) \\ &= P(\min(T_k) \leqslant t \mid \min(T_k) \in I)\, P(\min(T_k) \in I) \\ &= F_n(t \mid I)\, P(\min(T_k) \in I). \end{aligned}$$

Writing p_n for $P(\min(T_k) \in I)$, $p_{n-1}^{(i)}$ for $P(\min_{k \neq i}(T_k) \in I)$, and so on, we state the following bounds for the censored distribution function.

COROLLARY 4. *For all* $t \in I$

$$\begin{aligned} &\max_i \left[F_{n-1}^{(i)}(t \mid I)(p_{n-1}^{(i)}/p_n)\right] \\ &\quad \leqslant F_n(t \mid I) \\ &\quad \leqslant \min_{i,j} \left[F_{n-1}^{(i)}(t \mid I)(p_{n-1}^{(i)}/p_n) + F_{n-1}^{(j)}(t \mid I)(p_{n-1}^{(j)}/p_n) - F_{n-2}^{(i,j)}(t \mid I)(p_{n-2}^{(i,j)}/p_n)\right]. \end{aligned}$$

The above follows directly from Theorem 1 by simple algebra.

8. CONCLUDING REMARKS

The distributional bounds for n-channel processes presented in Theorem 1 are expressed in terms of marginal distributions of order $n-1$ and $n-2$. Obviously, repeatedly applying Theorem 1 to these marginals yields further inequalities for n-channel processes in terms of even lower order marginals. For example, a single application of Theorem 1 to the $(n-1)$-channel marginal $F_{n-1}(t)$ yields the following upper bounds for $F_n(t)$:

$$F_n(t) \leqslant F_{n-1}^{(j)}(t) + F_{n-2}^{(ik)}(t) - F_{n-3}^{(ijk)}(t). \tag{24}$$

A further application to $F_{n-1}^{(j)}(t)$ then implies

$$F_n(t) \leqslant F_{n-2}^{(ij)}(t) + F_{n-2}^{(jk)}(t) + F_{n-2}^{(ik)}(t) - 2F_{n-3}^{(ijk)}(t). \tag{25}$$

These inequalities being derived from Theorem 1 do not provide sharper bounds for $F_n(t)$. However, if the lower order marginals are available, inequalities (24) and (25) (and others, similarly obtained) provide a further consistency check on the parallel (first-terminating) model.

Moreover, the sharpness of the bounds in Theorem 1 has not been dealt with here. While it is a legitimate question, one reason for the lack of a general result seems to reside in the generality of the inequalities. Since no assumptions on stochastic independence or parametric form of the distributions are being made, sharpness of the bounds may vary widely as a function of those.

Another open issue is the statistical testing of the distribution inequalities proposed in this paper. While some bounds, e.g., the lower bound in Lemma 2, are amenable to the standard Kolmogorov–Smirnov approach, we have no solution to offer for those inequalities where one side constitutes a linear combination of distribution functions. Of course, in many empirical applications of the distribution inequalities proposed here, it may turn out that violations are so obvious that statistical testing would be pointless. Monte Carlo simulations using certain stochastically dependent multivariate distributions (cf. Devroye, 1986) might be helpful in better assessing the properties of the bounds.

9. APPENDIX

Proof of Lemma 3. Let $\mathbf{X} = (X_1, ..., X_p)$ be associated; define $S_k(x_k) = 1$ for $X_k > x_k$ $(k = 1, ..., p)$ and $S_k(x_k) = 0$ for $X_k \leqslant x_k$.

Thus, $S_k(x_k)$ is increasing in $\mathbf{X}$ and, by Property P5, $S_1(x_1), ..., S_p(x_p)$ are associated. Define increasing functions

$$f(\mathbf{S}) = S_1(x_1),\ g(\mathbf{S}) = S_2(x_2) \cdots S_p(x_p);$$

by the definition of association then

$$E[f(\mathrm{S})\,g(\mathrm{S})] \geqslant E[f(\mathrm{S})]\;E[g(\mathrm{S})],$$

which is equivalent to

$$P(X_1 > x_1, ..., X_p > x_p) \geqslant P(X_1 > x_1)\;P(X_2 > x_2, ..., X_p > x_p).$$

By property P2, $X_2, ..., X_p$ are associated; thus, repeated application of the above procedure finally yields

$$P(X_1 > x_1, ..., X_p > x_p) \geqslant \prod_k P(X_k > x_k),$$

i.e., PUOD; PLOD can be shown similarly. Q.E.D.

Acknowledgments

Some of the results of this paper are contained in an earlier techical report (Vorberg *et al.*, 1989) and were stimulated by discussions with Rainer Schmidt to whom we are most grateful. The authors further acknowledge helpful comments by the reviewers and the editor. Moreover, discussions with and/or comments by our colleagues Jerry Busemeyer, Hartmann Scheiblechner, Hans-Henning Schulze, Wolfgang Schwarz, and Rich Schweickert were most helpful.

References

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* Washington: V. H. Winston.

Ashby, F. G., Tein, J-Y., & Balakrishnan, J. D. (in press). Response time distributions in memory scanning. *Journal of Mathematical Psychology.*

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

Bamber, D. (1969). Reaction times and error rates for "same"–"different" judgments of multidimensional stimuli. *Perception & Psychophysics*, **6**, 169–174.

Block, H. W., & Sampson, A. R. (1983). Inequalities on distributions: Bivariate and multivariate. In S. Kotz, N. L. Johnson, and C. B. Read (Eds.), *Encyclopedia of statistical science* (Vol. 4, pp. 76–82). New York: Wiley.

Colonius, H. (1983). A characterization of stochastic independence by association, with an application to random utility theory. *Journal of Mathematical Psychology*, **27**, 103–106.

Colonius, H. (1986). Measuring channel dependence in separate activation models. *Perception & Psychophysics*, **40**, 251–255.

Colonius, H. (1988). Modeling the redundant signals effect by specifying the hazard function. *Perception & Psychophysics*, **43**, 604–606.

Colonius, H. (1990). Possibly dependent probability summation of reaction times. *Journal of Mathematical Psychology*, **24**, 253–275.

Devroye, L. (1986). *Non-uniform random variate generation.* New York: Springer-Verlag.

DIEDERICH, A. (1992a). *Intersensory facilitation with multiple stimuli: Race, superposition, and diffusion models for reaction time.* Frankfurt/M.: Verlag Peter Lang.

DIEDERICH, A. (1992b). Probability inequalities for testing separate activation models of divided attention. *Perception & Psychophysics*, **52**.

DIEDERICH, A., & COLONIUS, H. (1987). Intersensory facilitation in the motor component? A reaction time analysis. *Psychological Research*, **49**, 23–29.

EGETH, H., JONIDES, J., & WALL, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology*, **3**, 674–698.

ERIKSEN, C. W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception*, **44**, 191–193.

ESARY, J. D., PROSCHAN, F., & WALKUP, D. W. (1967). Association of random variables with applications. *The Annals of Mathematical Statistics*, **38**, 1466–1474.

FISHER, D. L. (1982). Limited channel models of automatic detection: Capacity and scanning in visual search. *Psychological Review*, **89**, 662–692.

GRICE, G. R., CANHAM, L., & BOROUGHS, J. M. (1984). Combination rules for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics*, **35**, 451–463.

HERSHENSON, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, **63**, 289–293.

HOLLAND, P., & ROSENBAUM, P. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, **14**, 1523–1543.

JOAG-DEV, K. (1983). Independence via uncorrelatedness under certain dependence structures. *The Annals of Probability*, **11**, 1037–1041.

JOAG-DEV, K., & PROSCHAN, F. (1983). Negative association of random variables, with applications. *The Annals of Statistics*, **11**, 286–295.

JOGDEO, K. (1978). On a probability bound of Marshall and Olkin. *The Annals of Statistics*, **6**, 232–234.

LEHMANN, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, **39**, 1137–1153.

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

MILLER, J. O. (1986). Time-course of coactivation in bimodal divided attention. *Perception & Psychophysics*, **40**, 331–343.

NEWMAN, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. In Y. L. Tong (Ed.), *Inequalities in statistics and probability*. Hayward, CA: Institute of Mathematical Statistics.

NEWMAN, C. M., RINOTT, Y., & TVERSKY, A. (1983). Nearest neighbors and Voronoi regions in certain point processes. *Advances in Applied Probability*, **15**, 726–751.

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

ROSENBAUM, P. (1984). Testing the conditional independence and monotonity assumptions of item response theory. *Psychometrika*, **49**, 425–435.

ROSENBAUM, P. (1988). Item bundles. *Psychometrika*, **53**, 349–359.

RUMELHART, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, **7**, 191–218.

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

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

THORNDYKE, P. W., & BOWER, G. H. (1974). Storage and retrieval processes in sentence memory. *Cognitive Psychology*, **6**, 515–543.

TONG, Y. L. (1980). *Probability inequalities in multivariate distributions.* New York: Academic Press.

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

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

TOWNSEND, J. T. (1990). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedlede buth they can (and should) be distinguished. *Psychological Science*, **1**, 46–54.

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

TVERSKY, A., RINOTT, Y., & NEWMAN, C. M. (1983). Nearest neighbor analysis of point processes: Applications to multidimensional scaling. *Journal of Mathematical Psychology*, **27**, 235–250.

ULRICH, R., & GIRAY, M. (1986). Separate activation models with variable base-times: Testability and decking cross-channel dependency. *Perception & Psychophysics*, **39**, 248–254.

VAN DER HEIJDEN, A. H. C., SCHREUDER, R., MARIS, L. & NEERINCS, M. (1984). Some evidence for correlated separate activation in a simple letter-detection task. *Perception & Psychophysics*, **36**, 577–585.

VORBERG, D. (1977). On the equivalence of parallel and serial models of information processing. Paper presented at the Tenth Annual Mathematical Psychology Meetings, Los Angeles, CA.

VORBERG, D. (1981). Reaction time distributions predicted by serial self-terminating models of memory search. In S. Grossberg (Ed.), *Mathematical Psychology and Psychophysiology*. Proceedings of the symposium in applied mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics (pp. 301–318). New York: American Mathematical Society.

VORBERG, D., COLONIUS, H., & SCHMIDT, R. (1989). Distribution inequalities for parallel models with unlimited capacity. Purdue Mathematical Psychology Technical Report 89-5.

VORBERG, D., & ULRICH, R. (1987). Random search with unequal search rates: Serial and parallel generalizations of McGill's model. *Journal of Mathematical Psychology*, **31**, 1–23.

WINSTON, P. H. (1992). *Artificial intelligence.* Reading, MA: Addison–Wesley.

Received: May 19, 1989