JOURNAL OF MATHEMATICAL PSYCHOLOGY 21, 93–123 (1980)

Decomposing the Reaction Time Distribution: Pure Insertion and Selective Influence Revisited[1]

F. GREGORY ASHBY AND JAMES T. TOWNSEND

Purdue University, West Lafayette, Indiana 47907

This paper investigates the consequences of extending the assumptions of pure insertion and selective influence (popular in RT theorizing) to the level of the distribution. In the case of pure insertion and under the additional assumption that the additive random variable is exponentially distributed, a solution is obtained which not only allows estimation of the exponential-rate parameter but also provides a test of the assumptions. The result is shown to be applicable not only when processing is serial but also for certain parallel models. In addition, discrimination between self-terminating and exhaustive search strategies is provided, and in the case of either, both parameter estimation and tests of the model are possible. Extensions to nonexponential models are investigated and a general method of moments solution is outlined. In the case of selective influence a general nonparametric alternative to Sternberg's additive factor method is developed. The problem of empirical estimation and application is then considered. Simulations which place bounds on the type I and II error are reported. Finally the first theorem is provided an illustrative application with data from a memory scanning experiment. The results provide some support for the double assumption of pure insertion and that the additive random variable is distributed exponentially.

THE ASSUMPTION OF PURE INSERTION

In 1868 Donders proposed estimating the duration of a cognitive process or stage by subtracting mean reaction times (RTs) from two tasks where it was thought that the first task requires all the stages of the second, plus one additional stage. This method of estimation was termed *the method of subtraction* and the postulate that two such tasks can be found has been called the *assumption of pure insertion.*

The method of subtraction has fallen into disfavor primarily because of its strong assumption of pure insertion and also because of the appearance of a method making the seemingly weaker assumption of *selective influence* (Sternberg, 1969). This latter assumption suggests that we again find two tasks, this time with the same number of stages but with some stages differing in duration across the tasks. Although there are other pos-

[1] Some of the results presented here were reported at the German Mathematical Psychology Meetings, University of Konstanz, 1977, and at the Tenth Annual Mathematical Psychology Meetings, UCLA, 1977. This research was begun at the Technical University of Braunschweig where both authors were on leave during the academic year 1976–1977. We are both indebted to the Deutsche Forschungsgemeinschaft and the Institute for Psychology. Parts of this research were supported by Purdue Research Foundation Grant XR0104 and by NSF Grant 76-84053.

Requests for reprints should be sent to F. Gregory Ashby, Ohio State University, Human Performance Center, 404B West 17th Avenue, Columbus, Ohio 43210.

0022-2496/80/020093-31$02.00/0

sibilities, it is usually postulated that if two experimental factors affect mean RT in an additive manner, then two separate stages exist that are distinctively influenced by the factors. Nevertheless, in certain experimental paradigms subtraction is still heavily relied upon. For instance, in the scanning literature (both memory and visual) it is often thought that adding one item to the to-be-scanned list effectively inserts a stage into the RT process.

Another example is in the letter identification task developed by Posner and his colleagues (e.g., Posner & Mitchel, 1967) in which a subject has to respond "same" or "different" depending on the identity of two simultaneously presented letters. In some conditions subjects are instructed to respond "same" only if the letters are physically identical (e.g., *A* and *A*) while in other conditions a "same" response is given if the two letters have the same name (e.g., *a* and *A*). It is often argued that processing to the level of name identity requires only additional decoding steps and thus that subtractive logic is viable (see, e.g., Hunt, 1978).

Another modern instance where the subtractive method might apply is in a detection task where stimulus intensity is varied across experimental conditions. In one condition, stimulus intensity would be set around threshold while in another it would be placed well above threshold. It may be that the second condition reduces to something like a simple RT task without the decision stage necessary to threshold level detection. If so, subtraction may be appropriate.

To make the method of subtraction more precise, assume that in one experimental condition we observe a random RT X and in another condition a random RT Z, where the expected value of Z, $E(Z) = E(X) + E(Y)$. Then subtraction simply recommends estimating $E(Y)$ as the difference in the observed means of Z and X, so that $\widehat{E(Y)} = \bar{Z} - \bar{X}$.

We thus might denote the assumption of this method as pure insertion *at the level of the mean.* Models which admit this possibility sometimes imply much more. That is, given the observable random variables X and Z and the unobservable Y, subtractive models sometimes assume

$$Z = X + Y, \tag{1}$$

where X and Y are stochastically independent. It is easily seen that additivity, not only of the means but of all cumulants, follows from (1). In general, however, tests of this prediction are not powerful. Consider the case of variances (the matter becomes progressively worse with higher cumulants). The variance of the estimate

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

is $2\sigma^4/n$, where σ^2 is the true population variance. RT variances are notoriously large and thus the standard deviation of the RT sample variance is apt to be quite large.

The assumption implicit in (1) might be termed pure insertion *at the level of the distribution.* The assumption is quite common to RT theorizing and is often associated

with mathematical tractability. It can be found in a great number of models postulating a system which performs a set of similar processing tasks serially (as, for example, in many memory scanning models). In fact, as we shall see later, the assumption is not limited to serial models; some parallel models can also be written in the form of Eq. (1).

The present paper begins by extending subtractive logic to the level of the distribution. Then a similar type of analysis is made under the somewhat weaker assumption of selective influence. Surprisingly simple yet powerful results can be obtained which allow not only estimates of the parameters of the models but also tests of their assumptions. Before proceeding further, we make clear some relevant terminology (see also Townsend, 1974). In Fig. 1 we see a hypothetical decomposition of the observable RT for one particular trial of an RT task in which four additive stages are included in the RT process and make up what is labeled the decomposable time. These are the stages for which subtractive logic is assumed acceptable. The base time includes all other processes and their stages for which subtraction seems inappropriate. For instance, almost all models would place a response-execution stage in this latter category.

Next we note that the decomposable time is itself decomposed into four *intercompletion times* (ICTs) representing the fact that the subtractive method does not require the decomposable time to be a strict serial subsystem of the RT process. The system may be strictly parallel or even hybrid. We thus define the jth ICT as the time between the completion of the $(j-1)$th and the jth additive stages. With a serial system each ICT represents the total processing time of a single stage. If search is instead parallel then all stages are begun simultaneously and are completed at the end of some ICT. Thus, for instance, the total processing time of the third stage completed will be the sum of the first three ICTs.

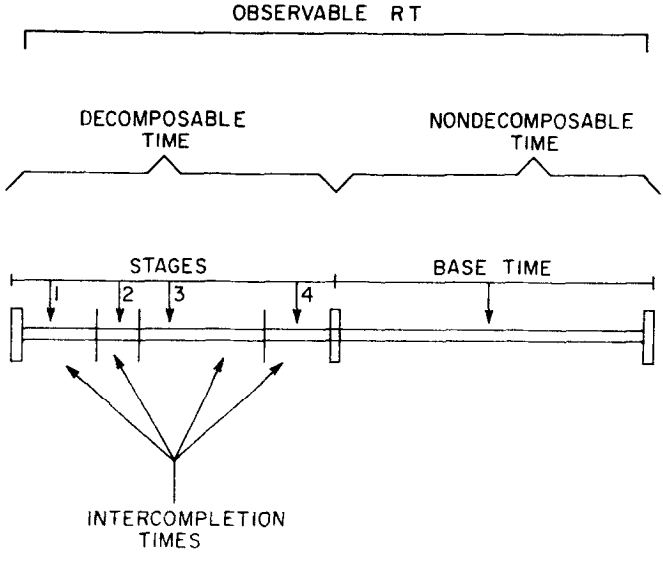

FIG. 1. Schematic representing a decomposition of a hypothetical RT.

Under this guise, the assumption of pure insertion at the level of the distribution becomes

$$RT_k = RT_{k-1} + ICT_k, \tag{2}$$

where RT_k is the random observable RT on trials when there are k ICTs, ICT_k is the random unobservable time of the kth ICT, and RT_{k-1} is independent of ICT_k.

As a concrete example of this framework consider a standard memory scanning task (e.g., Sternberg, 1966). In the language of a popular model of this task (see, e.g., Smith, 1968) the decomposable time would be labeled the comparison time and the number of ICTs would correspond to the number of items in the memory set. The nondecomposable base time would include stimulus encoding, response selection, and response execution.

This type of general model does not rule out the possibility that stages make use of partial information from preceding stages. In the case where stage times overlap, that is, where the processing time of one stage begins between the starting time and the completion time of some other stage, Townsend & Ashby (in press, Chap. 14) have shown that the above ICT definition leads to an additive decomposition of RT as in Eq. (2) for certain classes of automata and linear systems if the successive stages obey a condition of functional independence. McClelland (1979) has recently proposed a linear systems model for simple cognitive processes based on sequential banks of parallel integrators. Additivity is predicted in some cases but not in others.

Our goal in the present work is to estimate the probability density function of ICT_k given a set of RT data satisfying Eq. (2). This is accomplished most elegantly in the parametric case where distributional assumptions are made about ICT_k, but as will be seen nonparametric solutions can also be obtained.

When RT data is modeled at the level of the mean no distributional assumptions need be made, but of the researchers who have proposed models of the RT *distribution* and thus have been forced to make some distributional assumptions about the ICTs, many have selected the exponential density to model this random time between the completion of successive events or stages (e.g., Christie & Luce, 1956; Restle, 1961; McGill, 1963; McGill & Gibbon, 1965; Taylor, 1976; Townsend, 1972, 1974, 1976). The mathematical properties of this distribution are very simple and it turns out that if the added assumption of exponential ICTs is made the present analysis is greatly simplified. Further we shall see that if exponentiality is supported then not only are simpler estimates of the ICT densities obtained but also discrimination between self-terminating and exhaustive processing is possible. As motivation we mention that results will be considered below that render the exponential distribution somewhat less far fetched than is sometimes thought.

Assuming the Distribution of ICT_k is Exponential

Let us denote the density function of RT_k by $g_k(t)$ and the density of ICT_k by $f_k(t)$. Then Eq. (2) implies that $g_k(t)$ may be written as

$$g_k(t) = \int_0^\infty g_{k-1}(t - t_0) f_k(t_0)\, dt_0 = g_{k-1}(t) * f_k(t), \tag{3}$$

that is, $g_k(t)$ is the convolution of $g_{k-1}(t)$ and $f_k(t)$. Under the assumption of this section Eq. (3) can be rewritten as

$$g_k(t) = g_{k-1}(t) * V_k e^{-V_k t}, \qquad t > 0, \tag{4}$$

where V_k is the rate at which the kth ICT unfolds. Thus, an estimate of the kth ICT density, $f_k(t)$, requires us only to obtain an estimate of the parameter, V_k. For instance, a method of moments estimate would suggest

$$\hat{V}_k = \frac{1}{\overline{\mathrm{RT}}_k - \overline{\mathrm{RT}}_{k-1}}$$

since the mean of an exponential density is $E(t) = 1/V_k$. This estimate, while intuitive, ignores much available information in the RT data. Specifically, it provides no test of the assumptions inherent in Eq. (4). Theorem 1 below suggests a way around such limitations, essentially by solving Eq. (4) for the unknown parameter V_k.

THEOREM 1. *If and only if pure insertion at the level of the distribution holds together with the assumption of exponentially distributed ICTs (i.e., Eq. (4)), then*

$$V_k = \frac{g_k(t)}{G_{k-1}(t) - G_k(t)}, \qquad \textit{for any } t > 0,$$

where $G_k(t)$ is the cumulative distribution function of RT_k.

Proof. Taking the Laplace transform of both sides of Eq. (4) yields

$$\mathscr{G}_k(s) = \mathscr{G}_{k-1}(s) \frac{V_k}{V_k + s},$$

where $\mathscr{G}_k(s)$ is the transform of $g_k(t)$. Algebraic manipulation leads to

$$\mathscr{G}_k(s) = \frac{V_k}{s} [\mathscr{G}_{k-1}(s) - \mathscr{G}_k(s)].$$

Now $L\{\int_0^t g(t_0)\, dt_0\} = (1/s)\, \mathscr{G}(s)$ which implies that the inverse transform of the above is

$$g_k(t) = V_k[G_{k-1}(t) - G_k(t)]$$

from which the result easily follows. Thus, Eq. (4) implies the stated solution of V_k.

Assume $g_k(t) = V_k[G_{k-1}(t) - G_k(t)]$ for all $t > 0$. Then arguing as above, but in the reverse order, leads us to

$$g_k(t) = g_{k-1}(t) * V_k \exp(-V_k t), \qquad \text{for all } t > 0,$$

which implies both subtraction at the level of the distribution and the stronger condition of exponential ICTs. Q.E.D.

A result similar to Theorem 1 was proved by McGill *et al.* (1965). The difference is that in addition to the assumptions made here they also assumed that *all* RT-stage durations were exponentially distributed. In this case $g_k(t)$ is a general gamma density and the RT process can be modeled as a generalized birth process (see, e.g., Cox & Miller, 1965). Their result was then derived from the so-called Kolmogorov forward equation which characterizes stochastic processes of this kind.

Theorem 1 is a stronger result since it makes no distributional assumptions about the base time or even about $g_{k-1}(t)$. One way to intuit the result is to note that the addition of an independent random variable (e.g., the base time) to a birth process may affect the starting time or the finishing time of the process but it will not affect the internal structure of the process (i.e., the processing rates).

As a simple example of the application of Theorem 1 assume that the base time (i.e., the nondecomposable time) is uniformly distributed over the interval (0, 2] (i.e., $b(t) = \frac{1}{2}$, $0 < t \leqslant 2$ and $b(t) = 0$ otherwise) and that two ICTs are distributed exponentially, the first with rate $V_1 = 4$ and the second with $V_2 = 3$. Then

$$\begin{aligned} g_2(t) &= b(t) * f_1(t) * f_2(t) \\ &= \tfrac{1}{2}(1 + 3e^{-4t} - 4e^{-3t}) && 0 < t \leqslant 2 \\ &= \tfrac{3}{2}[(\tfrac{1}{3}e^6 + e^{-2} - \tfrac{4}{3})\, e^{-3t} + (e^8 - 1)(e^{-2}e^{-3t} - e^{-4t})] && t > 2 \\ &= 0 && \text{elsewhere} \end{aligned}$$

while

$$\begin{aligned} g_1(t) = f_1(t) * b(t) &= \tfrac{1}{2}(1 - e^{-4t}) && 0 < t \leqslant 2 \\ &= \tfrac{1}{2}(e^8 - 1)e^{-4t} && t > 2 \\ &= 0 && \text{elsewhere.} \end{aligned}$$

Even with such an oversimplified base time the observable RT distribution, $g_1(t)$, as seen in Fig. 2, still possesses an unlimited variate to the right and is not completely unreasonable as a model for many empirical RT distributions (e.g., Christie *et al.*, 1956; Taylor, 1965), particularly considering the usual experimental variability.

Now let us see if Theorem 1 can recover the exponential rate $V_2 = 3$ of the observable RT distribution $g_2(t)$ as it should. First for $0 < t \leqslant 2$

$$V_2 = \frac{g_2(t)}{G_1(t) - G_2(t)} = \frac{\tfrac{1}{2}(1 + 3e^{-4t} - 4e^{-3t})}{\tfrac{1}{2}(t + \tfrac{1}{4}e^{-4t} - \tfrac{1}{4}) - \tfrac{1}{2}(t + \tfrac{4}{3}e^{-3t} - \tfrac{3}{4}e^{-4t} - \tfrac{7}{12})}$$

$$= \frac{1 + 3e^{-4t} - 4e^{-3t}}{\tfrac{1}{3} + e^{-4t} - \tfrac{4}{3}e^{-3t}} = 3,$$

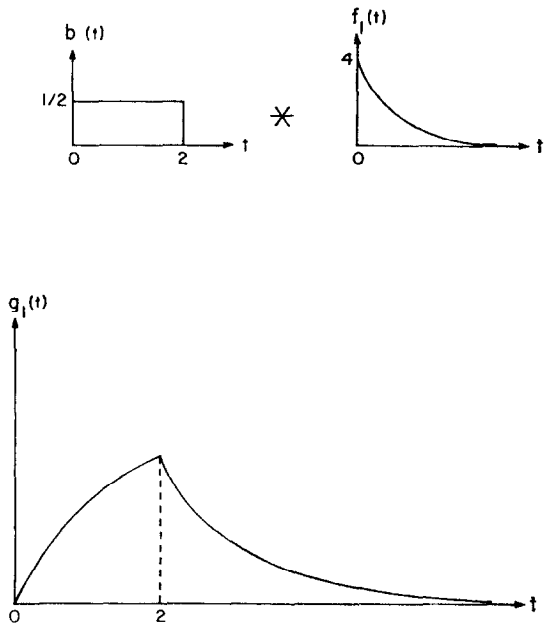

FIG. 2. Results of convolving a uniformly distributed base time with an exponentially distributed ICT.

as expected, and for $t > 2$,

$$V_2 = \frac{g_2(t)}{G_1(t) - G_2(t)}$$

$$= \frac{\frac{3}{2}[(\frac{1}{3}e^6 + e^{-2} - \frac{4}{3})\,e^{-3t} + (e^8 - 1)(e^{-2}e^{-3t} - e^{-4t})]}{1 - \frac{1}{8}(e^8 - 1)\,e^{-4t} - \frac{1}{2}[3(e^8 - 1)(\frac{1}{4}e^{-4t} - \frac{1}{3}e^{-2}e^{-3t}) - (\frac{1}{3}e^6 + e^{-2} - \frac{4}{3})\,e^{-3t} + 2]}$$

$$= \frac{3[\frac{1}{2}(\frac{1}{3}e^6 + e^{-2} - \frac{4}{3})\,e^{-3t} + \frac{1}{2}(e^8 - 1)(e^{-2}e^{-3t} - e^{-4t})]}{[\frac{1}{2}(e^8 - 1)\,e^{-4t} + \frac{1}{2}(e^8 - 1)\,e^{-2}e^{-3t} + \frac{1}{2}(\frac{1}{3}e^6 + e^{-2} - \frac{4}{3})\,e^{-3t}]}$$

$$= \frac{3[\frac{1}{2}(e^8 - 1)(e^{-2}e^{-3t} - e^{-4t}) + \frac{1}{2}(\frac{1}{3}e^6 + e^{-2} - \frac{4}{3})\,e^{-3t}]}{[\frac{1}{2}(e^8 - 1)(e^{-2}e^{-3t} - e^{-4t}) + \frac{1}{2}(\frac{1}{3}e^6 + e^{-2} - \frac{4}{3})\,e^{-3t}]}$$

$$= 3$$

and the second stage rate is successfully recovered.

Although these methods will work equally well with both serial and parallel models, the parallel models satisfying the assumptions may at first appear rather unintuitive. This need not always be the case, however. Consider a model where the decomposable component is an independent parallel unlimited capacity model and where the individual item processing times are exponentially distributed. Then it is well known (see, e.g.,

Townsend, 1972, 1974) that the ICTs will be independent and exponentially distributed.[2] For instance, if the individual item processing rate parameter is V then

$$\begin{aligned} g_1(t) &= b(t) * Ve^{-Vt} \\ g_2(t) &= b(t) * 2Ve^{-2Vt} * Ve^{-Vt} \\ g_3(t) &= b(t) * 3Ve^{-3Vt} * 2Ve^{-2Vt} * Ve^{-Vt} \\ &\vdots \quad \vdots \qquad\qquad \vdots \\ g_n(t) &= b(t) * nVe^{-nVt} * \cdots * Ve^{-Vt}. \end{aligned}$$

Although the density of each ICT changes with k, for the purposes we have in mind, the order of the ICTs is irrelevant since convolution is a commutative operation. Therefore, if we rearrange terms, the above equations can be rewritten as

$$\begin{aligned} g_1(t) &= b(t) * Ve^{-Vt} \\ g_2(t) &= g_1(t) * 2Ve^{-2Vt} \\ g_3(t) &= g_2(t) * 3Ve^{-3Vt} \\ &\vdots \quad \vdots \qquad \vdots \\ g_n(t) &= g_{n-1}(t) * nVe^{-nVt} \end{aligned}$$

so that the conditions of Theorem 1 are met. The fact that there is only one unknown parameter to be estimated allows a result stronger than that of Theorem 1.

COROLLARY. *If processing is independent, parallel unlimited capacity, and exponential with rate V, then*

$$V = \frac{g_k(t)}{kG_{k-1}(t) - kG_k(t)} \qquad t > 0, \quad k = 1, 2, \ldots, n.$$

Proof. Simply set $V_k = kV$ in Theorem 1 to obtain the result. Q.E.D.

It may be noted that the above parallel model is said to be of unlimited capacity because the rate V is constant as k is increased (e.g., Townsend, 1974).

Theorem 1 and it corollary should be, in practice, easy to apply; all we need available are the empirical density and distribution functions for at least two conditions, where the difference between the two conditions is the inclusion of an independent stage (or more precisely an independent ICT) to the RT process in one of them. For instance, in the typical memory scanning paradigm, $g_k(t)$ might be the RT density on target absent trials when k items are in the memory set while $g_{k-1}(t)$ would be the RT density on target absent trials when the memory set contains $k - 1$ items.

If the assumptions of Theorem 1 and the corollary are valid we need only select *any* time $t = t'$ and we can then estimate the processing rate of the kth ICT by $g_k(t')/[G_{k-1}(t') - G_k(t')]$. Of course, it will probably be a good idea to derive estimates of this rate at

[2] In fact, this will be true no matter what the level of the capacity is. Specifically, the ICTs of any independent parallel exponential process are independent and exponentially distributed (e.g., Townsend, 1972).

several different times to test the assumptions. Theorem 1 says these estimates should be the same for *all* values of $t > 0$. This suggests a plot of $g_k(t)/[G_{k-1}(t) - G_k(t)]$ vs t with, say, t as the abscissa. If the assumptions are correct, the plot should be linear, with slope zero and intercept V_k. Any sizable deviations from this form imply a falsification of one or more of the assumptions.

On the other hand, if the resulting plot is flat we know from the theorem that the assumptions must be exactly true, because it is an "if and only if" statement. No other set of assumptions can deliver a flat function. Of course, in emprirical applications the flatness of the resulting function will be a statistical question. As such, there may be models, not formally equivalent to those described by Theorem 1, capable of predicting functions close enough to being flat and linear that an empirical application of the theorem could not determine that they were not. This is essentially a question of the statistical power of such a test and will be discussed as such in a later section where the sampling variability of the Theorem 1 estimate of V_k is considered.

The problem of determining whether a given model can predict an approximately flat and linear function will probably be more difficult to answer for models not based on the notion of the discrete processing stage. One particular example which comes to mind here is McClelland's (1979) cascade model, mentioned earlier, which proposes a deterministic, but overlapping, flow of information through the processing stages. The randomness of the observed RTs arise from the addition of a normally distributed random variable to the proposed deterministic activation. It is unclear what sort of function such a model would predict but it is difficult at this point to imagine an exact equivalence mapping between the cascade model and the discrete stage model of Theorem 1 which would be necessary for the cascade model to generate truly flat linear functions.

Self-Terminating vs Exhaustive Processing

Consider RT paradigms where subjects search through some set of items for the presence of some target or critical item as in the memory scanning task. We continue to assume that the ICTs are exponential.

Until now we have considered only the case in which the observer searches through a stimulus array for the presence of a target item *not* contained in the presented array. This stipulation fairly ensures us that the observer will process *all* items in the display before he responds "no;" that is, that processing is exhaustive (e.g., Townsend, 1974; Ashby, 1976). What happens when the target item is contained in the display? In this case the observer may still process all items regardless of when the target is discovered (exhaustive search) or instead may terminate processing as soon as the critical item completes processing (self-terminating search). In the latter case, with k items in the display we can no longer be sure that the comparison time consists of k ICTs. Only on those trials on which the target is the *last* item completed will there be as many ICTs as items in the display.

Thus, under self-termination, the assumptions of the subtractive method are violated and our earlier results no longer apply. Fortunately we can circumvent this problem without too much difficulty. The result is a self-terminating counterpart of Theorem 1

which, as such, provides both an estimate of the kth ICT density function and a test of the requisite assumptions (namely subtraction, exponentially distributed ICTs, and self-termination). It will be assumed throughout that the target is randomly placed in the comparison set.

First, define the comparison time (or decomposable time) density function when there are a total of j ICTs as $h_j(t)$. The observed time on any given trial when j of the k items are processed will be a sample from the distribution

$$g_{k,j}(t) = h_j(t) * b(t), \qquad 1 \leqslant j \leqslant k,$$

where

$$h_j(t) = f_1(t) * f_2(t) * \cdots * f_j(t), \qquad 1 \leqslant j \leqslant k.$$

Under our assumption of exponentiality, $h_j(t) = C_1e^{-C_1t} * C_2e^{-C_2t} * \cdots * C_je^{-C_jt}$. When search is self-terminating, no matter what the search order, the observed distribution over an infinite number of trials will be a composite formed by a probability mixture of the $g_{k,j}$ for the various values of j. Thus,

$$g_k(t) = E_j[g_{k,j}(t)] = \sum_{j=1}^{k} \frac{1}{k} g_{k,j}(t) = \sum_{j=1}^{k} \frac{1}{k} [h_j(t) * b(t)]$$

$$= \left[\frac{1}{k} \sum_{j=1}^{k} h_j(t)\right] * b(t), \tag{5}$$

due to the linearity of the convolution operation. We can use this result to show that if search is self-terminating Theorem 1 no longer holds; that is, the ratio of g_k to $G_{k-1} - G_k$ is not constant over t.

LEMMA 1. *If the conditions of Theorem* 1 *are met and in addition processing is self-terminating, then on target present trials it is* never true *that*

$$\frac{g_k(t)}{G_{k-1}(t) - G_k(t)} = W_k, \qquad \textit{for all } t > 0.$$

Proof. It follows from Theorem 1 that $g_k(t) = C_1e^{-C_1t} * C_2e^{-C_2t} * \cdots * C_ke^{-C_kt} * b(t)$ but due to the self-terminating character of the comparisons, it is also the case that

$$g_k(t) = \left\{\sum_{j=1}^{k} \frac{1}{k} [B_1e^{-B_1t} * B_2e^{-B_2t} * \cdots * B_je^{-B_jt}]\right\} * b(t).$$

Ridding the above expressions of $b(t)$ via the Laplace transform and its inverse and setting these equal to one another, it follows that

$$C_1e^{-C_1t} * \cdots * C_ke^{-C_kt} = \sum_{j=1}^{k} \frac{1}{k} [B_1e^{-B_1t} * \cdots * B_je^{-B_jt}], \qquad t > 0,$$

that is, a uniform probability mixture of general gamma distributions, where the number

of stages runs from 1 to k, must be equivalent to a single general gamma distribution with k stages. This is impossible as we show by evaluating an equation involving the Laplace transforms of the two sides of the above expression.

$$\prod_{i=1}^{k} \frac{C_i}{C_i + s} = \frac{1}{k} \sum_{j=1}^{k} \prod_{i=1}^{j} \frac{B_i}{B_i + s}, \qquad -\infty \leqslant s \leqslant +\infty, \quad k < \infty.$$

We rule out the possibility that k is infinite as absurd. By multiplying through by the denominator terms we arrive at an equation of the form: $D_{2k-1}s^{2k-1} + D_{2k-2}s^{2k-2} + \cdots + D_1 s + D_0 = 0$, for all $s \in (-\infty, +\infty)$, where the D's are in general made up of polynomial combinations of the B's and C's. As indicated, this equation must be satisfied for all s so by the method of undetermined coefficients all coefficients have to identically equal 0. But, $D_{2k-1} = (1/k)B_1$ implying $B_1 = 0$. Now $B_1 = 0$ only when the mean first-stage processing time is infinite which has been ruled out. We have thus shown that both hypotheses cannot hold simultaneously in a nondegenerate fashion. Q.E.D.

It should be noted that we have not shown that no distributions $f_1, f_2, \ldots, f_k$ exist such that the distribution

$$\frac{1}{k} \sum_{j=1}^{k} [f_1(t) * \cdots * f_j(t)]$$

is equivalent to a general gamma distribution. In fact, at least for $k = 2$, it can be shown that they do exist but we have not yet proven it in the general case.

Suppose now that $g_k(t)/[G_{k-1}(t) - G_k(t)]$ vs time, when plotted for the *target-absent* data, is flat, so that the Theorem 1 assumptions are supported. We now repeat this procedure with the target-present data. If the plot is again flat we can, because of the above lemma, rule out self-termination; within the class of models under consideration an exhaustive search is supported. If the plot is not flat we can rule out an exhaustive search and thus we might suspect self-termination. Of course, there are other possibilities, for example, search could be neither exhaustive nor self-terminating but something else, or the ICT distribution might have changed (i.e., no longer exponential). In any case, a direct test of the self-terminating hypothesis would appear valuable. Specifically if exhaustive processing is not supported, we desire a way to *verify* self-termination and also to recover the processing rates; that is, we desire a result similar to Theorem 1 in the case that search is self-terminating. Now Theorem 1 is really more a result about the $g_{k,j}(t)$ than about the $g_k(t)$. In the special case of an exhaustive search, the two functions are equal when j is set equal to k and interpretation of Theorem 1 becomes straightforward. Thus, what we need in order to apply Theorem 1 when search is self-terminating is a means to relate the unobservable $g_{k,j}(t)$ to the observable $g_k(t)$. Theorem 2 provides a solution to this problem.

THEOREM 2. *If the conditions of Theorem* 1 *are met and, in addition, processing is self-terminating, then for all* $t > 0$

$$V_k = \frac{kg_k(t) - (k-1)\, g_{k-1}(t)}{2(k-1)\, G_{k-1}(t) - (k-2)\, G_{k-2}(t) - kG_k(t)}.$$

Proof. Now

$$g_{k,k}(t) = h_k(t) * b(t)$$
$$= k \sum_{j=1}^{k} \frac{1}{k} [h_j(t) * b(t)] - (k-1) \sum_{j=1}^{k-1} \frac{1}{k-1} [h_j(t) * b(t)],$$

which for self-termination reduces, through Eq. (5) to

$$g_{k,k}(t) = kg_k(t) - (k-1)\, g_{k-1}(t).$$

Likewise

$$G_{k,k}(t) = kG_k(t) - (k-1)\, G_{k-1}(t).$$

Because of the equality $g_{k,k}(t) = g_k(t)$, the result of Theorem 1 can be written as

$$V_k = \frac{g_{k,k}(t)}{G_{k-1,k-1}(t) - G_{k,k}(t)}.$$

Substituting the above self-terminating counterparts into this expression gives us the desired result. Q.E.D.

We have now a method which allows us to observe the processing rates of each ICT and at the same time to test the assumptions necessary for this observability. In addition, on target-present trials, we can discriminate between self-terminating and exhaustive processing. In the case of either we can observe the target-present processing rates.

Relaxing the Exponential Assumption

Earlier it was stated that the assumption of exponential ICTs is not absolutely necessary since the ICT density can be estimated without this restriction. There are at least two general approaches we can take with regard to this problem. In the first we can make a specific nonexponential distributional assumption about the form of the ICT density. Then we can attempt to solve Eq. (3) for that parameter of $f_k(t)$ which identifies the distribution. Under the assumption of exponentiality we solved easily for V_k but in general, most solutions, in addition to being more difficult to obtain, will more importantly be more intransigent to implement with empirical data. In addition, many distributions are characterized by more than one parameter.

However, there is an alternative approach, similar to the method of moments, which can sometimes be easily realized. We illustrate it here for the special case when the kth ICT has a gamma distribution with n_k stages and rate V_k. In this case Eq. (3) becomes

$$g_k(t) = g_{k-1}(t) * \frac{(V_k t)^{n_k - 1}}{(n_k - 1)!} V_k e^{-V_k t}, \qquad t > 0.$$

If we now assume $g_k(t)$ has a well-defined moment generating function (m.g.f.) $\mathcal{G}_k(\theta)$, then the m.g.f. of the above expression is

$$\mathcal{G}_k(\theta) = \mathcal{G}_{k-1}(\theta) \left(\frac{V_k}{V_k - \theta}\right)^{n_k}.$$

Taking logarithms of both sides gives us

$$\log \mathcal{G}_k(\theta) = \log \mathcal{G}_{k-1}(\theta) + n_k \log \left(\frac{V_k}{V_k - \theta}\right)$$

or

$$\log \left(\frac{V_k}{V_k - \theta}\right) = \frac{1}{n_k}[\log \mathcal{G}_k(\theta) - \log \mathcal{G}_{k-1}(\theta)]. \tag{6}$$

Now $\log \mathcal{G}(\theta)$ is the cumulant generating function of $g(t)$ and thus the derivative of (6) with respect to θ and evaluated at $\theta = 0$ is

$$\frac{1}{V_k} = \frac{1}{n_k}(\mu_k - \mu_{k-1}), \tag{7}$$

where μ_k represents the mean of the observable g_k distribution. Similarly,

$$\frac{d^2}{d\theta^2} \log \left(\frac{V_k}{V_k - \theta}\right)\Big|_{\theta=0} = \frac{d^2}{d\theta^2}\left[\frac{1}{n_k} \log \mathcal{G}_k(\theta) - \frac{1}{n_k} \log \mathcal{G}_{k-1}(\theta)\right]\Big|_{\theta=0}$$

reduces to

$$\frac{1}{V_k^2} = \frac{1}{n_k}(\sigma_k^2 - \sigma_{k-1}^2), \tag{8}$$

where σ_k^2 is the variance of $g_k(t)$.

Thus in (7) and (8) we have two equations and two unknowns. Simple algebra leads us to observable estimates of the two unknown parameters of the kth ICT, namely

$$\hat{V}_k = \frac{\hat{\mu}_k - \hat{\mu}_{k-1}}{\hat{\sigma}_k^2 - \hat{\sigma}_{k-1}^2}$$

and

$$\hat{n}_k = \frac{(\hat{\mu}_k - \hat{\mu}_{k+1})^2}{\hat{\sigma}_k^2 - \hat{\sigma}_{k-1}^2}.$$

These estimates hold for any base time, but do require subtraction at the level of the distribution, and that the ICTs are distributed gamma, although each may have a different number of stages and a different rate. Note that realistic estimates of V_k and n_k can be obtained only if both the mean and variance of the observable RT distributions increase monotonically with k. This restriction is, in fact, implied by pure insertion at the distributional level as we earlier noted.

Reliable estimates of higher-order moments of RT distributions require careful experimental procedure and a fairly large sample size (e.g., Ratcliff, 1979). However, arguments based on RT variances are becoming more and more prevalent (e.g., Schneider & Shiffrin, 1977; Rossmeissl, Theios, & Krunnfusz, 1979) and thus the above estimates may become increasingly more feasible.

If the ICTs are gamma distributed and pure insertion holds at the distributional level then we should be able to estimate V_k and n_k using third and fourth cumulants instead of first and second. If the assumptions of pure insertion and gamma distributed ICTs are true the two sets of estimates should agree, thus providing a test of the general model. Of course the estimation problems with third and fourth cumulants are even more severe than for variances. Thus, at present, such a test may not be practical. However, recent results in moment estimation (e.g., Shiffrin & Thompson, 1979) offer some hope that, at least in some cases, more reliable moment estimates may become available.

Let $\kappa_{r,k}$ be the rth cumulant of the observable RT distribution when k elements are processed. Then successively differentiating (6), three and four times, respectively, yields

$$\frac{2}{V_k^3} = \frac{1}{n_k}(\kappa_{3,k} - \kappa_{3,k-1})$$

and

$$\frac{6}{V_k^4} = \frac{1}{n_k}(\kappa_{4,k} - \kappa_{4,k-1}).$$

Again we solve for V_k and n_k:

$$\hat{V}_k = 3\,\frac{\hat{\kappa}_{3,k} - \hat{\kappa}_{3,k-1}}{(\hat{\kappa}_{4,k} - \hat{\kappa}_{4,k-1})}$$

and

$$\hat{n}_k = \frac{27}{2}\,\frac{(\hat{\kappa}_{3,k} - \hat{\kappa}_{3,k-1})^4}{(\hat{\kappa}_{4,k} - \hat{\kappa}_{4,k-1})^3}.$$

These two sets of estimates provide a test of the assumptions, for if they are true it must be the case that

$$V_k = \frac{\mu_k - \mu_{k-1}}{\sigma_k^2 - \sigma_{k-1}^2} = 3\,\frac{(\kappa_{3,k} - \kappa_{3,k-1})}{(\kappa_{4,k} - \kappa_{4,k-1})}$$

and

$$n_k = \frac{(\mu_k - \mu_{k-1})^2}{\sigma_k^2 - \sigma_{k-1}^2} = \frac{27}{2}\,\frac{(\kappa_{3,k} - \kappa_{3,k-1})^4}{(\kappa_{4,k} - \kappa_{4,k-1})^3}.$$

This method will work well with many distributions. An especially simple case arises if one or more of the RT components are approximately normally distributed, since all cumulants of a normal distribution other than the first and second equal zero. Thus the higher cumulants of the RT distribution will be unbiased estimates of the cumulants of the nonnormally distributed component distributions (e.g., see Taylor, 1965).

So far we have not dropped the third (distribution) assumption; we have merely examined the consequences of substituting some other distribution for the exponential. In general, an analytically simple estimate of the ICT density can be obtained even when no distributional assumptions are made.

LEMMA 2. *If pure insertion at the level of the distribution holds, then*

$$f_k(t) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}[g_k(t)]}{\mathcal{F}[g_{k-1}(t)]}\right\} = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}[G_k(t)]}{\mathcal{F}[G_{k-1}(t)]}\right\},$$

where $\mathcal{F}$ represents the Fourier transform.

Proof. If

$$g_k(t) = g_{k-1}(t) * f_k(t)$$

then it is also true that

$$G_k(t) = G_{k-1}(t) * f_k(t).$$

Taking Fourier transforms of both sides of these two equations and solving for $f_k(t)$ yields the Lemma. Q.E.D.

Thus we have an estimate of the kth ICT density function which is completely observable and yet makes no distributional assumptions. The investigator simply divides the Fourier transform of the empirical distribution (or density) function when there are k ICTs by the transform of the empirical distribution (density) function when there are $k - 1$ ICTs and takes the inverse transform of this result. Here the method of Fast Fourier Transforms will prove useful (e.g., McGillem & Cooper, 1974). Fourier methods employing empirical RT densities have been utilized in the past, usually with indifferent success as the resulting estimates generally turn out to be very noisy (e.g., Green & Luce, 1971). Employing cumulative distribution functions as in Lemma 2 should reduce this problem somewhat as distribution function estimates are more stable than density estimates.

Under these same Lemma 2 assumptions, Bloxom (1979) has developed an estimate of $f_k(t)$ based on cubic splines which does not seem as susceptible as the Fourier methods to the noise or higher-frequency components of the empirical density estimate and thus appears to be a promising alternative to the estimate of Lemma 2.

THE ASSUMPTION OF SELECTIVE INFLUENCE

The assumptions of the subtractive method are admittedly very strong and can be expected to hold only in certain special circumstances. There are several ways one might weaken the assumptions yet still retain enough specificity to generate testable predictions. Probably the most popular of these is to assume selective influence rather than pure insertion (e.g., Sternberg, 1969). Under selective influence, if we are given an RT process consisting of a base time and k ICTs then we attempt to find an experimental factor which *affects* only one ICT and nothing else (rather than deleting it as in subtraction).

In this section, we extend this assumption to the level of the distribution in much the same way in which we earlier extended pure insertion. The developments, however, are logically independent. In the beginning, we make no distributional assumptions about any of the RT components.

Sternberg (1969) established the additive factor method as a technique to analyze RT data from experiments in which selective influence is thought to hold. Suppose we have several levels of each of two different experimental factors. Then the manner in which the additive factor method is usually implemented is to perform an analysis-of-variance (ANOVA) on the RT data. If the two factors are found to be additive, then the *assumption* is that the two factors must be influencing separate stages in the RT process. If the two factors are found to interact, it is assumed that they must be influencing at least one stage in common.

This method, although widely popular, has from its inception received criticism (see, e.g., Theios, 1973; Pachella, 1974; Taylor, 1976). For instance, it is deemed unlikely that all of the assumptions of the ANOVA model are satisfied with typical RT data (e.g., normality, homogeneity of variance). Also the large RT variances and small number of subjects often used indicate low statistical power. This suggests that the probability of incorrectly deciding that two factors are additive may be unreasonably large.

We now outline a new nonparametric method for studying RT processes when the assumption of selective influence is appropriate. In a fashion analogous to our treatment of pure insertion we can extend selective influence to the level of the distribution by assuming that when additivity holds the processing time of the stage selectively influenced is stochastically independent of the processing times of the other stages. This extra assumption allows fairly strong nonparametric tests of the hypothesized additivities which do not require the questionable assumptions of ANOVA.

The assumption that selective influence holds at the level of the distribution is stronger than assuming it holds at the mean level. Conceivably, a factor might affect the mean of only one stage but, say, the variances of several. While statistically this may be unassailable, it is not immediately clear whether it is always in the spirit of an assumption of selective influence, since the factor in question is indeed affecting more than one stage. Whether or not it does make sense is a qualitative consideration which will depend on specific empirical circumstances.

To illustrate the technique assume we have two levels of two experimental factors, A and B, and each factor influences a different stage of processing at the distributional level. Then we may write the probability density function when factor A is at level i and factor B is at level j as

$$g_{A_iB_j}(t) = b(t) * f_{A_i}(t) * f_{B_j}(t), \tag{9}$$

where $b(t)$ is the density function of all stages not influenced by either factor A or factor B. Under these experimental circumstances the following result is obtained.

THEOREM 3. *If the assumption of selective influence at the level of the distribution holds (as in Eq. (9), then for all* $t > 0$

$$g_{A_1B_1}(t) * g_{A_2B_2}(t) = g_{A_1B_2}(t) * g_{A_2B_1}(t)$$

or alternatively

$$G_{A_1B_1}(t) * g_{A_2B_2}(t) = G_{A_1B_2}(t) * g_{A_2B_1}(t).$$

Proof. The Laplace transform of (9) is

$$\mathscr{G}_{A_iB_j}(s) = \mathscr{B}(s)\,\mathscr{F}_{A_i}(s)\,\mathscr{F}_{B_j}(s),$$

which leads to

$$\frac{\mathscr{F}_{A_1}(s)}{\mathscr{F}_{A_2}(s)} = \frac{\mathscr{G}_{A_1B_1}(s)}{\mathscr{G}_{A_2B_1}(s)} = \frac{\mathscr{G}_{A_1B_2}(s)}{\mathscr{G}_{A_2B_2}(s)}$$

and

$$\frac{\mathscr{F}_{B_1}(s)}{\mathscr{F}_{B_2}(s)} = \frac{\mathscr{G}_{A_1B_1}(s)}{\mathscr{G}_{A_1B_2}(s)} = \frac{\mathscr{G}_{A_2B_1}(s)}{\mathscr{G}_{A_2B_2}(s)}.$$

The result follows from either expression. Q.E.D.

Theorem 3 provides a test of the assumption of selective influence. One collects RT data from the four conditions, estimates the RT density functions, performs the numerical convolutions and compares the results in some appropriate way such as via a Kolmogorov–Smirnov two-sided test. The Kolmogorov–Smirnov test uses the maximum absolute deviation of the two empirical distribution functions as its test statistic. It has consistently been found to be a more powerful goodness-of-fit test than the more widely known χ^2 test (e.g., Massey, 1951). The Kolmogorov–Smirnov test for the equality of two distribution functions is a nonparametric test and thus Theorem 3 provides a test of additivity which does not require specifying the *form* of the RT density as in ANOVA.

Applications of Theorem 3 are currently being performed and at least one early application (Ashby, 1979) indicates that selective influence at the distributional level may sometimes be a reasonable assumption.

An interesting and potentially important question which we might ask about Theorem 3 is whether the implication holds also in the reverse (ascending) direction. That is, does the result of Theorem 3 imply the RT decomposition of Eq. (9) and in so doing selective influence at the distributional level? The answer, as it turns out, is "yes" and "maybe not," respectively. Towsend *et al.* (in press, Chap. 14), under somewhat different motivations, solved a functional equation which is essentially the same as the one in Theorem 3. Their solution indicates that the function $g_{A_iB_j}$ is indeed constrained to the sort of decomposition evidenced in Eq. (9). However, it turns out that the functions f_{A_i}, f_{B_j}, and b of (9) are not constrained to be true probability density functions. Thus there seem to be analytic models other than of the type we have been considering here, which can predict the result of Theorem 3. Whether any of these have realistic processing interpretations is unclear. For instance, we have elsewhere shown that a large and eminently reasonable class of independent parallel models cannot be written in this form (Townsend *et al.*, in press, Chap. 14). Thus it may be that the class of plausible (in the sense of their processing interpretations) models predicting Theorem 3 is precisely the class of models we considered when writing Eq. (9); namely, those assuming selective influence at the distributional level.

An extension of this approach to three factors is straightforward. If three experimental factors selectively influence three separate stages at the distributional level, then the associated RT density in condition $A_iB_jC_k$ is given by

$$g_{A_iB_jC_k}(t) = b(t) * f_{A_i}(t) * f_{B_j}(t) * f_{C_k}(t). \tag{10}$$

Instead of the one testable prediction (of Theorem 3) generated when two additive factors are postulated the completely additive three-factor model generates 6 independent testable predictions when each factor has two levels.

LEMMA 3. *If three experimental factors selectively influence three RT stages at the distributional level (as in Eq.* (10)*), then for all* $t > 0$

(i) $G_{A_1B_1C_1}(t) * g_{A_2B_2C_1}(t) = G_{A_1B_2C_1}(t) * g_{A_2B_1C_1}(t)$

(ii) $G_{A_1B_1C_2}(t) * g_{A_2B_2C_2}(t) = G_{A_1B_2C_2}(t) * g_{A_2B_1C_2}(t)$

and

(iii) $G_{A_1B_1C_1}(t) * g_{A_2B_1C_2}(t) = G_{A_1B_1C_2}(t) * g_{A_2B_1C_1}(t)$

(iv) $G_{A_1B_2C_1}(t) * g_{A_2B_2C_2}(t) = G_{A_1B_2C_2}(t) * g_{A_2B_2C_1}(t)$

and

(v) $G_{A_1B_1C_1}(t) * g_{A_1B_2C_2}(t) = G_{A_1B_1C_2}(t) * g_{A_1B_2C_1}(t)$

(vi) $G_{A_2B_1C_1}(t) * g_{A_2B_2C_2}(t) = G_{A_2B_1C_2}(t) * g_{A_2B_2C_1}(t)$.

Restriction (i) is essentially the same result as was obtained in Theorem 3 for the case when factor C is held constant at level 1 while in (ii) factor C is at level 2. Thus restrictions (i) and (ii) together reflect the additivity of factors A and B. Similarly (iii) and (iv) result from the additivity of factors A and C while restrictions (v) and (vi) arise from the additivity of factors B and C.

Now suppose that we perform an experiment with all possible combinations of these different factor levels and that we estimate the appropriate density and distribution functions. We are now in a position to perform the tests suggested in Lemma 3.[3] If none of the six predictions can be rejected then, of course, the completely additive model of Eq. (10) becomes the model to beat in this paradigm and is thus indirectly supported. If they are all rejected, then additive factor theory assumes that all factors affect some stage in common. There are, of course, other possibilities. For instance, the factors might influence separate stages which interact with each other through some feedback loop.

[3] When performing these tests, care must be taken to control the type I error level. That is, the individual test type I error level should be selected so that the error level with respect to the collection of all tests made is not unreasonably large.

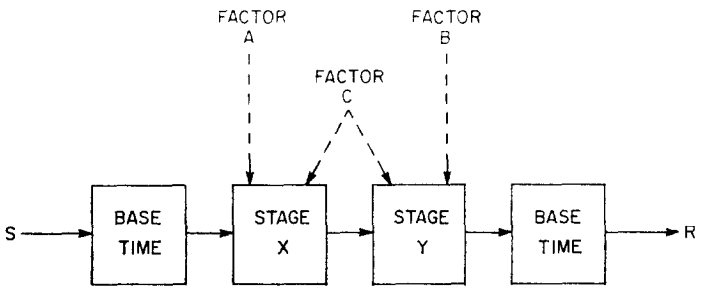

FIG. 3. The hypothetical effects of three experimental factors on the stages of an RT process. It is assumed that factors A and B are additive and that factor C interacts with each of factors A and B. The predicted RT densities for the different experimental conditions are given in Eq. (11).

Suppose now that an application of Lemma 3 rejects predictions (iii) through (vi) but not predictions (i) and (ii). The problem now is to postulate a model consonant with this pattern of results. One possibility is to assume, as in Fig. 3, that factors A and C influence a stage in common as do factors B and C but that factors A and B influence separate stages. Under such a model the predicted RT density function in experimental condition $A_iB_jC_k$ is

$$g_{A_iB_jC_k}(t) = b(t) * f_{A_iC_k}(t) * f_{B_jC_k}(t). \tag{11}$$

It is easily shown that this equation predicts (i) and (ii) of Lemma 3.

It is important to note that even though Eq. (11) implies stochastic independence of the densities $f_{A_iC_k}(t)$ and $f_{B_jC_k}(t)$, in general the stage X and Y processing times could still evidence correlations when the data is compiled over the C conditions. For instance, assume factor C represents degree of subject motivation which is controlled by varying reward magnitudes. Now suppose that both stage X and Y processing times are faster when motivation (i.e., the reward magnitude) is high. Then we may still be able to get a stochastically independent RT decomposition along the lines of Eq. (11) and yet in conditions (via C) when the stage X processing time is short the stage Y time will also be short (and vice versa). What is required by Eq. (11) is that for any given level of factor C the stage X and Y processing times be stochastically independent. If, though, we average over levels of C this no longer need be true. The stage X and Y processing times can, at this level, be dependent.

It may also be possible to attack the thorny problem of errors using this general methodology. For instance in Eq. (10) suppose C_1 refers to correct responses and C_2 to incorrect responses. Then under the assumption that the errors arise from stages not influenced by either of the additive factors A or B, Lemma 3 should hold. If, on the other hand, some of the six predictions are rejected, then by observing the pattern of rejections one can still hypothesize the stage or stages which form the locus of the processing errors. This general line of research is currently being elaborated and more deeply investigated.

The preceding material outlines a general methodology for analyzing RTs when selective influence is thought to hold. The approach is freed of many assumptions inherent

to the additive factor method and the ANOVA model. Specifically, no distributional assumptions about the RT need be made as is the case with ANOVA. It need not be assumed that the RT variances in the different experimental conditions are all equal. In exchange we are forced to assume that the RT can always be decomposed into a sum of independent random variables within any fixed experimental conditions. This does not mean that the individual stage processing times must be independent when conditions are varied. In general, they need not be. What it does mean, however, is that if additivity is found at the level of the mean then the additivity must also hold for all higher moments.

Theoretically, an examination of the means is not necessary with this method. If additivity exists at the distributional level, it necessarily exists at the mean level although the converse is, of course, not true. Practically speaking, however, one may wish to incorporate a means analysis in conjunction with the above approach as a time-saving strategy since firmly established interaction at the mean level could rule out an additivity at the distributional level.

Distributional Assumptions and Selective Influence

A myriad of possibilities present themselves if we permit distributional assumptions about the RT components. The methods of solution are similar to those we utilized when pure insertion was assumed and for this reason we will not attempt an exhaustive investigation. However, to illustrate the types of results one might expect to obtain, let us consider the case where some experimental factor A selectively influences a stage (i.e., an ICT) of the RT process whose completion time is exponentially distributed. Specifically assume that under level one of factor A the stage completion time density is Ve^{-Vt} and under level two it is We^{-Wt}. We assume all other stages are unaffected by factor A.

Then the observable density function when factor A is at level A_1 is

$$g_{A_1}(t) = b(t) * Ve^{-Vt} \tag{12}$$

while at level A_2,

$$g_{A_2}(t) = b(t) * We^{-Wt}. \tag{13}$$

In these expressions $b(t)$ is the density function of all stages not affected by factor A. This representation admits the following result.

THEOREM 4. *If the completion time of some stage X is exponentially distributed and is independent of the completion time of all other stages, and if an experimental factor A affects only the rate of stage X, then*

$$\left[\frac{g_{A_1}(t)}{G_{A_2}(t) - G_{A_1}(t)}\right] = \frac{V}{W}\left[\frac{g_{A_2}(t)}{G_{A_2}(t) - G_{A_1}(t)}\right] + V, \qquad \textit{for all } t > 0,$$

where V is the rate of stage X under level one of factor A and W is the rate under level two of A.

Proof. Eqs. (12) and (13) imply

$$\mathscr{G}_{A_1}(s) = \mathscr{B}(s)\left(\frac{V}{V+s}\right) \quad \text{and} \quad \mathscr{G}_{A_2}(s) = \mathscr{B}(s)\left(\frac{W}{W+s}\right)$$

which together imply

$$\mathscr{G}_{A_2}(s) = \frac{W}{V}\mathscr{G}_{A_1}(s)\left(\frac{V+s}{W+s}\right).$$

Now writing $(V+s)/(W+s) = 1 + (V-W)/(W+s)$ and some algebraic manipulation leads to

$$\frac{1}{s}[\mathscr{G}_{A_2}(s) - \mathscr{G}_{A_1}(s)] = \frac{1}{V}\mathscr{G}_{A_1}(s) - \frac{1}{W}\mathscr{G}_{A_2}(s)$$

from which the result easily follows. Q.E.D.

Theorem 4 suggests that a plot of

$$\frac{g_{A_1}(t)}{G_{A_2}(t) - G_{A_1}(t)} \text{ vs } \frac{g_{A_2}(t)}{G_{A_2}(t) - G_{A_1}(t)}, \quad \text{for } t > 0$$

should yield a linear function with slope V/W and intercept V if the assumptions of the theorem are valid. We can thus use Theorem 4 in much the same way as Theorem 1. We perform the experiment and estimate the density and distribution functions. We next plot $g_{A_1}(t)/[G_{A_2}(t) - G_{A_1}(t)]$ vs $g_{A_2}(t)/[G_{A_2}(t) - G_{A_1}(t)]$, which allows us both to test the assumptions of the theorem and if they are supported (i.e., if the plot is linear with positive slope and intercept) to estimate both V and W. The estimate of V is given by the intercept of the regression line while the estimate of the inverse of W is the slope of the regression line divided by the intercept.

As mentioned earlier, the specific import of results of this type will depend on circumstances of the empirical application one has in mind. At any rate, we see that it is sometimes possible to both estimate model parameters and test assumptions. This appears to be the primary benefit of extending assumptions such as pure insertion and selective influence to the distributional level. At the mean level parameter estimates can often be obtained but seldom is a test of the assumptions of the model possible.

Simulations, Type I and II Errors, and an Empirical Application

We turn briefly now to empirical applications of Theorem 1. Applications of the other results, and in particular the method developed when selective influence is assumed, will be the target of future efforts.

Before one begins applying Theorem 1, there are a few things about which we need to be reassured. One which comes to mind is the specificity of the results. If, for example,

the ICTs are distributed other than exponential, will the deviations from a flat linear function, when we plot

$$\frac{g_k(t)}{G_{k-1}(t) - G_k(t)} \text{ vs } t,$$

be significant enough for us to safely conclude that at least one of the assumptions has failed?

Ideally, of course, we would like to be able to perform a significance test on, say, the slope of the least-squares regression line fit to this set of data. Unfortunately, since we have only one data point for each value of t, standard regression theory does not provide such a test. A second alternative would be to obtain standard errors of the estimated slopes and then to construct confidence intervals or our own significance tests. Again, it unfortunately appears that the problem of calculating such standard errors is analytically intractable since it involves the exceedingly complex statistic $\hat{g}_k(t)/[\hat{G}_{k-1}(t) - \hat{G}_k(t)]$. This state of affairs leaves us in somewhat of a lurch, a common enough situation when one encounters new and slightly exotic statistics. In our case, an attempt is made to estimate a standard error by performing a series of simulations and then calculating the standard deviation on the resulting regression-line slopes.

Monte Carlo simulations were performed in which sets of data were generated fulfilling all or only some of the earlier assumptions. In all of these simulations the density functions, $g_k(t)$, were then estimated following Parzen (1962) with a Gaussian kernel so that

$$\hat{g}_k(t; n) = \frac{1}{nh} \sum_{j=1}^{n} \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{t - t_j}{h}\right)^2\right],$$

where the t_j are the observed RTs and n is the number of observations. If some care is taken in the selection of h, the estimate can be shown to be uniformly consistent provided $g_k(t)$ is uniformly continuous (see Parzen, 1962). We can now use the known convergence properties (see, e.g., Roussas, 1973), of the estimate

$$\hat{G}_k(t; n) = \frac{1}{n} [\text{the number of } t_1, t_2, \ldots, t_n \leqslant t]$$

together with the g_k estimates to investigate the properties of the estimate

$$\hat{V}_k = \frac{\hat{g}_k(t; n_2)}{\hat{G}_{k-1}(t; n_1) - \hat{G}_k(t; n_2)} \tag{14}$$

First, for convenience, assume $n_1 = n_2 = n$. Next following the development of Rice & Rosenblatt (1976) of properties of various hazard function estimators we note that

$$\frac{\hat{g}_k(t; n)}{\hat{G}_{k-1}(t; n) - \hat{G}_k(t; n)}$$

$$= \frac{\hat{g}_k(t; n)}{G_{k-1}(t) - G_k(t) + [G_k(t) - \hat{G}_k(t; n)] - [G_{k-1}(t) - \hat{G}_{k-1}(t; n)]}$$

$$= \frac{\hat{g}_k(t; n)}{G_{k-1}(t) - G_k(t)} \sum_{j=0}^{\infty} (-1)^j \left\{ \frac{[G_k(t) - \hat{G}_k(t; n)] - [G_{k-1}(t) - \hat{G}_{k-1}(t; n)]}{G_{k-1}(t) - G_k(t)} \right\}^j$$

$$= \frac{\hat{g}_k(t; n)}{G_{k-1}(t) - G_k(t)} \left\{ 1 + \sum_{j=1}^{\infty} (-1)^j \left(\frac{[G_k(t) - \hat{G}_k(t; n)] - [G_{k-1}(t) - \hat{G}_{k-1}(t; n)]}{G_{k-1}(t) - G_k(t)} \right)^j \right\}.$$

Now we use the properties of the estimate $\hat{G}_j(t; n)$; specifically we know that

$$\sup | G_j(t) - \hat{G}_j(t; n)| = 0_p \left(\frac{1}{n^{1/2}} \right).$$

In other words, $n^{1/2}(\sup | G_j(t) - \hat{G}_j(t; n)|)$ is bounded for all n implying that $\hat{G}_j(t; n)$ converges in probability to $G_j(t)$ at a rate faster than $1/n^{1/2}$ converges to zero. Thus, incorporating this bound into the above expansion,

$$\frac{\hat{g}_k(t; n)}{\hat{G}_{k-1}(t; n) - \hat{G}_k(t; n)} = \frac{\hat{g}_k(t; n)}{G_{k-1}(t) - G_k(t)} \left[1 + 0_p \left(\frac{1}{n^{1/2}} \right) \right]$$

so the properties of the estimate (14) follow directly from the properties of the density estimate $\hat{g}_k(t; n)$.

In our first simulations all assumptions hold exactly. Our goal here is to get some idea as to the standard error or the variability in the slope of a regression line fit to the $\hat{V}_k$ vs t plot, when the assumptions of Theorem 1 are formally satisfied. This knowledge can then be used as a general guide in empirical applications. If the absolute value of the slope of a regression line fit to empirical data is much larger than those found in these simulations, then one may tentatively conclude that one or more of the assumptions of Theorem 1 do not hold for this particular set of data. Toward this end, it is desirable to make the manipulable details (e.g., parameter values) of the simulations conform as closely as possible to our expectations with regard to empirical data, and we have attempted to accomplish this.

The empirical application to be reported after the simulations utilizes data from a standard memory-scanning experiment which was reported by Townsend & Roos (1973). In this type of experiment, the classically quoted mean scan time for one item is 25 to 30 msec (e.g., Sternberg, 1966). Therefore, in our simulations we chose a mean of 25 msec for the ICT whose rate was to be recovered. Another parameter to be set was n, the number of trials. The Townsend *et al.* (1973) study used an n of about 250 so this should be the minimum value used in our simulations. We should expect, however, the variability in the slope of the regression line to decrease with increases in n. Thus if a larger n is used, the trend should be toward a narrower range of acceptable slopes and thus a more conservative test. If with a larger n the slope from the empirical application agrees with those from the simulations, then the assumptions of Theorem 1 gain support. So, in an attempt to make the results of these simulations generalize to memory-scanning (or visual-search) applications where more than 250 observations per condition are collected, the fairly large values of $n = 1000$ was selected. Some simulations with smaller

n were performed to gain a rough idea as to the effect of sample size on the variability of the regression line slope (and intercept).

In all simulations the base time was assumed uniformly distributed from 200 to 400 msec. The first ICT was exponentially distributed with a mean of 12 msec and the second ICT (to be recovered) had an exponential distribution with a mean of 25 msec. The results are contained in Table 1. There the slopes of the least-squares regression lines (fit to the $\hat{V}_k$ vs t plot) and the resulting estimates of the second ICT mean, $1/V_2$, are given. The "mean" estimate is taken as the inverse of the mean of all estimates generated by applying Theorem 1 at different possible values of t.

TABLE 1

Results of Computer Simulations in which All Assumptions are Fulfilled

Simulation number	$(n_1 \overset{n}{=} n_2 = n)$	$1/V_2$	Estimate of $1/V_2$	Slope of regression line
1	25	25	26.25	-9.1×10^{-5}
2	50	25	27.86	3.6×10^{-4}
3	100	25	14.6	2.4×10^{-5}
4	250	25	13.39	-6.9×10^{-5}
5	500	25	25.84	1.6×10^{-5}
6	1000	25	27.78	-2.4×10^{-5}
7	1000	25	25.00	-1.3×10^{-5}
8	1000	25	28.82	-4.2×10^{-5}
9	1000	25	23.47	3.5×10^{-5}
10	1000	25	22.88	-5.5×10^{-5}
11	2000	25	23.98	7.0×10^{-6}

Although there is some variability, overall the results appear quite good. Perhaps most surprising is the relatively small effect n seems to have on the slope of the regression line. Even with only 25 observations per condition, the slope is still on the same order of magnitude as when 1000 observations are generated. The larger slope for $n = 50$ indicates that there will probably be more variation for the smaller values of n but nevertheless, the results do appear promising. Somewhat more puzzling are the poor estimates of $1/V_2$ obtained in simulations 3 and 4. It may be that the method provides somewhat more variability in the intercept than in the slope of the regression line. It is also possible that taking the mean of all the estimates is not the optimal method of estimating V_2. An alternative approach, which appears to have some potential, is to integrate both sides of Theorem 1 and plot the result vs time. The result should be a linear function with slope V_2 since

$$\int_0^t V_k \, dx = \int_0^t \frac{g_k(x)}{G_{k-1}(x) - G_k(x)} \, dx$$

implies

$$\int_0^t \frac{g_k(x)}{G_{k-1}(x) - G_k(x)}\, dx = V_k t. \tag{15}$$

We will see more of this estimate somewhat later.

A perusal of Table 1 indicates that an acceptable decision rule (for the null hypothesis that the joint assumptions of pure insertion at the level of the distribution and exponentially distributed ICTs hold) is to reject if the absolute value of the regression line slope is greater than or equal to 1.0×10^{-4}, since almost all of the slopes from the simulations fall in this interval. This strategy is also consonant with an approach based on standard errors. The sample standard deviation of the slopes obtained from all the simulations in which the sample size is at least as large as in the Townsend *et al.* (1973) study (i.e., simulations 4 through 11) is 3.62×10^{-5}. Thus the acceptance boundaries, $\pm 1.0 \times 10^{-4}$, encompass a region approximately 2.75 standard deviations on either side of zero. This strategy, then, would seem to ensure a fairly small type I error.

What then of the type II error? Can models, other than those satisfying the assumptions of Theorem 1, generate slopes with absolute value less than our criterion of 1.0×10^{-4}?

To attempt an answer to this question, the distribution of the second ICT was changed from the exponential. All simulations were performed with an n of 1000. Table 2 presents

TABLE 2

Results of Computer Simulations in which Subtraction at the Level of the Distribution Holds but the 2nd ICT is Not Distributed Exponentially

Simulation number	Second ICT density	Parameter values	Slope of regression line
12	Gamma $f_2(t) = \frac{(Vt)^{m-1}}{(m-1)!} Ve^{-Vt};\ t > 0$	$m = 2$, $V = \frac{1}{12.5}$	2.24×10^{-4}
13	Gamma	$m = 5$, $V = \frac{1}{5}$	1.81×10^{-4}
14	Uniform $f_2(t) = \frac{1}{V},\ 0 \leqslant t \leqslant V$; $= 0$ otherwise	$V = 50$	3.83×10^{-4}
15	Weibull $f_2(t) = ap(pt)^{a-1} e^{-(pt)^a};\ t > 0$	$a = 1.5$, $p = \frac{1}{25}$	3.27×10^{-4}

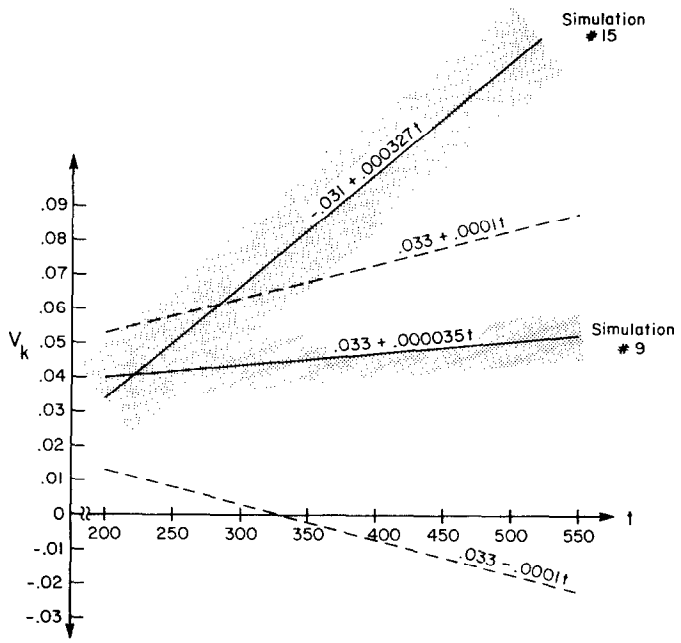

FIG. 4. Graphical representation of two typical Monte Carlo simulations. In simulation 9 the assumptions of Theorem 1 are exactly satisfied while simulation 15 represents a violation of those assumptions. The shaded regions reflect the variability of the data about the regression lines and the broken lines enclose the acceptance region for the null hypothesis of pure insertion at the distributional level and exponentially distributed ICTs.

the results.[4] A quick glance at the regression line slopes reveals that in all cases they fall in the rejection region. Thus for each of these simulations our decision would be that at least one of the assumptions of Theorem 1 had been violated. Our test, therefore, appears to have reasonable power.

An example of the decision rule we have adopted is illustrated in Fig. 4 for simulations 9 and 15. In simulation 9 (see Table 1) the assumptions of Theorem 1 are exactly satisfied and the resulting regression line slope appears typical of such simulations. Simulation 15 (see Table 2) represents a violation of the assumptions. The tubes enclosing the regression lines in Fig. 4 reflect the variability in the points about the regression line. The edges of the tubes are at plus and minus the mean absolute deviation of the points about the lines. Finally, the broken lines in the figure (with slopes $\pm 1.0 \times 10^{-4}$) define the acceptance region for the null hypothesis of pure insertion at the distributional level and exponentially distributed ICTs. As can be seen the regression line for simulation 15 is clearly outside of the acceptance region.

As an empirical application of Theorem 1 we utilized data from a standard memory-scanning experiment reported by Townsend *et al.* (1973). The memory set contained

[4] These densities all have increasing hazard functions. Densities with decreasing hazard functions have extremely high tails and for this reason appear unreasonable as models of human RT.

from one to five alphanumeric characters. Each of three subjects participated in 500 trials with each memory set size, 250 of these entailed "yes" responses and 250 "no" responses. The analyses reported below deal only with the target absent or "no" data. These should be sufficient to test the double assumption of pure insertion at the distributional level and exponentially distributed ICT's. The question of whether the method can discriminate between self-terminating and exhaustive search strategies will be a target for future efforts.

Table 3 contains the results of applying Theorem 1 to these data. The stars on some rows indicate that the sample RT variances were larger for set size $k - 1$ than for set size k. Of course, if these were population variances this decrease would immediately falsify pure insertion at the level of the distribution.

TABLE 3

Results of Applying Theorem 1 to the Data of a Memory-Scanning Experiment[a]

	Slope of regression line	Estimate of $1/V$	Estimate of $1/V$ from Eq. (15) (after integration)	$\overline{RT}_k - \overline{RT}_{k-1}$
Subject 1				
* ICT 2	3.56×10^{-4}	−31.25	−11.85	15
ICT 3	-7.7×10^{-5}	21.28	17.59	7
* ICT 4	3.5×10^{-5}	32.26	37.64	28
* ICT 5	-2.7×10^{-5}	35.71	28.69	43
Subject 2				
* ICT 2	-9×10^{-6}	71.43	59.72	2
* ICT 3	-7.5×10^{-5}	125.0	84.76	17
ICT 4	$<6 \times 10^{-7}$	45.45	43.19	25
* ICT 5	-7.6×10^{-5}	20.41	15.24	16
Subject 3				
ICT 2	1.49×10^{-4}	20.41	19.3	29
* ICT 3	-1.48×10^{-4}	35.71	17.29	16
ICT 4	8×10^{-6}	∞	286.48	31
ICT 5	2.8×10^{-5}	34.48	22.5	35

[a] That is, Townsend and Roos (1973).

An examination of the slopes in Table 3 shows that in all but three cases the slopes of the regression lines fell in the interval $\pm 1.0 \times 10^{-4}$, which we earlier tentatively agreed upon as a conservative (because of the larger sample size used in the simulations) acceptance region for the hypothesis of pure insertion at the level of the distribution and exponentially distributed ICTs.[5] Thus, each of these regression lines rises or falls less

[5] The fits of the regression lines were found to be adversely affected by outliers from the extreme tails of the RT densities. For this reason extremely long RTs were not utilized in the regression analysis. However, the discarded data always comprised less than 1 % of the RTs.

steeply than the broken lines of Fig. 4. An analysis based on the estimated standard error yields similar conclusions. We earlier calculated that the standard deviation of the regression line slopes in the simulated data was 3.62×10^{-5}. Half of the slopes reported in Table 3 therefore lie within one of these standard deviations from the origin. Three others lie within 2.2 standard deviations and these nine cases represent all of the slopes falling in the interval $\pm 1.0 \times 10^{-4}$ and so they are the ones for which we tentatively accepted the assumptions of Theorem 1. The three slopes falling outside this acceptance region all lie more than four standard deviations from the origin.

An illustration of what the worst case and best case actually look like is given in Figs. 5 and 6. In the first of these a plot of $\hat{g}_2(t)/[\hat{G}_1(t) - \hat{G}_2(t)]$ vs t is given for subject 1. The resulting regression line which was fit to these data has a slope of 3.56×10^{-4} and was thus the steepest of all the regression lines fit. The regression line fit to the subject 2, ICT 4 data of Fig. 6 was about 600 times shallower, having a slope of less than 6×10^{-7}. As can be seen, although the subject 1, ICT 2 data of Fig. 5 exhibits much more variability, both sets of data appear quite flat and do not, for instance, appear to display consistent nonlinear trends.

These are surprising yet propitious results. They are surprising when we consider that in over half of the cases (7 out of 12) the sample variances were smaller for set size k than for $k - 1$. They are propitious in the sense that they suggest there may be more order in the complete RT distribution data (as opposed to means alone) than is commonly believed. The exponential distribution is often derided as being too simplistic, but the present results caution against dismissing it without further hearing.

The estimates of $1/V$ are in most cases reasonable, although it appears there may be more variability here than in the slopes. There appears little in Table 3 to favor either of the two estimates over the other, but it is expected that the estimate obtained after an integration has been performed (see Eq. (15)) will be more stable, since integration is itself a smoothing operation.

In any event, somewhat more troubling is a comparison of the estimates obtained from, say, Eq. (15), with the more naive estimates of $1/V$ obtained by subtracting the appropriate means (as in the rightmost column of Table 3). While these estimates are in some cases quite close in a great many they are not. There are at least several possible reasons for this discrepancy. First recall that the simulations reported in Table 1 indicated that with increases in n the regression slopes may converge more quickly than the intercepts (i.e., the $1/V_2$ estimates). Specifically note that for simulation 4 the slope is well within the acceptance region while the V_2 estimate is poor. Simulation 4 utilized an n of 250 which is the same as in the Townsend and Roos study. Thus it may be that while 250 trials per data point is enough to test the null hypothesis of pure insertion at the level of the distribution and exponentially distributed ICTs it may not be enough to use Theorem 1 to reliably estimate the ICT rate, V_k. We can see from Table 1 that an n of 1000 should easily be large enough for this purpose and we have an indication that an n of 500 may be adequate. However, a satisfactory answer to this question necessitates performing many more simulations.

Another possibility we must not forget is that $\overline{\mathrm{RT}}_k - \overline{\mathrm{RT}}_{k-1}$ is, of course, a random variable. While the variance of the statistic $\overline{\mathrm{RT}}_k$ is not normally considered prohibitive,

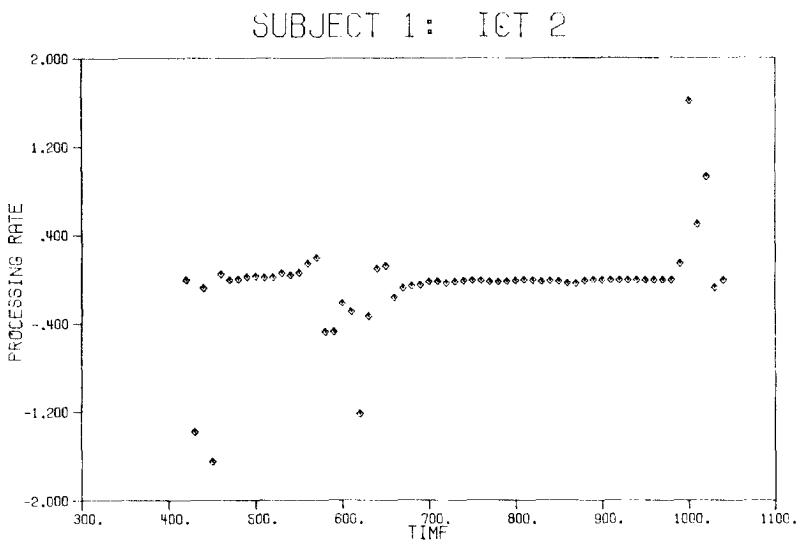

FIG. 5. A plot of $\hat{g}_2(t)/[\hat{G}_1(t) - \hat{G}_2(t)]$ vs t for subject 1. A flat linear plot supports the double assumption of pure insertion at the distributional level and that the inserted stage duration has an exponential distribution. The regression line fit to this data, with a slope of 3.56×10^{-4}, was the steepest of all empirical applications.

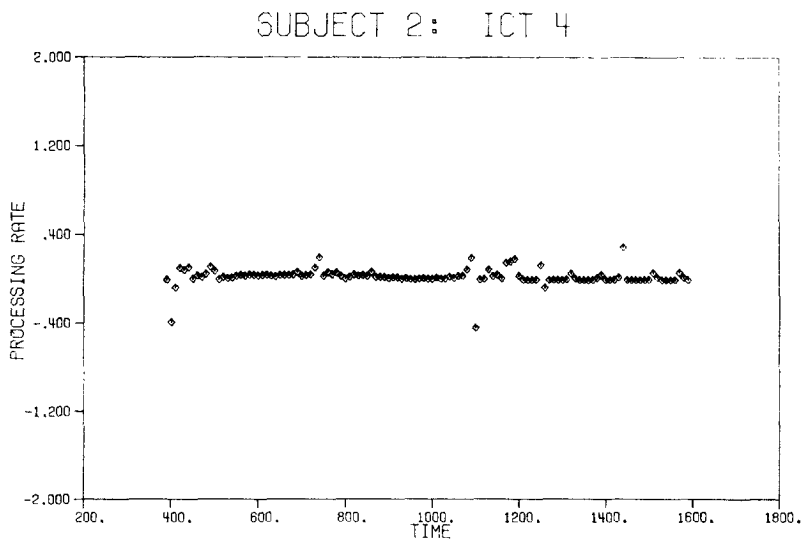

FIG. 6. A plot of $\hat{g}_4(t)/[\hat{G}_3(t) - \hat{G}_4(t)]$ vs t for subject 2. The regression line fit to this data had a slope of less than 6×10^{-7} and was the shallowest of all the regression lines fit to empirical data.

it must be kept in mind that the variance of $\overline{RT}_k - \overline{RT}_{k-1}$ is roughly twice as large. As a rough indicant, the average sample variance for the subjects in the Townsend *et al.* (1973) study was about 25,000 msec^2. Thus the variance of the estimate $\overline{RT}_k - \overline{RT}_{k-1}$ is roughly

$$\text{var}(\overline{RT}_k - \overline{RT}_{k-1}) \doteq \frac{2(25000)}{250} \text{ msec}^2 = 200 \text{ msec}^2$$

and hence one standard deviation is just over 14 msec. We thus see that over half of the estimates obtained from Eq. (15) lie within one standard deviation of the estimate $\overline{RT}_k - \overline{RT}_{k-1}$.

All in all, the application, though tentative, seems a success. The double assumption of pure insertion at the level of the distribution and exponentially distributed ICTs received support in a memory scanning task. What is needed now are more tests of these assumptions in more varied experimental situations and that will be a target for future efforts.

REFERENCES

ASHBY, F. G. Pattern matching: Self-terminating or exhaustive processing? Unpublished master's thesis, Purdue University, 1976.

ASHBY, F. G. Assuming selective influence: An alternative to additive factor theory. Paper presented at Twelfth Annual Mathematical Psychology Meeting, Brown University, August 1979.

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

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

COX, D. R., & MILLER, H. D. *The theory of stochastic processes*. London: Methuen, 1965.

DONDERS, F. C. On the speed of mental processes. Translated by W. G. Koster. In W. G. Koster (Ed.), *Attention and performance II*. Amsterdam: North-Holland, 1969.

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

HUNT, E. Mechanics of verbal ability. *Psychological Review*, 1978, **85**, 109–130.

MASSEY, F. J., JR. The Kolmogorov–Smirnov test for goodness of fit. *American Statistical Association Journal*, 1951, **46**, 68–78.

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

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

MCGILL, W. J., & GIBBON, J. The general gamma distribution and reaction times. *Journal of Mathematical Psychology*, 1965, **2**, 1–18.

MCGILLEM, C. D., & COOPER, G. R. *Continuous and discrete signal and system analysis*. New York: Holt, Rinehart & Winston, 1974.

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

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

POSNER, M., & MITCHELL, R. Chronometric analysis of classification. *Psychological Review*, 1967, **74**, 392–409.

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

RESTLE, F. *Psychology of judgment and choice.* New York: Wiley, 1961.

RICE, J., & ROSENBLATT, M. Estimation of the log survivor function and hazard function. *Sankhyā: The Indian Journal of Statistics, Series A*, 1976, **38**, 60–78.

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

ROUSSAS, G. G. *A first course in mathematical statistics.* Reading, Mass.: Addison–Wesley, 1973.

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

SHIFFRIN, R., & THOMPSON, M. Moments of random variables on regenerative random processes. Paper presented at Twelfth Annual Mathematical Psychology Meeting, Brown University, August 1979.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

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

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

TAYLOR, D. A. Stage analysis of reaction time. *Psychological Bulletin*, 1976, **83**, 161–191.

TAYLOR, D. H. Latency models for reaction time distributions. *Psychometrika*, 1965, **30**, 157–163.

THEIOS, J. Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation, Vol. 7.* New York: Academic Press, 1973.

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

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

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

TOWNSEND, J. T., & ASHBY, F. G. *Stochastic Modeling of Elementary Psychological Processes.* In press.

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

RECEIVED: June 5, 1979