The ex-Wald distribution as a descriptive model of response times

WOLFGANG SCHWARZ
University of Nijmegen, Nijmegen, The Netherlands

We propose a new quantitative model of response times (RTs) that combines some advantages of substantive, process-oriented models and descriptive, statistically oriented accounts. The ex-Wald model assumes that RT may be represented as a convolution of an exponential and a Wald-distributed random variable. The model accounts well for the skew, shape, and hazard function of typical RT distributions. The model is based on two broad information-processing concepts: (1) a data-driven processing rate describing the speed of information accumulation, and (2) strategic response criterion setting. These concepts allow for principled expectations about how experimental factors such as stimulus saliency or response probability might influence RT on a distributional level. We present a factorial experiment involving mental digit comparisons to illustrate the application of the model, and to explain how substantive hypotheses about selective factor effects can be tested via likelihood ratio tests.

Attempts to quantitatively account for response time (RT) data may roughly be classified into either of two broad categories: one of substantive, process-oriented models and one of descriptive, statistically oriented accounts. Process-oriented RT models are based on detailed substantive assumptions concerning the number, nature, and temporal scheduling of processing stages; often, they specify the dynamics of information accumulation over time (see, e.g., Ratcliff, 1978; Ratcliff & Murdock, 1976; Ratcliff & Rouder, 1998; Schwarz, 1993, 1994; P. L. Smith, 1995). The ability of these models to fit empirical data sets is an important boundary condition, but their ultimate success is more typically seen to rest on their ability to explain complex data patterns from different conditions, or different experiments, in a conceptually coherent way. A typical example is Ratcliff's (1978) detailed account of set size effects in recognition memory in terms of a race of parallel diffusion processes, a model that also explains effects of constant versus varied mapping (Strayer & Kramer, 1994). Quite generally, such models give detailed accounts not only of mean RT, but also of its complete distribution, and they are able to address the relation of mean correct latency and mean error latency, and of how both covary with response accuracy. The price at which those advantages are bought is that most process-oriented models are formally and computationally demanding, and their implementation in specific experimental contexts often requires flexible adjustments and nonstandard elaborations. For example, in the context of diffusion models of RT, the incorporation of psychologically important model features such as

state- or time-dependent drift rates (Heath, 1992; P. L. Smith, 1995), probabilistic randomizations of response barriers (Ratcliff & Rouder, 1998), or transitions from reflecting to absorbing response barriers (Schwarz & Stein, 1998) requires a firm background in formal diffusion theory and does not easily lead to readily applied standard tools for data analysis. These factors are very likely a main reason why, despite their undisputed advantages, detailed process-oriented models are perhaps more often discussed than actually applied to data. As an example, Ratcliff's (1978) diffusion model of memory search has been extremely influential in recognition memory research, but it is also true that only relatively few contributors from outside Ratcliff's lab have ever applied this model to their own data.

In contrast to substantive, process-oriented models, other accounts simply aim at an accurate description and an economical summary of empirical RT distributions. From a formal and computational point of view, these models are often less demanding, and their application to data sets is a standard routine (cf. Cousineau & Larochelle, 1997; Dawson, 1988; Heathcote, 1996). A prominent example of this class of descriptive RT models is the "ex-Gaussian" model, originally described by Hohle (1965). It assumes that RT is distributed as is

$$T = D + M,$$

where D is a normally distributed random variable and M is an exponential random variable, independent of D. The normal component D is sometimes thought to arise from a large number of serially organized stages; if their durations roughly conform to the assumptions of the central limit theorem (Feller, 1966), then their overall duration is approximated by a normal random variable. Given that empirical RT distributions are not symmetrical, the normal component is convolved with an independent ex-

Correspondence concerning this article should be addressed to W. Schwarz, University of Nijmegen, NICI, Postbus 9104, 6500 HE Nijmegen, The Netherlands (e-mail: schwarz@nici.kun.nl).

ponential component M, which generates the required skew. The assumption of an additive exponential RT component is supported by independent evidence from (1) approximations to the tail of the log RT survivor function or its derivative the hazard function (McGill & Gibbon, 1965; Ueno, 1992), (2) RT deconvolution approaches (Burbeck & Luce, 1982; Luce, 1986), and (3) statistical procedures specifically designed to test for additive exponential RT components (Ashby, 1982; Ashby & Townsend, 1980).

Initially attempts were made to interpret D and M as the durations of two serially organized stages and to identify them with specific component processes such as "perception and decision" versus "organization and execution of the motor response" (Hohle, 1965, p. 384). However, for over 35 years, it has proved difficult to reach an agreement as to which of the two components represents which processing stage. For example, McGill and Gibbon (1965) interpreted the exponential component as a residual motor latency, in direct conflict with the original interpretation of Hohle, who took it to represent the perception and decision latency. As a consequence of these stage identification problems, part of the more recent research has abandoned substantive interpretations of the ex-Gaussian model altogether. Thus, the ex-Gaussian model is often fitted to data without any restrictions on its parameters, and without specific expectations about which experimental manipulation might affect which model parameter(s). This cautious attitude is perhaps best illustrated by a quote from Heathcote, Popiel, and Mewhort (1991, pp. 346–347; for a very similar position, see Blanco & Alvarez, 1994, p. 83): "We treat the ex-Gaussian as a descriptive first-order account of response latency. Because we do not propose the ex-Gaussian as a model of cognitive processes, we do not need a cognitive attribution for its parameters." Consistent with this view, the ex-Gaussian model is sometimes used as a purely descriptive tool for preliminary data reduction (as, e.g., in Blanco & Alvarez, 1994; Ratcliff, 1978; D. G. Smith & Mewhort, 1998). Other applications (Balota & Spieler, 1999; Heathcote et al., 1991; Plourde & Besner, 1997) specifically stress the use of the ex-Gaussian model to statistically evaluate factor effects beyond the level of the mean, such as differences in variance or skew across conditions.

Despite this modest attitude, several problems with the ex-Gaussian model of RT remain. First, contrary to the cautious position of Heathcote et al. (1991), in many recent contributions the ex-Gaussian model has been used not simply as a purely descriptive, data-reductionist tool to describe RT effects; rather, researchers have clearly seen a "need [for] a cognitive attribution for its parameters," as one might expect of models used in cognitive psychology, even when they are meant to be first approximations. As an example, Rohrer and Wixted (1994) fit the ex-Gaussian model to free recall latencies; they interpreted its normal component as an early search-initiation stage, with the exponential component reflecting an ongoing memory search process. As another example, Balota and Spieler (1999) discuss an interpretation of the ex-Gaussian model in which the component D is related to stimulus-driven automatic (nonanalytic) processes, with the exponential component reflecting more central, attention demanding (analytic) processes. Many other substantive interpretations of the ex-Gaussian model have been described—see, for example, Hockley and Murdock (1987), Wixted and Rohrer (1993), or Juhel (1993).

Second, if a major aim of the ex-Gaussian analysis is simply to compare higher central moments of empirical RT distributions such as the standard deviation or the standardized skew across conditions, then statistical standard methodology with an associated established estimation and sampling theory (cf. Stuart & Ord, 1987) seems preferable to strong ad hoc models whose statistical properties are not very well understood. It is true that, as is sometimes argued, the standard error of moment estimates increases with the order of the moments, but the ex-Gaussian distribution is of course no exception to this general rule.

Third, RT distributions have repeatedly been reported to exhibit peaked hazard functions under many conditions: they first increase and then decrease toward a non-zero asymptotic level (Maddox, Ashby, & Gottlob, 1998; P. L. Smith, 1995; Ueno, 1992). It is often assumed that this behavior is due to a peaked hazard function of the premotor component of RT because it is also observed when the residual motor time is deconvolved out of the RT distribution (Burbeck & Luce, 1982; Luce, 1986). However, the hazard function of the ex-Gaussian distribution is strictly increasing, because its normal component has a strictly increasing hazard function, and its exponential component, a constant hazard function (cf. Luce, 1986). Thus, the ex-Gaussian model lacks an important functional property characteristic of many empirical RT distributions.

Finally, although (as summarized above) there is independent evidence that empirical RT distributions do show an additive exponential component, the assumption of a normal component seems less convincing for several reasons. First, as Luce (1986) pointed out, reference to the central limit theorem is vague; if just a few steps of unequal variances are involved, the normal approximation is likely to be inaccurate. Sometimes the normal component is also justified by the reproductive property of the normal distribution (e.g., Rohrer & Wixted, 1994); it should be noted, though, that this is a property that the normal distribution shares with any other stable distribution (cf. Feller, 1966). Furthermore, conceptually the normal distribution is not completely adequate for dealing with real time latency data, which is also suggested by the fact that it extends over the entire real line, whereas durations are necessarily nonnegative. It is sometimes argued that this represents no real

problem because the parameters are usually chosen so that the probability $P(D < 0)$ is kept small. However, this argument is potentially misleading, because in typical RT research one deals not with a single realization, but with many replications of D. Thus, even if $P(D < 0)$ is small for any single realization, the probability that negative values of D arise increases considerably for the many replications of D obtained in typical RT experiments.

The aim of this paper is to propose a compromise between substantive, process-oriented models and descriptive, statistically oriented accounts that retains at least some advantages of both approaches; a related aim is to describe techniques to apply this new model to example data, and to evaluate its statistical fit. Our starting point is the following consideration: Given that we are prepared to accept an essentially descriptive model of RT that is parsimonious and sufficiently flexible to account for empirical latency distributions, then why should we not use component distributions that at least roughly capture substantive concepts of the cognitive mechanisms that we try to describe?

A basic concept to explain human latency data rests on the assumption that the decision to initiate and to execute a specific overt response does not arise holistically, in an all-or-none fashion, but is rather preceded by a stage during which response-related information gradually accumulates over time. One of the simplest nontrivial formal descriptions of gradual accumulation of information consists of diffusion models with independent increments. The idea here is to model the amount of sampled information as a stochastic process characterized by a mean information gain per time unit, called the *drift rate* of the process. The amount of evidence required before a response is initiated and executed may then be represented as an evidence criterion (or response barrier) at which the process terminates when it reaches that critical level for the first time. Over the last 40 years, many variants of this basic conceptualization have been developed and successfully applied by many authors (Emerson, 1970; Laming, 1968; Link, 1992; Ratcliff, 1978; Ratcliff & Rouder, 1998; Schwarz, 1990, 1992; P. L. Smith, 1995; Stone, 1960; Strayer & Kramer, 1994). In the following section, we will describe a relatively simple version of this general concept that might help researchers to overcome some of the problems with the ex-Gaussian model reviewed above.

The Ex-Wald Model

The ex-Wald model is also a member of the serial (additive) family $T = D + M$ of RT models, and it also assumes the random variable M to be exponentially distributed, with rate γ, independent of D. The difference with the ex-Gaussian model is that it assumes that the decision latency, D, follows a Wald distribution.[1] Formally, any Wald-distributed random variable may be represented as the time required in a time- and space-homogeneous (i.e., Wiener) diffusion process to reach for the first time

a certain fixed level. It thus is by definition nonnegative, and its hazard function may be shown to be first increasing and then decreasing (cf. Luce, 1986). In the present context, we interpret the diffusion process as a process of accumulating noisy partial information over time. The speed of this accumulation process is governed by the drift rate, μ, the mean gain of information per time unit. The fixed level of the process that needs to be reached before a response is initiated and executed is interpreted as an evidence criterion, a. The exponential component M is thought to summarize all processes following (and possibly preceding) the decision stage, D (see Figure 1).

Two factors then determine the distribution of D: the drift rate, μ, of the diffusion process representing the data-driven component of the task, and the distance, a, of the evidence criterion from the starting point of the sampling process representing strategic criterion-setting (cf. Strayer & Kramer, 1994). In the ex-Wald model, the interpretation of these two quantities is conceptually similar to the well-known parameters d' and c of signal detection theory (see Macmillan & Creelman, 1991): like c in signal detection theory, the parameter a is considered to be a response criterion that the subject strategically adjusts to the current experimental conditions. For example, in a go/no-go RT experiment, the accumulated evidence required for a subject to initiate a response will be expected to increase as the a priori probability of a go signal decreases. More generally, any partial preinformation about the stimulus to be presented might be expected to lead to a corresponding adjustment of the evidence criterion. This is illustrated in Figure 1 for two criterion levels, $a_1 < a_2$.

In contrast, the drift rate μ, like the parameter d' in signal detection theory, is considered a parameter not under voluntary control of the subject; rather, it is determined by the experimental conditions, and it summarizes the average amount of response-relevant information that can be extracted during a given time unit. For example, in a go/no-go RT experiment, the rate with which evidence is accumulated will be expected to depend on the perceptual or cognitive saliency of the signal presented. This is illustrated in Figure 1 for two levels $\mu_1 > \mu_2$ of the drift rate. Thus, increases of μ and decreases of a will both decrease RT, but for quite different reasons, just as a high hit rate in signal detection theory may be due to any combination of a high sensitivity and/or a lax response criterion. A further parallel with the parameters c and d' in signal detection theory is that in most experiments without trial-to-trial manipulations of a (e.g., by cues), the evidence criterion would be expected to be blockwise constant, whereas the drift rate μ may vary within a block according to the stimuli presented.

The most important formal properties of the ex-Wald model are derived in the Appendix. In addition, Figure 2 summarizes and illustrates the shapes of the Wald and the exponential component distributions, and that of their
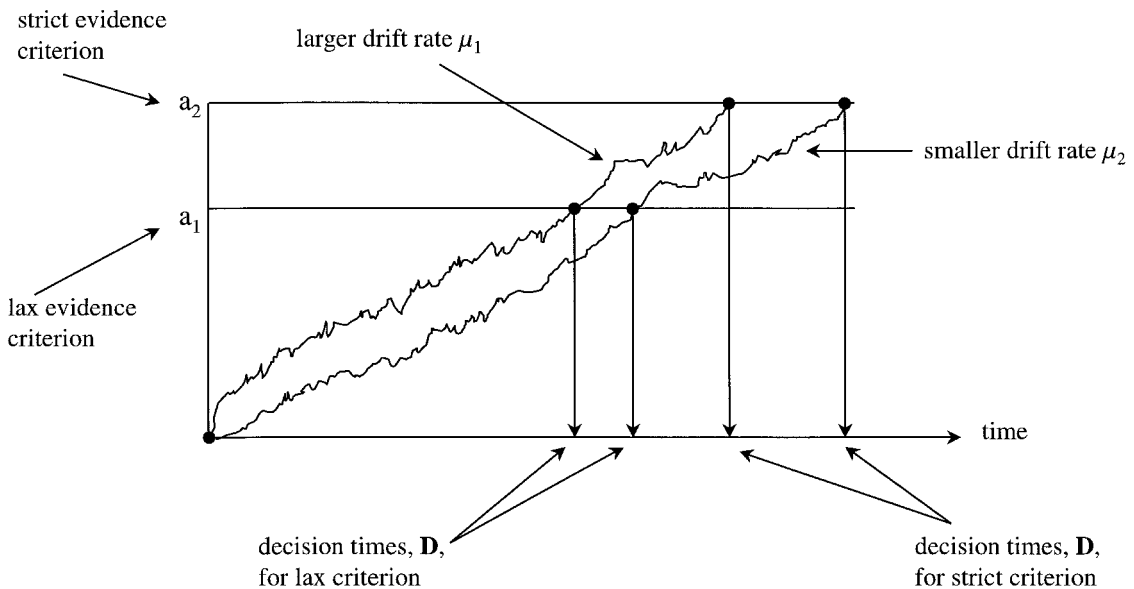
Figure 1. Two typical sample paths representing stochastic accumulation of information over time. On the average, less time will be needed to reach the evidence criterion for a larger drift rate ($\mu_1 > \mu_2$), and also for a lax evidence criterion ($a_1 < a_2$). The ex-Wald model assumes that overall RT is the sum of the first-passage time to the evidence criterion plus an independent exponential random variable.

associated convolution. It also illustrates how the parameters of the component distributions influence the location and shape of their convolution. Clearly, when compared with detailed process-oriented models, the information-processing notions of the ex-Wald model seem quite broad. However, as mentioned above, the ex-Wald model is proposed here, not to replace those more detailed, but computationally more demanding attempts, but rather as a compromise to combine at least some advantages of concept-oriented modeling with the formal and computational simplicity of, for example, the ex-Gaussian model.

Application of the Ex-Wald Model to Mental Digit Comparison

As an illustration, we applied the ex-Wald model to a go/no-go RT experiment involving mental digit comparisons (cf. Dehaene, 1997). Specifically, we looked at the effect of numerical distance under different levels of the a priori probability of go trials. It is well known that in two-choice tasks, mean RT decreases with increasing numerical distance, but it is an open question how this effect of numerical distance might interact with the a priori go probability in a go/no-go task. For example, from the geometry in Figure 1, it would be predicted that the difference in RT between conditions with different drift rates should be larger for a stricter evidence criterion (cf. Strayer & Kramer, 1994). If the a priori go probability selectively influences the evidence criterion, a, and the numerical distance selectively influences the drift rate, μ, then the distance effect should be larger for smaller a priori go probabilities. However, our main aim with this illustration is not to present a new original contribution to mental comparisons but to describe relevant statistical procedures for the application of the ex-Wald model.

In each trial of a go/no-go digit comparison task, the participant is presented with a single test digit. A buttonpress was required (go trial) if the test digit was numerically larger than the fixed standard of 5; if the test digit was smaller than 5, participants had to withhold a response (no-go trial). Two factors were varied orthogonally. The go stimulus could be either the test digit 6 or 9, corresponding to a numerical distance to the standard of 1 versus 4. This factor was varied randomly within blocks. Second, in different sessions, subjects were informed that the a priori probability, p_{go}, of a go trial could be either .50 or .75. The expectation was, first, that just as with larger/smaller judgments in choice RT tasks, mean RT would decrease with increases in the numerical distance to the standard; thus, mean RT was expected to be shorter for the test digit 9 than for the test digit 6. Second, we expected mean RT in go trials to be shorter for the p_{go} = .75 condition than for the p_{go} = .50 condition.

Essentially four distributions were then obtained in our experiment: those for the test digit 6 and the test digit 9 under each of the two levels of p_{go}. In a first application, the ex-Wald model was fit without any restrictions on its parameters to each of these four distributions to see (1) how well it would account for the distribution of the observed latencies, and (2) which of its parameters would best reflect the variations of the two experimental factors.

Second, a more specific assumption within the framework of the ex-Wald model would be that variations of p_{go} will selectively influence the evidence criterion, a,
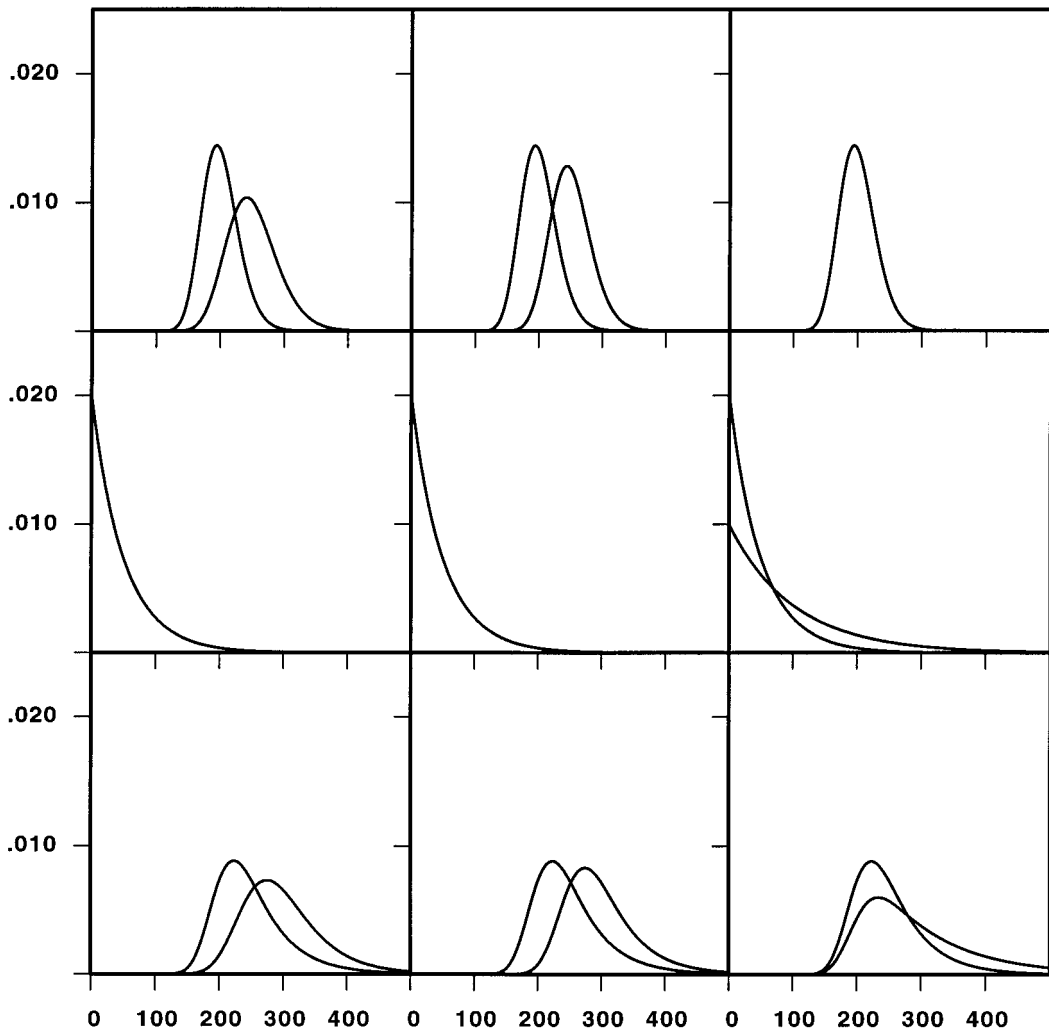
Figure 2. The top row of panels shows Wald, the middle row shows exponential component densities, and the bottom row shows their convolutions. The three vertical columns of panels illustrate the effects of selective parameter variations, starting from a constant baseline ($a = 100$, $\mu = 0.50$, $\sigma = 1$, $\gamma = 1/50$). The leftmost column of panels illustrates the influence of differences of the drift rate ($\mu = 0.40$ vs. $\mu = 0.50$) of the Wald density (the two densities in the top left panel) on the resulting convolutions (the two densities in the bottom left panel). The middle column of panels illustrates the influence of differences of the evidence criterion ($a = 100$ vs. $a = 125$) of the Wald density (the two densities in the top middle panel) on the resulting convolution (the two densities in the bottom middle panel). The rightmost column of panels illustrates the influence of differences of the exponential rate parameter ($\gamma = 1/50$ vs. $\gamma = 1/100$) of the exponential density (the two densities in the middle right panel) on the resulting convolution (the two densities in the bottom right panel). Note that all three parameter variations yield a main effect equal to 50 msec.

while variations of the numerical distance (1 vs. 4) of the go stimulus to the standard will selectively influence the drift rate, μ. Clearly, this model is more specific than (i.e., is a special case of) the unrestricted model. Therefore, its fit will necessarily be worse than the fit of the unrestricted model; the question, however, is whether this difference is statistically significant. Below, we will describe how this hypothesis may be addressed via likelihood ratio tests.

Method

Participants. Two paid male right-handed students, 18 and 19 years of age, participated in five sessions of about 1 h.

Stimuli and Apparatus. The digits 2, 3, 6, and 9 were used as stimuli. The size of the digits was 1.2×1.0 cm; the viewing distance was approximately 60 cm. They were displayed on a 60-Hz video monitor, the display timing being synchronized with the video refresh cycle.

Table 1
Mean RTs and RT Standard Deviations for Each Combination
of p_{go} = .50, .75, and Numerical Distance = 1, 4

p_{go}	Distance 1		Distance 4	
	M	SD	M	SD
.50	360	62	330	53
.75	324	58	302	52

Note—Mean RTs for p_{go} = .50 and p_{go} = .75 are based on n = 400 and n = 600 individual RTs, respectively.

Response latencies were recorded to the nearest millisecond, using an external response keyboard connected to the computer's parallel port.

Procedure. Each trial started with the presentation of the fixation point for a random duration of 150–300 msec, which was then followed by presentation of a single digit (2, 3, 6, or 9). The task of the participants was to push the button of the response keyboard with the right index finger if the number was numerically larger than 5 (digits 6, 9) but to withhold a response if it was smaller than 5 (digits 2, 3). Thus, for the go digit 6, 9 the factor *numerical distance* to the fixed standard was equal to 1, 4, respectively. The no-go digits 2, 3 were chosen to prevent participants from focusing their attention on single discriminative form features like, for example, the vertical strokes that would have signaled a no-go trial had we used other digits such as, for example, 1, 4. The display was terminated either by a response (go trials) or else after 1,000 msec (no-go trials). The next trial started 1,200 msec later. No feedback was given for correct responses; false alarms were followed by an acoustic error message. Participants were instructed to respond as fast as possible on go trials but to avoid false alarms. Blocks were separated by breaks which could be terminated by the subject after a minimum duration of 30 sec. During the break the number of false alarms and the mean RT of the preceding block was displayed on the screen, together with the mean RT of the fastest block so far.

Both subjects participated in five sessions of 22 blocks on 5 consecutive days. A single block consisted of four warm-up trials, followed by 40 regular trials in a random order generated by the computer, subject to the boundary condition that for p_{go} = .50 within each block, each digit (2, 3, 6, 9) occurred exactly 10 times, whereas for p_{go} = .75, the digits 2, 3 occurred exactly five times each, and the digits 6, 9 occurred exactly 15 times each. On day one, considered practice, Participant A did 11 blocks with p_{go} = .50, followed by 11 blocks with p_{go} = .75. On days 2 and 5 he then did 22 blocks with p_{go} = .75, and on Days 3 and 4, 22 blocks with p_{go} = .50. Participant B worked under the same schedule but with the values of p_{go} interchanged. Both participants were always informed about the current value of p_{go}. The last 20 blocks of each of the last four sessions entered into the final data analysis. Thus, for both participants, 400 and 600 replications were obtained for each go stimulus

(6, 9) under the condition of p_{go} = .50 and .75, respectively.

Results

Since our main aim is to illustrate the application of the ex-Wald model, we focus in the following on the data of Participant A; all important observations and conclusions were confirmed by the data of Participant B.

Response times. Mean RTs for each combination of p_{go} = .50, .75 and numerical distance = 1, 4 are shown in Table 1 (see also Figure 3). In a preliminary analysis, the data were evaluated statistically by a within-subjects two-factor ANOVA based on the 40 block means obtained under each of the 2 × 2 factor level combinations. Clearly, these results permit no generalization to hypothetical subject populations.

The variation of the a priori probability of a go trial yielded a large and significant main effect of 32 msec on RT [$F(1,156)$ = 162.53, MS_e = 255.90, p < .001]: responses were significantly faster for p_{go} = .75 (mean RT, 313 msec) than for p_{go} = .50 (mean RT, 345 msec). Similarly, numerical distance showed a significant main effect of 26 msec [$F(1,156)$ = 414.95, MS_e = 255.90, p < .001], with mean RTs of 342 and 316 msec for the distances of 1 (digit 6) and 4 (digit 9), respectively. Finally, the distance effect was slightly larger (30 msec) for the p_{go} = .50 condition than for the p_{go} = .75 condition (22 msec), but this difference was too small for the interaction to be significant [$F(1,156)$ = 2.20, MS_e = 255.90, p = .140].

Error rates. All together, 18 false alarms were observed. For the condition of p_{go} = .50, four errors occurred for each no-go stimulus (2, 3) in 400 trials, corresponding to error rates of 1%. For p_{go} = .75, in 200 trials, four false alarms were observed for the digit 2, and six for the digit 3, corresponding to error rates of 2% and 3%. No miss (i.e., a failure to respond to a go stimulus) was observed in any of the four conditions.

Unrestricted Maximum Likelihood Fit
of the Ex-Wald Model

Let <ij> denote the 2 × 2 experimental conditions, where the index i = 1, 2 refers to the factor levels p_{go} = .50, .75, respectively, and the index j = 1, 2 refers to the numerical distance of 1, 4, respectively. Maximum like-

Table 2
Unrestricted Ex-Wald Parameter Estimates and
Goodness-of-Fit for Each Factor Level Combination

p_{go}	Distance	$\hat{a}$	$\hat{\mu}$	$1/\hat{\gamma}$	χ^2	df	p
.50	1	108	0.320	22	8.86	8	0.354
.50	4	111	0.356	18	11.18	6	0.083
.75	1	97	0.321	22	3.92	8	0.864
.75	4	98	0.348	20	15.91	7	0.026
Total					39.87	29	0.086

Note—Bin width of categories used for χ^2 test was 25 msec. When necessary, categories were combined so that each expected frequency exceeded 5. Degrees of freedom result as the number of categories minus one minus the number parameters (three) estimated, $df = n_{cat} - 4$.
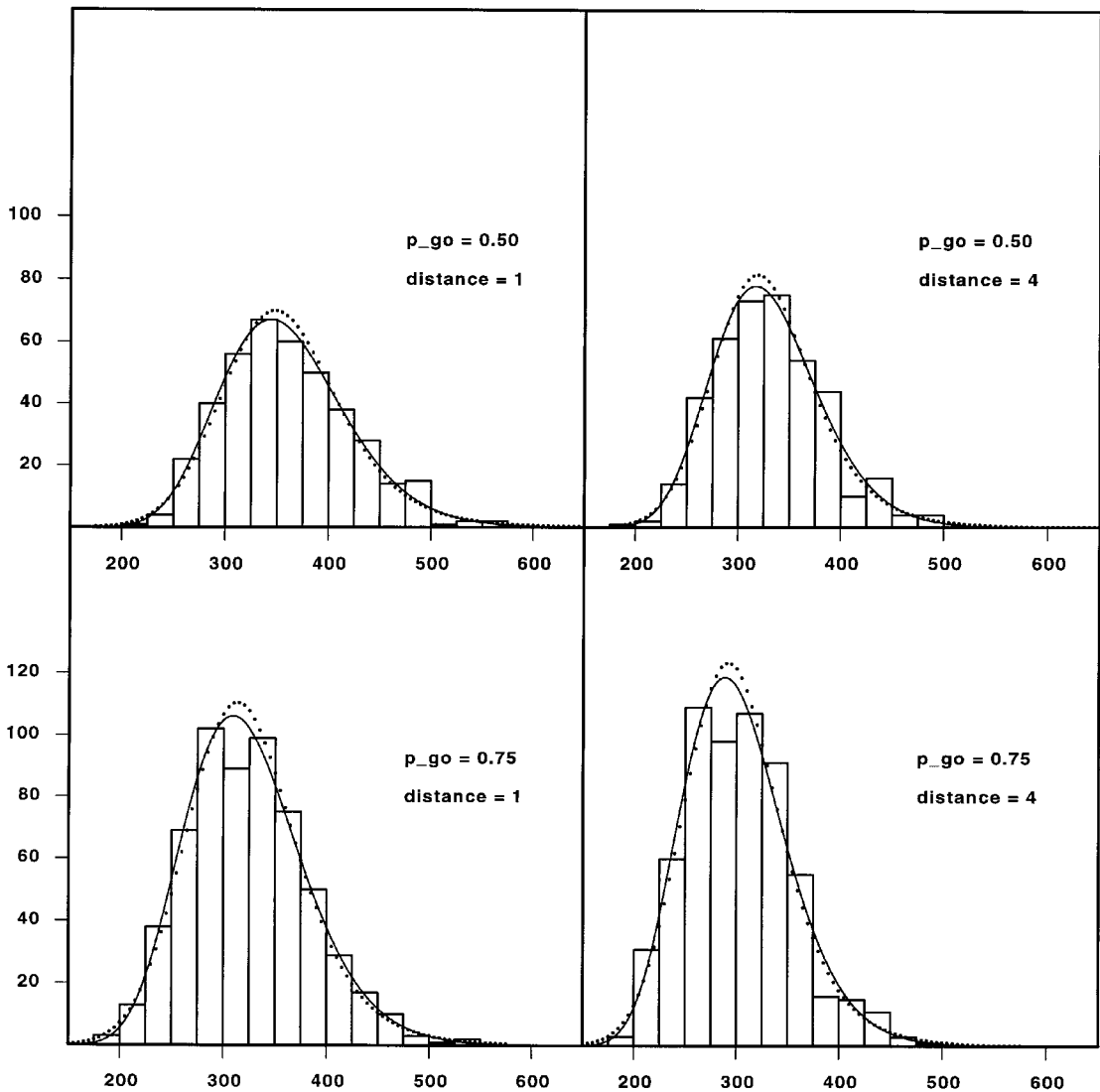
Figure 3. Histograms of RT from the 2 × 2 experimental conditions. The top row shows data for $p_{go} = .50$, the bottom row for $p_{go} = .75$; the left panels refer to a numerical distance of 1, the right panels to a numerical distance of 4. Bin width is equal to 25 msec; the number of data per histogram is 400 in the top row, and 600 in the bottom row. Abscissa: time in milliseconds. Ordinate: frequency count per bin. Theoretical ex-Wald (solid lines) and ex-Gaussian (dots) densities with unrestricted maximum likelihood parameters (see Table 2) are scaled so that they cover the same areas as the histograms.

lihood estimates of the three parameters of the ex-Wald model (scaled by $\sigma = 1$) were found by maximizing the log likelihood of the data:

$$\ln L = \sum_{k=1}^{n_d} \ln h\left(t_{ijk} \mid \mu_{ij}, \sigma_{ij} = 1, a_{ij}, \gamma_{ij}\right)$$

separately for each condition $<ij>$ as a function of the parameters $<\mu_{ij}, a_{ij}, \gamma_{ij}>$. Here, ln stands for the natural logarithm, the function h is the ex-Wald density derived in the Appendix (Equation A5), the t_{ijk} are the individual RTs observed under condition $<ij>$, and the number of data points, n_d, was equal to 400 and 600 for $p_{go} = .50$ and .75, respectively. The resulting maximum likelihood param-

eter estimates are given in Table 2, and the associated fit is shown graphically in Figure 3 (solid lines). The following observations may be made.

First, all four distributions shown in Figure 3 exhibit the positive skew typical of RTs and are well described by the ex-Wald model. To statistically evaluate these fits, RTs were first sorted into bins of width 25 msec, and observed and expected frequencies were compared via the χ^2 statistic; in this calculation, bins were collapsed as was necessary until expected frequencies were at least equal to five (cf. Hoel, 1971). Degrees of freedom (df) for each condition are given as the number of categories minus one minus the number (i.e., three) of parameters esti-

mated for each distribution. Results are given in Table 2; they confirm that the RT distributions from three conditions are in excellent agreement with the ex-Wald model; only for the fastest condition (p_{go} = .75, numerical distance = 4) is some discrepancy visible, although it does not reach significance at a level of α = .01. Overall, combining the values of χ^2 and the dfs across conditions, the fit of the ex-Wald model is very satisfactory. As a comparison, the ex-Gaussian model was fitted to the same data; for the unrestricted maximum likelihood estimates of the ex-Gaussian parameters, a reasonable overall fit was obtained, although the value of χ^2 = 57.55 for df = 30 is significant (p = .002). These fits of the ex-Gaussian model are indicated by the dots in Figure 3.

Second, the unrestricted maximum likelihood parameters of the ex-Wald model obtained support the substantive hypotheses put forth above: estimates of the evidence criterion mainly reflect variations of p_{go}, whereas estimates of the drift rate mainly reflect variations of the numerical distance. This is consistent with the interpretation, suggested by previous related work (cf. Schwarz & Stein, 1998; Smith & Mewhort, 1998), that partial numerical information about the digit presented is gradually accumulated at a rate that increases with the numerical distance of that digit to the standard. Also, the amount of evidence required before a go response is initiated increases with decreases of the a priori probability that a go response will be required on any given trial.

An important substantive question, then, is how well the ex-Wald model can account for the present data under the additional restrictions that (1) variations of p_{go} are exclusively reflected in changes of the evidence criterion a, that (2) variations of the numerical distance are exclusively reflected in changes of the drift rate μ, and that (3) neither variation has an effect on the exponential rate parameter, γ. This question will be addressed in the next section.

Evaluating Psychological Hypotheses: Likelihood Ratio Tests

Formally, in the unrestricted model fit, the ex-Wald distribution was fitted to the data by maximizing the likelihood using the most general parameter structure $<\mu_{ij}, a_{ij}, \gamma_{ij}>$; that is, in each condition $<ij>$ each parameter was determined without any restrictions. Within the ex-Wald model, the parameter restrictions outlined above represent an assumption of selective factor influence on the model parameters. More specifically, if p_{go} and numerical distance selectively influence only the evidence criterion a and the drift rate μ, respectively, then the restricted parametric model structure becomes $<\mu_j, a_i, \gamma>$. The indices in this structure indicate that μ depends on the numerical distance only, that a depends on p_{go} only, and that γ does not depend on either factor.

It is intuitively clear that relative to the unrestricted (or less restricted) model fit, the maximum of the likelihood satisfying additional boundary conditions posed on the parameters is necessarily smaller than or equal to the maximum of the likelihood without (or with fewer) such restrictions, even if the restricted model actually holds. On the other hand, the unrestricted hypothesis would be supported when the maximum likelihood under the restricted model is much smaller than for the unrestricted model. The essential problem, then, is to decide between these two cases, and it may be solved via likelihood ratio tests. Denote as L_0, L_1 the maximized likelihood under the restricted and unrestricted model, respectively. If the restricted model actually holds, then the likelihood ratio statistic

$$\lambda = -2 \cdot \ln \frac{L_0}{L_1}$$

has an approximate χ^2 distribution with its degrees of freedom equal to the number of parameters that are determined by the restricted model (for a detailed exposition of likelihood ratio tests, see Hoel, 1971). To illustrate, the unrestricted ex-Wald model fitted above has altogether 12 free parameters—namely, the triple $<\mu_{ij}, a_{ij}, \gamma_{ij}>$ for each of the four conditions $<ij>$. The restricted model, on the other hand, has only five free parameters—namely, $<\mu_1, \mu_2, a_1, a_2, \gamma>$. Thus, the parameter restrictions determine $12 - 5 = 7$ parameters, and if the restricted model holds, then the likelihood ratio statistic λ will follow an approximate χ^2 distribution with df = 7.

To apply these results to the present experiment, we maximized the likelihood of the data under the restrictions of selective factor influence on the model parameters, as explained above. Specifically, this involved maximizing the likelihood of the data as a function of the five parameters $<\mu_1, \mu_2, a_1, a_2, \gamma>$. The resulting restricted maximum likelihood estimates of μ_j were 0.326, 0.355 for numerical distance = 1, 4, respectively. Maximum likelihood estimates of a_i were 110, 98 for p_{go} = .50, .75, respectively. Finally, the maximum likelihood estimate of γ, when restricted to be identical across all conditions, was 1/22.

Next, the restricted (L_0) and unrestricted (L_1) maximum likelihoods were compared via the likelihood ratio statistic, λ. For the present data, $\hat{\lambda}$ = 2.04 for df = 7. This value of $\hat{\lambda}$ is fairly small in comparison with a χ^2 distribution with df = 7 (p = .958), indicating that the restricted maximum likelihood L_0 is not significantly smaller than L_1. This conclusion is also supported when we compare expected and observed frequencies in the same way as for the unrestricted model fit. A slightly increased χ^2 value of 41.47 was found; however, because fewer parameters are now estimated from the data, this increase is compensated by a simultaneous increase of the degrees of freedom to df = 35. Thus, according to the χ^2 criterion, the restricted five-parameter model also fits the data quite well.

In conclusion, then, the present data are perfectly compatible with the assumptions (1) that p_{go} selectively influences the evidence criterion set by the subject (but not the drift rate), (2) that numerical distance selectively

influences the drift rate with which information is accumulated (but not the evidence criterion), and (3) that γ is not influenced by either of these two factors.

Discussion

The present article presents a simple quantitative account of RT distributions, the ex-Wald model. The model is based on a distinction between data-driven accumulation processes and strategic processes assumed to be under voluntary control of the subject. In our application to numerical judgments, the data support the interpretation that response probability selectively influenced the evidence criterion while numerical distance affected the rate with which information was accumulated to pass that criterion.

We are well aware that the ex-Gaussian model, in particular, also provides flexible descriptions of RT distributions which usually lead to satisfactory statistical fits, even though for the present data the fit of the ex-Gaussian model happens to be slightly poorer than that of the ex-Wald model. However, although the ex-Wald model is of similar descriptive flexibility and has comparable computational demands, we would like to emphasize that it embodies at least in a broad sense important information-processing concepts such as gradual information accumulation, and strategic criterion setting. These concepts have already often proved very useful in the context of more elaborate random walk and diffusion models of RT. In contrast, for the ex-Gaussian model, there is usually no clear theoretical expectation regarding which factors should influence which model parameters.

Although in the present application the assumption of selective factor influence on the parameters is well supported, it is quite likely that in other applications a factor may influence not just a single parameter but, say, both μ and a. If we again draw the conceptual parallel to signal detection theory, such a finding would be analogous to a factor that influences not just d' but also the response criterion c—a finding that may complicate our understanding and interpretation of this factor but that in itself is not evidence against the basic signal detection model. Analogously, nonselective factor influence on its parameters is not evidence against the ex-Wald model; rather, it would indicate that the influence of this factor is more complex than in the simpler present case. In fact, in the case of nonselective influence, significant likelihood ratio tests based on the ex-Wald model would provide an important diagnostic indicating that a simple selective influence interpretation was not supported.

Two important aspects of the ex-Wald model proposed here should be stressed specifically. First, as explained above, the Wald distribution is skewed, it extends over nonnegative values only, it has a peaked hazard function, and it is, owing to its relation to diffusion processes, conceptually very attractive. It should be noted, though, that the Wald distribution itself has repeatedly been shown to yield quite poor fits to empirical RT distributions (Burbeck & Luce, 1982; Luce, 1986). This is probably

not too surprising, given the independent evidence, summarized above, for an additive exponential RT component. Thus, our basic result (Equation A5) for the exponential-plus-Wald convolution density is an important and necessary improvement of earlier, less successful accounts involving the Wald distribution. Second, it is precisely the relevant results for the ex-Wald density derived in the Appendix that make possible the application of statistical standard techniques such as maximum likelihood parameter estimation, and likelihood ratio tests. As discussed above, the lack of standard techniques is likely to be a major obstacle to the application of more detailed process-oriented RT models, which have, despite their conceptual attractiveness, been used far less often than readily applied descriptive RT models such as the ex-Gaussian.

Another important aspect of the ex-Wald model is closely related to a research strategy prominent in many areas of cognitive psychology: the idea that we may better understand the nature of hypothesized perceptual or cognitive processes by learning about the nature of experimental factors that influence and modify them (cf. Sternberg, 1998). In the context of this research strategy, psychological hypotheses typically imply that a specific experimental factor selectively influences a specific model mechanism, represented by a specific model parameter; a related assumption is that two specific factors might influence different model mechanisms. The ex-Wald model admits the statistical evaluation of such hypotheses for the broad concepts of strategic criterion setting, and the efficiency with which response-related information can be accumulated. An example of this, considered above, is provided by the hypotheses that in mental digit comparison the factor of numerical distance selectively influences the drift rate, and that the a priori go probability selectively influences the evidence criterion. Within the ex-Wald model, the general statistical approach to test these hypotheses using RTs from factorial experiments is based on likelihood ratio tests. Note, however, that the likelihood ratio approach presupposes that the model parameters represent, at least in a broad sense, certain information-processing concepts, so that meaningful hypotheses of selective factor influences can be derived from substantive considerations. Formally, one could of course likewise use likelihood ratio tests for the ex-Gaussian model, for example. However, precisely because there is no clear cognitive interpretation of its parameters, there would seem to be little point to such an approach; perhaps this is one of the reasons why likelihood ratio tests have in fact played no role in the context of the ex-Gaussian model.

The decomposition into effects on the drift rate versus effects on the evidence criterion may be contrasted with factorial additivity on the level of mean RT (cf. Sternberg, 1998). From the results given in the Appendix (or graphically from Figure 1), it is clear that the effects of factors selectively influencing the drift rate versus the evidence criterion will not combine additively but interact

(cf. Strayer & Kramer, 1994). For example, in the present experiment, the numerical distance effect is predicted to be larger for the condition $p_{go} = .50$ than for $p_{go} = .75$. Descriptively, the distance effect observed was in fact slightly (8 msec) larger for the low-probability condition, but the difference was too small to be significant. This pattern illustrates and explains a more general finding: For typical effect sizes (e.g., a numerical distance effect of 26 msec), the difference in drift rates, although real, is likely to be small, so that the interaction contrast predicted by the model is also fairly small, relative to typical standard errors of the cell means. In this situation, a conventional ANOVA test of factor interaction is likely to have low power. This implies that the absence of a significant interaction is not very strong evidence against the type of model illustrated in Figure 1. In fact, maximum likelihood model fits and likelihood ratio tests of the type illustrated above provide a more precise test of the hypothesis of selective influence on the parameters μ and a.

Relative to more detailed substantive models, an important limitation of the ex-Wald model is its lack of a mechanism to explain why and when errors occur in choice RT experiments. Just like the ex-Gaussian model, it cannot predict error rates, error latencies, or the relation between response speed and accuracy in choice RT experiments. This limitation is perhaps slightly obscured in the present application to a go/no-go paradigm not involving a choice between two different responses. In fact, in a go/no-go situation, false alarms may be explained by the assumption that no-go stimuli have a zero (or near-zero) drift rate, so that occasionally the evidence criterion would be reached during the period (1,000 msec in the present experiment) in which responses to the no-go stimulus would be registered. This explanation predicts that more false alarms should be observed with the lower evidence criterion in the condition $p_{go} = .75$. Descriptively, the error rates reported above conform to this prediction, although errors were too infrequent to confirm this result statistically.

In conclusion, the ex-Wald model provides a flexible description of RT distributions that avoids many of the problems associated with other descriptive models of RT. It also offers a broad cognitive interpretation of its parameters, and this interpretation may be tested using standard statistical tools, such as maximum likelihood estimation of its parameters and likelihood ratio tests of selective influence. The model is certainly not intended to replace more elaborate process-oriented models (with their associated more elaborate demands), but it is an advantageous choice when a researcher wants to analyze RT distributions by using a simple descriptive model.

REFERENCES

ASHBY, F. G. (1982). Testing the assumptions of exponential, additive reaction time models. *Memory & Cognition*, **10**, 125-134.

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

BALOTA, D. A., & SPIELER, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, **128**, 32-55.

BLANCO, M. J., & ALVAREZ, A. A. (1994). Psychometric intelligence and visual focussed attention: Relationships in nonsearch tasks. *Intelligence*, **18**, 77-106.

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

CHHIKARA, R. S., & FOLKS, J. L. (1989). *The inverse Gaussian distribution*. New York: Marcel Dekker.

COUSINEAU, D., & LAROCHELLE, S. (1997). PASTIS: A program for curve and distribution analyses. *Behavior Research Methods, Instruments, & Computers*, **29**, 542-548.

COX, D. R., & MILLER, H. D. (1965). *The theory of stochastic processes*. London: Chapman & Hall.

DAWSON, M. R. W. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods, Instruments, & Computers*, **20**, 54-57.

DEHAENE, S. (1997). *The number sense*. Oxford: Oxford University Press.

DERENZO, S. E. (1977). Approximations for hand calculators using small integer coefficients. *Mathematics of Computation*, **31**, 214-225.

EMERSON, P. L. (1970). Simple reaction time with Markovian evolution of Gaussian discriminal processes. *Psychometrika*, **35**, 99-109.

FELLER, W. (1966). *An introduction to probability theory and its applications* (Vol. II). New York: Wiley.

HEATH, R. A. (1992). A general nonstationary diffusion model for two-choice decision making. *Mathematical Social Sciences*, **23**, 283-309.

HEATHCOTE, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments, & Computers*, **28**, 427-445.

HEATHCOTE, A., POPIEL, S., & MEWHORT, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, **109**, 340-347.

HOCKLEY, W. E., & MURDOCK, B. B., JR. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, **94**, 341-358.

HOEL, P. G. (1971). *Introduction to mathematical statistics* (4th ed.). New York: Wiley.

HOHLE, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, **69**, 382-386.

JOHNSON, N. L., KOTZ, S., & BALAKRISHNAN, N. (1994). Inverse Gaussian (Wald) distributions. In N. L. Johnson, S. Kotz, & N. Balakrishnan (Eds.), *Continuous univariate distributions* (2nd ed., Vol. 1, pp. 259-297). New York: Wiley.

JUHEL, J. (1993). Should we take the shape of reaction time distributions into account when studying the relationship between RT and psychometric intelligence? *Personality & Individual Differences*, **15**, 357-360.

LAMING, D. R. J. (1968). *Information theory of choice reaction time*. New York: Wiley.

LINK, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Erlbaum.

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

MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

MADDOX, W. T., ASHBY, F. G., & GOTTLOB, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Perception & Psychophysics*, **60**, 620-637.

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

PLOURDE, C. E., & BESNER, D. (1997). On the locus of the word frequency effect in word recognition. *Canadian Journal of Psychology*, **51**, 181-194.

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

RATCLIFF, R., & MURDOCK, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, **83**, 190-214.

RATCLIFF, R., & ROUDER, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347-356.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and inter-response time in free recall. *Memory & Cognition*, **22**, 511-524.

Schwarz, W. (1990). Stochastic accumulation of information in discrete time: Comparing exact results and Wald approximations. *Journal of Mathematical Psychology*, **34**, 229-236.

Schwarz, W. (1992). The Wiener process between a reflecting and an absorbing barrier. *Journal of Applied Probability, 29*, 597-604.

Schwarz, W. (1993). A diffusion model of early visual search: Theoretical analysis and experimental results. *Psychological Research*, **55**, 200-207.

Schwarz, W. (1994). Diffusion, superposition, and the redundant-targets effect. *Journal of Mathematical Psychology*, **38**, 504-520.

Schwarz, W., & Stein, F. (1998). On the temporal dynamics of digit comparison processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1275-1293.

Seshadri, V. (1993). *The inverse Gaussian distribution.* Oxford: Oxford University Press, Clarendon Press.

Smith, D. G., & Mewhort, D. J. K. (1998). The distribution of latencies constrains theories of decision times: A test of the random-walk using numerical comparisons. *Australian Journal of Psychology*, **50**, 149-154.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, **102**, 567-593.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive psychology: Vol. 4. Methods, models, and conceptual issues* (pp. 703-863). Cambridge: Cambridge University Press.

Stone, M. (1960). Models of choice reaction time. *Psychometrika*, **25**, 251-260.

Strayer, D. L., & Kramer, A. F. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 318-341.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics: Vol. 1. Distribution theory* (5th ed.). London: Griffin.

Ueno, T. (1992). Sustained and transient properties of chromatic and luminance systems. *Vision Research*, **32**, 1055-1065.

Wixted, H. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1024-1039.

NOTE

1. Also called the Inverse Gaussian distribution; for a review of the different terminologies and notations, see Johnson, Kotz, and Balakrishnan (1994, chap. 15).

APPENDIX

In the following, we derive and summarize important formal results for the ex-Wald model. Formally, we thus consider the random variable T defined by

$$T = D + M, \tag{A1}$$

where D is the first-passage time through a level $a > 0$ in a Wiener diffusion process starting at $x = 0$, with drift $\mu > 0$ and variance parameter $\sigma^2 > 0$; M is an exponential random variable, independent of D, with rate $\gamma > 0$. To make transparent the relation to diffusion models, we retain the general parametric notation $<\mu, \sigma, a>$ for the random variable D, although one of these parameters can be fixed without loss of generality. A convenient normalization, used in the present article, is $\sigma = 1$. For more information about properties of the random variable D, we refer to the detailed monographs of Chhikara and Folks (1989) or Seshadri (1993), and to Cox and Miller (1965) or Johnson, Kotz, and Balakrishnan (1994).

I. Notation and Preliminaries

The density, f, of the random variable D is

$$f\left(w \mid \mu, \sigma, a\right) = \frac{a}{\sigma\sqrt{2\pi w^3}} \cdot \exp\left[-\frac{(a - \mu w)^2}{2\sigma^2 w}\right], \tag{A2}$$

and its cumulative distribution function (CDF), F, is

$$F\left(w \mid \mu, \sigma, a\right) = \Phi\left(\frac{\mu w - a}{\sigma\sqrt{w}}\right) + \exp\left(\frac{2a\mu}{\sigma^2}\right) \cdot \Phi\left(-\frac{\mu w + a}{\sigma\sqrt{w}}\right), \tag{A3}$$

where $\Phi(\ )$ is the cumulative standard normal distribution function.

Finally, the density, g, of the exponential random variable M is

$$g\left(w \mid \gamma\right) = \gamma \exp(-\gamma w). \tag{A4}$$

II. The Density and Cumulative Distribution Function of T

We first derive the density, h, of the random variable T as defined above. We consider the case that $\mu^2 > 2\gamma\sigma^2$. As we argue below, the remaining case is of little practical interest to the present applications. Thus, let

$$k \equiv \sqrt{\mu^2 - 2\gamma\sigma^2} \geq 0.$$

The convolution integral for the density h is

$$
\begin{aligned}
h(t\mid\mu,\sigma,a,\gamma) &= \int_0^t f(w\mid\mu,\sigma,a)\cdot g(t-w\mid\gamma)\,dw \\
&= \int_0^t \frac{a}{\sigma\sqrt{2\pi w^3}}\cdot\exp\left[-\frac{(a-\mu w)^2}{2\sigma^2 w}\right]\cdot\gamma\exp\left[-\gamma(t-w)\right]dw \\
&= \gamma\exp(-\gamma t)\cdot\int_0^t \frac{a}{\sigma\sqrt{2\pi w^3}}\cdot\exp\left[-\frac{(a-\mu w)^2-2\gamma\sigma^2 w^2}{2\sigma^2 w}\right]dw.
\end{aligned}
$$

We next convert the integrand into an expression having the form of the density f, so that we can express the last integral in terms of the CDF F. To this end, we complete the square in the numerator of the fraction within exp( ), and we factor out of the integral terms independent of the integration variable, w. The result is

$$
h(t\mid\mu,\sigma,a,\gamma) = \gamma\exp\left[-\gamma t+\frac{a(\mu-k)}{\sigma^2}\right]\cdot F(t\mid k,\sigma,a), \tag{A5}
$$

where k, F are defined above. All computations required for maximum likelihood parameter estimation and likelihood ratio tests are based on Equation A5.

Using a theorem of Ashby and Townsend (1980) about convolutions involving exponential random variables, the CDF, H, of the random variable T is

$$
\begin{aligned}
H(t\mid\mu,\sigma,a,\gamma) &= F(t\mid\mu,\sigma,a)-\frac{1}{\gamma}\cdot h(t\mid\mu,\sigma,a,\gamma) \\
&= F(t\mid\mu,\sigma,a)-\exp\left[-\gamma t+\frac{a(\mu-k)}{\sigma^2}\right]\cdot F(t\mid k,\sigma,a).
\end{aligned} \tag{A6}
$$

Equations A5 and A6 hold for $\mu^2 > 2\gamma\sigma^2$, or, in terms of the expectation, $1/\gamma$, of M:

$$
E[M] = \frac{1}{\gamma} \geq 2\cdot\left(\frac{\sigma}{\mu}\right)^2.
$$

The lower limit on the right-hand side is equal to $2\mathrm{Var}[D]/E[D]$, a value that for most reasonable choices of parameters cannot exceed very few milliseconds. To illustrate, if D has a mean of 250 msec and a standard deviation of 25 msec, then the lower limit for $E[M]$ would be equal to 5 msec. Clearly, this is much smaller than realistic estimates of $E[M]$ would be in most applications. The case of $\mu^2 < 2\gamma\sigma^2$ requires a more elaborate mathematical treatment and will therefore be covered in a separate publication.

III. Moments of T

The expectation of the random variable T is:

$$
E[T] = \frac{a}{\mu}+\frac{1}{\gamma}; \tag{A7}
$$

its variance is

$$
\mu_2 = \mathrm{Var}[T] = \frac{a\sigma^2}{\mu^3}+\frac{1}{\gamma^2}; \tag{A8}
$$

and the third central moment is

$$
\mu_3 = E\left[\left(T-E[T]\right)^3\right] = 3\frac{a\sigma^4}{\mu^5}+\frac{2}{\gamma^3}. \tag{A9}
$$

From the last two results, the usual dimensionless measure of skew, $\mu_3/(\mu_2)^{3/2}$ (cf. Stuart & Ord, 1987), is easily computed. Clearly, for the ex-Wald model, the skew is generally positive, which follows from the fact that the skew of both D and M is positive.

These results could in principle be used to provide suitable starting values for any maximum likelihood parameter search algorithm by equating the first three theoretical and empirical moments. However, the resulting system of equations tends to be complex, because it is nonlinear in the parameters. A simpler procedure starts from the observation that $1/\gamma$ is necessarily bounded from above by (and usually considerably smaller than) the RT standard deviation. For a given trial value of $1/\gamma$, the remaining parameter starting values are then easily determined to match the first two moments.

<div align="center">

APPENDIX (Continued)

</div>

IV. Computational and Numerical Aspects

As is indicated by Equations A5 and A6, computations involving the density, h, and the CDF, H, of T can all be reduced to computations involving the function F, as given by Equation A3, which in turn essentially requires the cumulative normal distribution function, Φ. Thus, the computational demand is comparable to computing the density of the ex-Gaussian model, which also involves Φ.

The first summand in Equation A3 presents no numerical problems; however, the second summand is a product the first factor of which can be become quite large, whereas the second factor can become quite small. For example, for the parameters of the order found for the experimental data in this paper, products such as $\exp(78) \cdot \Phi(-14.5)$ are typical. Many current personal computers are accurate to 15 decimal places, which implies that numbers like $\Phi(-14.5)$ are represented as zero (and so are, thus, products involving these numbers); in fact, with 15 significant places, $\Phi(z)$ is represented as zero for $z < -7.93$. It is therefore essential to use numerical implementations of the function F that take these aspects into account.

The problem can be handled by a convenient approximation to Φ due to Derenzo (1977), who shows that for $z > 5.50$

$$1 - \Phi(z) = \Phi(-z)$$

$$= \frac{1}{z \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{z^2}{2} - \frac{0.94}{z^2}\right) + \varepsilon(z)$$

where the relative error of the term $\varepsilon(z)$ is bounded by

$$\frac{|\varepsilon(z)|}{1 - \Phi(z)} \leq 4.0 \times 10^{-4}.$$

The important advantage of Derenzo's approximation in the present context is that it combines the two factors to be multiplied in such a way that it keeps the overall argument of the resulting exponential function within moderate ranges. Specifically, applying Derenzo's approximation to the second summand in Equation A3, we obtain

$$\exp\left(\frac{2a\mu}{\sigma^2}\right) \cdot \Phi\left(-\frac{\mu w + a}{\sigma\sqrt{w}}\right) \approx \frac{\sigma\sqrt{w}}{(\mu w + a)\sqrt{2\pi}} \cdot \exp\left[-\frac{(\mu w - a)^2}{2\sigma^2 w} - 0.94 \cdot \frac{\sigma^2 w}{(\mu w + a)^2}\right]$$

for

$$\frac{\mu w + a}{\sigma\sqrt{w}} > 5.50. \tag{A10}$$