Analysis of Response Time Distributions: An Example Using the Stroop Task

Andrew Heathcote, Stephen J. Popiel, and D. J. K. Mewhort
Queen's University at Kingston
Kingston, Ontario, Canada

The shape of a response time (RT) distribution can be described by a 3-parameter model consisting of the convolution of the normal and exponential distributions, the ex-Gaussian. Analyses based on mean RT do not take the distribution's shape into account and, for that reason, may obscure aspects of performance. To illustrate the point, the ex-Gaussian model was applied to data obtained from a Stroop task. Mean RT revealed strong interference but no facilitation, whereas the analysis based on the ex-Gaussian model showed *both* interference and facilitation. In short, analyses that do not take the shape of RT distributions into account can mislead and, therefore, should be avoided.

Response time (RT) distributions typically have a positively skewed unimodal shape that contains information that cannot be derived from the distribution's mean and variance. A number of studies using a variety of tasks have exploited the extra information to test models (Hacker, 1980; Hockley, 1984; Ratcliff, 1978, 1979). A distributional analysis was used, for instance, to reject the class of models for recognition memory that assumes serial processing at a constant rate: Such models predict mean RT (M_{RT}) but do not account for the shape of the distribution in both the study-test and prememorized-list paradigms (Hockley & Corballis, 1982; Ratcliff & Murdock, 1976). In spite of its proven utility, however, the literature appears to treat a distributional analysis as an esoteric supplement to the traditional analysis, namely, the analysis of M_{RT}.

In this article, we argue that a distributional analysis should not be treated as a supplementary technique. Rather, we contend that RT measures should always be analyzed using a distributional analysis and that the traditional analysis of M_{RT} risks serious misinterpretation of the data. We start by describing difficulties associated with an analysis of M_{RT}. Next, we discuss two techniques for distributional analysis, one based on moments and one based on a model that characterizes the distribution of RT scores. Finally, we illustrate the kind of misinterpretation associated with an analysis of M_{RT} by comparing it with an analysis that takes the shape of the RT distribution into account.

Analysis of Skewed Data

RT data are notoriously skewed. There are two possible explanations for the skew: On one hand, the processes of interest may yield skewed data, and if so, an analysis that takes the distribution's shape into account is required. On the other hand, the processes of interest yield symmetrical data, whereas skew reflects nuisance variables such as a lapse of attention or an eye blink. If skew reflects nuisance variables, techniques such as data trimming or rescaling can be used to remove the nuisance scores.[1] Most investigators act as if skew reflects nuisance variables—in fact, they trim the data.

Trimming the Data

The most brutal response to skew is to trim the data, that is, to drop some data from the analysis. Data are trimmed to eliminate extreme (i.e., nuisance) values and, thereby, to create a distribution closer to the normal form. Although trimming data is common practice, investigators do not agree which values warrant trimming: Some investigators trim trials with a latency longer than a fixed value; others trim trials with a latency longer than a fixed number of standard deviations from the mean.

This research was supported by a grant from the Natural Science and Engineering Research Council of Canada (AP-318) to D. J. K. Mewhort and by a Canadian Commonwealth Fellowship to Andrew Heathcote. The work was presented at the Annual Meeting of the Canadian Psychological Association, June 1989.

We acknowledge the assistance of the office of Research Services, Queen's University, and we thank E. E. Johns, B. E. Butler, and P. C. Dodwell for a critical reading of an earlier draft of the manuscript of this article. We thank Queen's University Statlab for help in the development of the distribution analysis.

Correspondence concerning this article should be addressed to Andrew Heathcote, who is now at the Department of Psychology, Northwestern University, Evanston, Illinois 60208. Electronic mail may be sent to Andrew@nurr.psych.nwu.edu (internet).

[1] The skew associated with RT raises a related statistical question. The normality assumption underlying most parametric statistical analyses is violated by RT data. To correct the violation, scale transformations such as the speed transform or a logarithmic transform can be applied. Such techniques may correct a *statistical* problem but do not correct the *semantic* problem, that is, the meaningfulness of a measure of central tendency.

Psychological Bulletin, 1991, Vol. 109, No. 2, 340–347

When is it reasonable to trim data? As noted earlier, it seems reasonable to trim trials to avoid specific nuisance variables—for example, to drop posterror trials (e.g., Rabbitt & Rodgers, 1977) or to drop trials on which the subject has been distracted.

The trouble starts when the criterion for trimming is derived from the trial's value. Here, one must face the possibility that the process of interest may have been responsible for the extreme score. One can expect extreme but valid values whenever a process yields a skewed distribution: By trimming on the basis of the trial's value, one risks tossing the baby out with the bath water.

Extreme values are more likely to reflect the process of interest when that process yields a skewed distribution than when it yields a symmetric one. Hence, researchers who use trimming to correct a skewed distribution, when the skew is produced by the processes of interest, do so in exactly the wrong situation. In short, people are most likely to resort to trimming when trimming is most dangerous.

Transforming the Data

The problem of skew can be addressed by using a data transformation to normalize the distribution, that is, to get rid of the skew by making the distribution symmetrical (e.g., Mead, 1988, pp. 283 ff). A normalizing transformation is nonlinear and, as a result, may cause a loss of information about the underlying processes (see Anderson, 1961, Figure 1; Wainer, 1977).

Sternberg (1969a) described the undesirable loss of information associated with data transformations in relation to his additive-factors method:

> Additivity will in general be destroyed by nonlinear transformation of measurements.... Furthermore, the median is inappropriate for our purpose because it is not, in general, additive. (For example, the median of a sum of components need not be the sum of the component medians.) (pp. 286–287)

Although Sternberg described the problem in terms of additive-factors logic, we believe that the problem applies more generally.

Rescaling normalizes a distribution by increasingly discounting data points as their values increase. When skew is produced by nuisance processes, one must discount in proportion to the number of data points generated by nuisance processes. The rescaled data may not describe the processes of interest correctly if one has not discounted at the correct proportions; in short, rescaling requires a principled criterion for discounting data points. Lacking a model of the RT distribution, there is no guarantee that any particular transformation incorporates the correct criterion, even when it normalizes the distribution. Hence, rescaling is not an adequate response to the problem of skew in RT distributions.

Representing the Data

The mean is the algebraic midpoint in the distribution of data. As a result, when the distribution is skewed, it misrepresents central tendency because it gives extreme values too much weight. Hence, M_{RT} can be misleading.

For skewed distributions, some authorities recommend the median as a measure of central tendency (e.g., Hays, 1981, p. 158; but see Miller, 1988). The use of the median reduces the weight given to extreme values, but it does not solve all of the problem. A measure of central tendency is intended as a summary of data. When the distribution is skewed, the mean, median, and mode do not converge, and the concept of central tendency is unclear. We contend that central tendency is a meaningful concept only for symmetrical distributions.

Summary

We have argued that a conventional analysis of means or medians may be misleading. Likewise, data should not be trimmed because there is no clear criterion with which to distinguish outliers from valid extreme scores. Data transformations may be misleading and often discard valuable information. In short, a procedure is needed that preserves information without introducing distortions. A distributional analysis solves the problem: It preserves all information available in the experiment and provides a clear description of the behavior of interest.

Methods of Distributional Analysis

To exploit the information contained in an RT distribution, one must describe the distribution's characteristic shape. Skew and kurtosis (derived from higher order moments) describe the shape of a distribution without assuming an explicit model for it (Sternberg, 1964, 1969b). As Ratcliff (1979) documented, however, a description based on higher order moments suffers from two major drawbacks. First, the sampling variance associated with estimates of higher order moments is large: To obtain stable estimates, several thousand RTs must be collected. Second, estimates for higher order moments are very sensitive to extreme scores. Hence, even when a large number of observations has been collected, the estimates of the moments may be contaminated by only a few outliers.

An alternative approach is to assume an explicit model for the shape of the RT distribution. Characterization of RT distributions avoids the problem of misrepresenting central tendency by providing a complete account of the entire distribution; it also provides more information with which to constrain models.

Early attempts to characterize the distribution of RT scores were motivated by the hope that a distributional analysis would help to specify the nature of the processes underlying the observed scores (McGill, 1963). Hohle (1965), for example, proposed a decomposition of RT into decision versus other components and derived a mathematical description of the RT distribution. In his analysis, the distribution of the residual component is normal and that of the decision component is exponential. The resultant model for the RT distribution, assuming that the normal and exponential components are independent, is the convolution of the normal and the exponential functions, the ex-Gaussian distribution. For values of σ and τ greater than zero, the ex-Gaussian distribution is defined by the following equation, its probability density function:

$$f(t|\mu, \sigma, \tau) = \frac{1}{\tau(2\Pi)^{1/2}} \exp\,[\sigma^2/2\tau^2 - (t-\mu)/\tau] \times \int_{-\infty}^{(t-\mu)/\sigma-(\sigma/\tau)} \exp\,(-y^2/2)^{dy} \quad (1)$$

where t is time, μ is the mean of the normal component, σ^2 is the variance of the normal component, and τ is the parameter (both mean and standard deviation) of the exponential component. Intuitively, μ reflects the mode and τ reflects the size of the tail of an RT distribution.

The rationale underlying Hohle's (1965) decomposition of RT is problematic because his attribution of the exponential component to decision processes and the normal component to residual processes remains unproven. Nevertheless, as Luce (1986, p. 439) noted," for reasons unknown, the ex-Gaussian . . . fits response-time distributions well." Ratcliff and Murdock (1976), for example, found that the ex-Gaussian provided a better account of RT distributions than is provided by either the gamma or the lognormal distributions. Thus, the questionable theoretical status of its derivation notwithstanding, the ex-Gaussian remains a useful model of the RT distribution.

Moments as Functions of the Ex-Gaussian's Parameters

Ratcliff (1979, p. 456) has shown that the mean and variance of an RT distribution, M_{RT} and $\hat{s}^2_{RT}$, can be expressed as functions of the ex-Gaussian's parameters. In particular,

$$M_{RT} = \mu + \tau \tag{2}$$

and

$$\hat{s}^2_{RT} = \sigma^2 + \tau^2 \tag{3}$$

for σ and τ greater than zero.[2] Within the limits indicated by Equations 2 and 3, two RT distributions with the same M_{RT} may have different degrees of positive skew. Similarly, two RT distributions with the same $\hat{s}^2_{RT}$ may have quite different shapes. Unless the distribution is normal, a traditional RT analysis, using two parameters, is ambiguous. The ambiguity can be resolved with the three parameters provided by the ex-Gaussian analysis.

Fitting the Ex-Gaussian

In contrast to methods based on higher order moments, one can obtain stable estimates for the ex-Gaussian distribution's parameters without Herculean effort (Ratcliff & Murdock, 1976). Stable estimates require as few as 100 RT observations for each condition, more observations than are usually collected in experiments aimed at determining M_{RT}, but an order of magnitude fewer than are required for stable estimates of higher order moments. Moreover, if large numbers of RT observations cannot be obtained from each subject—perhaps because data collection is too expensive or because the number of stimuli available in a particular class is too small—Ratcliff (1979) has shown that stable estimates of the distribution's parameters may be obtained by averaging across subjects using a technique (Vincent averaging) that preserves the component distributions' shapes (see Woodworth & Schlosberg, 1954, pp. 535–536).

The first step toward obtaining parameter estimates is to define a goodness-of-fit function, that is, a function that reflects how well the model fits the data. Then choose values of the model's parameters that maximize the fit, as defined by the goodness-of-fit function. A goodness-of-fit function commonly used in the behavioral sciences is chi-square. Smaller values of chi-square indicate a better fit; therefore, in fitting a model to data using chi-square, we choose parameters that minimize chi-square.

Because the number of possible combinations of parameters is large, finding the best combination can be difficult. Fortunately, the problem, often called the *optimization problem*, has been the subject of extensive study in the mathematical literature. Of the techniques available, the SIMPLEX algorithm is preferred for fitting the ex-Gaussian, mainly because it is robust. An excellent practical review of techniques for optimization (including SIMPLEX) is provided by Press, Flannery, Teukolsky, and Vetterling (1988, pp. 274–334).

When fitting frequency distributions, chi-square is seldom used as a goodness-of-fit function. To calculate chi-square one must first calculate observed and expected frequencies. Unfortunately, the value of chi-square depends on the way the data are aggregated when the latter frequencies are calculated. We use a likelihood function instead of chi-square as a goodness-of-fit function because it does not require grouping of the data. Ratcliff and Murdock (1976) listed other desirable properties of likelihood estimation.

A likelihood estimator is constructed as follows: First, calculate the probability of each data point by using the model's density function. Second, multiply the probabilities associated with each data point. The product is the value of the likelihood estimator. We choose model parameters that maximize the likelihood function given the observed data.

For the ex-Gaussian model, the probability density function, $f(t|\mu, \sigma, \tau)$, determines the probability of an observed RT. To determine the value of the likelihood function for given values of the model parameters, calculate $f(t|\mu, \sigma, \tau)$ for each data point by setting t equal to each observed RT in turn. Instead of determining the product of the values, take the sum of their natural logarithms. Maximizing the sum of the logarithms is equivalent to maximizing the product. We used the SIMPLEX algorithm to choose the model parameters that maximize the sum of the logarithms.

Having estimated the best fitting parameters, we wanted to know how good the fit was. We obtained an assessment of the fit by superimposing the estimated ex-Gaussian function on a histogram of the corresponding data. The quality of the fit depends, of course, on the way the data were aggregated, that is, on the width of the histogram's bins. Although the likelihood estimator provides a precise quantitative assessment of the model's fit, we were not able to perform inferential testing using that estimator because we do not know its distribution. Hence, we used chi-square to assess the quality of fit, keeping in mind the aggregation problem.

The process of fitting the model can be computationally expensive. Because we used a number of computational tricks to speed the process, our program can compute best fitting param-

[2] The identities are derived using the method of moments. We do not know that the identities hold between sample estimates of the mean and variance and ex-Gaussian parameters derived using a maximum likelihood estimator. It has been our experience, however, that the identities approximately hold.

eters in reasonable time on a relatively slow machine (e.g., an IBM personal computer [PC]). Interested readers may obtain a copy of our program for the IBM PC from us.[3]

An Example Using the Stroop Effect

As noted, we contend that a distributional analysis should not be treated as a supplementary technique. To illustrate the kind of misinterpretation that a distributional analysis can obviate, we performed a Stroop color-naming study and compared the results of a traditional analysis with an analysis based on an ex-Gaussian characterization of the RT distributions. We used the Stroop effect both because it is well documented as a phenomenon in its own right (e.g., Dyer, 1973; La Heij, 1988; La Heij, Van der Heijden, & Schreuder, 1985; Neumann, 1984) and because it is widely used as a tool with which to examine other phenomena (e.g., Cheesman & Merikle, 1984).

Method

Subjects. Eight students from Queen's University participated in the experiment. Subjects were paid $5 for their participation.

Apparatus and stimuli. Stimuli were presented on a Zenith Data Systems ZVM-1330 color monitor controlled by a Zenith Z-158 PC. Responses were obtained using a Sony dynamic headset microphone attached to a Vero voice key. Millisecond timing and the synchronization of stimulus presentation with screen refresh were achieved using Heathcote's (1988) programs. Word stimuli were RED and GREEN; neutral stimuli were XXX and XXXXX. Viewing distance was 0.9 m: RED and XXX subtended 0.55°; GREEN and XXXXX subtended 0.93° horizontally.

Design. A 2 × 3 factorial design was used. The stimuli were presented in either red or green uppercase letters, and there were three presentation conditions: (a) congruent—RED displayed in red or GREEN displayed in green, (b) incongruent—RED displayed in green or GREEN displayed in red, and (c) neutral—XXX or XXXXX in red or green.

Subjects performed 18 practice trials (6 from each of the three conditions) and 720 experimental trials. The experimental trials were administered in 12 blocks of 60 correctly answered trials. The order of conditions was randomized within blocks. The three conditions were presented equally often, and the three- and five-character neutral stimuli were presented equally often within the neutral condition. Trials scored as errors or as equipment failures were repeated later in the block. Error trials were followed by "dummy" trials to eliminate slow posterror responses. RTs were recorded from all trials except the dummy trials. Experimental trials resumed when a correct response was made on a dummy trial.

Procedure. Subjects were asked to name the color of the stimulus; the instructions stressed both speed and accuracy. The subjects were asked to ignore the shape of the stimulus while keeping it in focus and in central vision. When a fixation cross appeared at the center of the screen, the subject initiated a trial by depressing a push button. The fixation cross remained for 500 ms and was replaced by the stimulus. The middle character of each stimulus was positioned to cover the position of the fixation cross. Stimuli remained for approximately 100 ms.

The experimenter classified each response as correct, as incorrect, or as an equipment failure. Equipment failures occurred when the voice key closed before the subject responded or when it failed to close. If the subject did not respond within 3 s, the message RESPONSE TOO SLOW was displayed for 1 s, and the trial was repeated later in the block. An incorrect response was scored if the subject mispronounced the correct color name. An incorrect response triggered feedback in the form of the message INCORRECT RESPONSE. At the end of each block, a message reported the number of incorrect responses in the block.

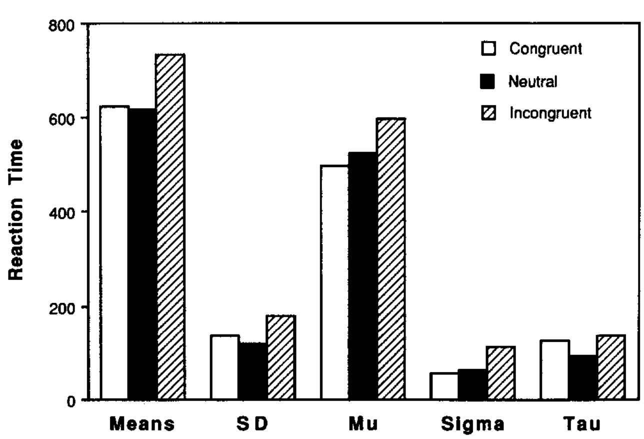

Figure 1. Mean RT (M_{RT}), standard deviation ($\hat{s}_{RT}$), μ, σ, and τ averaged over subjects and stimulus color.

Results

The best fitting ex-Gaussian parameters were obtained for each subject in each condition using maximum likelihood estimation, and chi-squares were obtained for each fit. Analyses of variance were performed on M_{RT}, $\hat{s}_{RT}$, percentage of error, and the three ex-Gaussian parameters. There was no interaction between stimulus color (red or green) and presentation condition (congruent, incongruent, or neutral) in any of the measures (all Fs < 1). Consequently, all further analyses were collapsed across stimulus color.

The traditional analysis. Mean RT, $\hat{s}_{RT}$, and the three ex-Gaussian parameters, averaged across subjects and stimulus color, are presented in Figure 1. As shown in Figure 1. M_{RT} in the incongruent condition was longer than in the neutral condition, $F(1, 7) = 50.62$, $p < .001$; that is, the experiment yielded a classic Stroop effect.

Mean RT in the congruent condition, however, did not differ from that in the neutral condition, $F(1, 7) = 1.37$, $p > .25$. Thus, congruency yielded neither interference nor facilitation. Note that we use the terms *interference* or *facilitation* whenever a measure for either the congruent or the incongruent condition is, respectively, larger or smaller than the same measure from the neutral condition. We apply the terms *interference* and *facilitation* to all measures, not just to M_{RT}.

Standard deviation of RT in the neutral condition was smaller than $\hat{s}_{RT}$ in both congruent condition, $F(1, 7) = 11.58$, $p < .02$, and the incongruent condtion, $F(1, 7) = 149.73$, $p < .0001$. Thus, both the congruent and incongruent conditions exhibited interference. $\hat{s}_{RT}$ in the congruent condition was less than in the incongruent condition, $F(1, 7) = 23.51$, $p < .001$.

Ex-Gaussian analysis. Kolmogorov one-sample tests (see

---

[3] Please forward a formatted floppy disk plus $10 for postage and handling to Andrew Heathcote, Department of Psychology, Northwestern University, Evanston, Illinois 60208.

Conover, 1980) indicated that each of the 48 distributions of data (one for each combination of subjects and conditions) differed significantly from the normal (all $Zs > 10.0$, $ps < .001$). Frequency histograms for a representative subject are shown in Figure 2. The figure also shows the corresponding fitted ex-Gaussian functions. The ex-Gaussian model was rejected (at the 1% level) in 4 of the 48 fits, a frequency that indicates that the ex-Gaussian model provides a good description of the data.

Figure 3 presents the average shape of the RT distribution across subjects; it shows the probability density distributions obtained by Vincent averaging over subjects and stimulus color for the congruent, the neutral, and the incongruent conditions. The Vincent average was obtained as follows: First, we described each subject's distribution by calculating successive values bracketing 5% of the responses (semideciles), and then we averaged the semideciles to obtain the curve shown in Figure. 3.

Note that both the congruent and incongruent conditions exhibit greater skew than the neutral condition: The mode of the congruent condition is less than that for the neutral condition, that which, in turn, is less than that for the incongruent

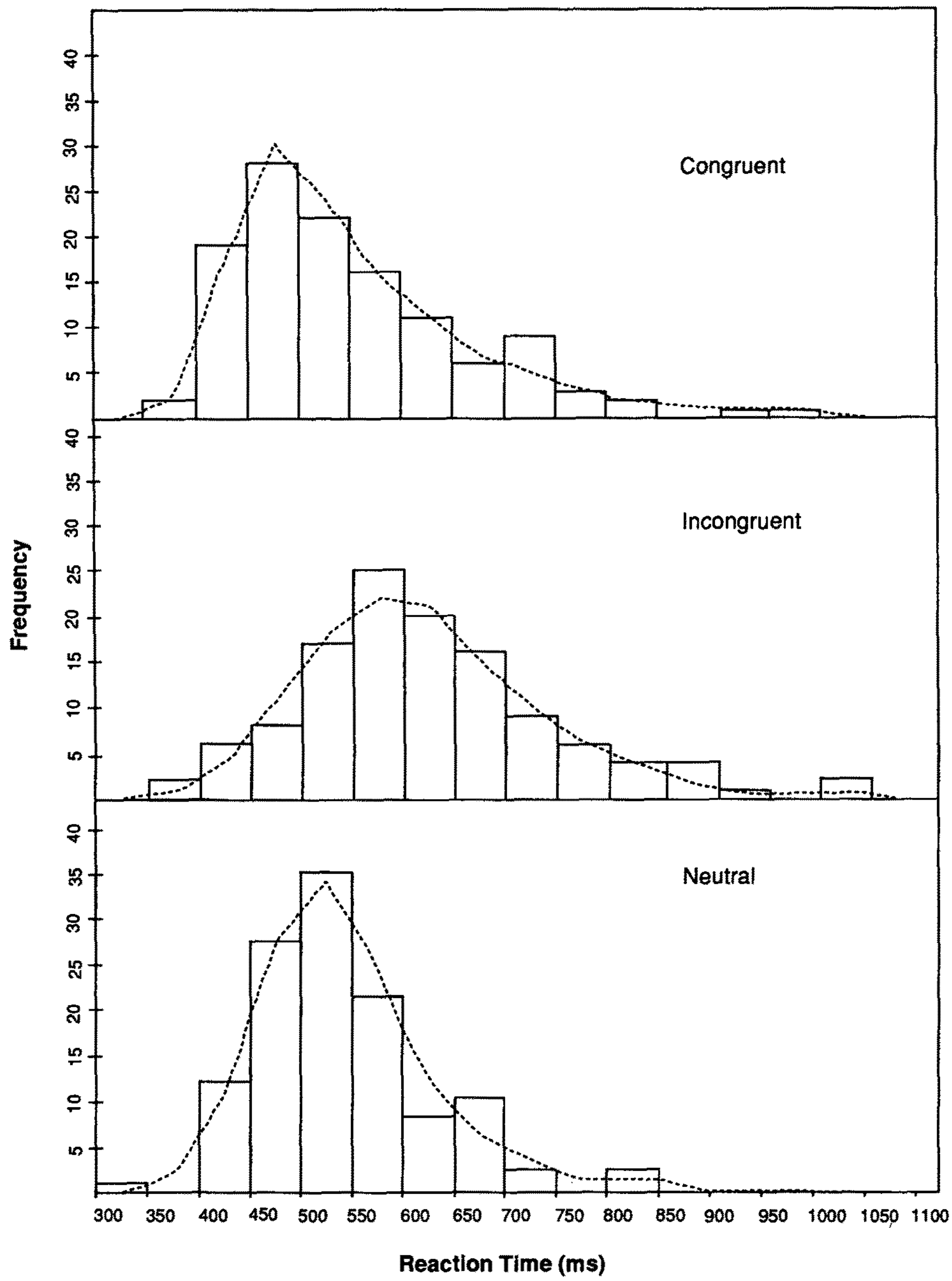

Figure 2. Frequency histograms with fitted ex-Gaussian functions for a representative subject. (The top panel presents the data for the congruent condition, the middle panel presents the data for the incongruent condition, and the bottom panel presents the data for the neutral condition.)

condition. Figure 3 also shows the corresponding ex-Gaussian functions. As is clear in Figure 3, the ex-Gaussian distribution provides a good fit to the averaged data. When read in conjunction with the mean values of the ex-Gaussian parameters (Figure 1), Figure 3 illustrates the relation between a distribution's shape and its parameter's values. Note, in particular, that larger values of τ correspond to larger tails and that larger values of μ correspond to larger modes.

In contrast to the analysis of M_{RT}, analysis of μ showed both interference in the incongruent condition, $F(1, 7) = 16.52$, $p < .01$, and facilitation in the congruent condition, $F(1, 7) = 9.48$, $p < .05$. Sigma (σ) showed interference, $F(1, 7) = 26.77$, $p < .005$, but did not differ between the neutral and the congruent conditions, $F(1, 7) = 2.80$, $p > .10$. For τ, there was interference for both congruent and incongruent conditions; that is, τ for both the congruent and incongruent conditions was greater than for the neutral condition, $F(1, 7) = 9.69$, $p < .05$, and $F(1, 7) = 21.13$, $p < .005$, respectively, and τ for the congruent and incongruent conditions did not differ, $F(1, 7) < 1$.

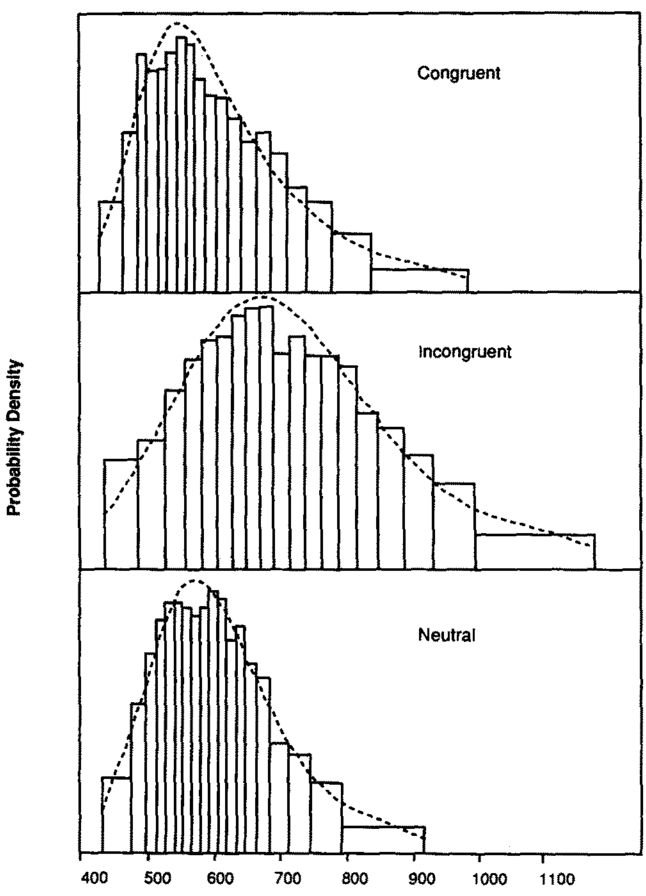

Figure 3. Probability density histograms produced by Vincent averaging over subjects and stimulus color; the figure also shows the fitted ex-Gaussian functions. (The top panel presents the data for the congruent condition, the middle panel presents the data for the incongruent condition, and the bottom panel presents the data for the neutral condition.)

Error analysis. Errors in the incongruent condition were significantly more frequent than errors in the neutral condition, $F(1, 7) = 25.35$, $p < .005$, and in the congruent condition, $F(1, 7) = 23.51$, $p < .005$. Errors in the congruent and neutral conditions did not differ, $F(1, 7) < 1$. The means were 0.26%, 4.71%, and 0.21% for congruent, incongruent, and neutral conditions, respectively. As errors were positively correlated with M_{RT}, our data are not seriously confounded by a speed–accuracy trade-off.

Sign tests. Because our data did not meet the distributional and homogeneity assumptions associated with a conventional F test, we also performed sign tests (which are based on the number of subjects who show the difference) for all of the 18 comparisons described so far. We obtained the same pattern of significance with one exception; namely, the sign test indicated that σ for the neutral condition was significantly greater than σ for the congruent condition ($p = .035$). The corresponding parametric test was not significant, $F(1, 7) = 2.80$, $p > .10$. In light of the close correspondence between the two methods of inferential analysis, it is clear that our results are stable across subjects.

Practice effects. An ex-Gaussian analysis would be of little value if its outcome is an artifact of practice. An ex-Gaussian analysis demands a large number of trials, and hence practice effects are likely. To check the stability of the ex-Gaussian analysis across practice, we analyzed subjects' performance as a function of trials. To look at practice effects, we divided the data for the 12 blocks into three groups of 4 blocks: Each group contained approximately 40 data points per condition. Because 40 data points are too few to fit the ex-Gausian model directly, we averaged across subjects within the three groups of 4 blocks by means of Vincent averaging. Vincent averaging permitted us to obtain stable parameter estimates with fewer data points per subject. The cost of the procedure is that we lost subject variance and, hence, could not perform inferential testing.

Table 1 reports M_{RT}, $\hat{s}_{RT}$, and the three ex-Gaussian parameters averaged over stimulus color. The values are shown for each condition and group of blocks, one row per condition per group of blocks.

All parameters tended to decrease with practice. With one exception, however, the relationships among the conditions shown in Figure 1 remained stable across practice. The exception concerned τ: The pattern in the overall data did not emerge until the second group of four blocks. It is clear, therefore, that the ex-Gaussian analysis is stable across practice.

Discussion

The results can be summarized as follows: The congruent condition exhibited facilitation in μ, interference in both τ and $\hat{s}_{RT}$, and no effect in the other measures. The incongruent condition exhibited interference in all measures and was larger than the congruent condition in all measures but τ. All measures indicate that an incongruent color word interferes with color naming. The conclusion for congruent color words is less clear.

In terms of M_{RT}, congruency did not affect color naming. A reasonable conclusion is that a congruent word has no effect on color naming. That interpretation is wrong: $\hat{s}_{RT}$ was larger in the congruent condition than in the neutral condition. We ac-

Table 1
Mean RT (M_{RT}), Standard Deviation ($\hat{s}_{RT}$), μ, σ, and τ as a Function of Practice and Presentation Condition

Practice and presentation condition	Parameter				
	M_{RT}	$\hat{s}_{RT}$	μ	σ	τ
Blocks 1–4					
C	626.2	112.3	522.9	54.3	103.4
I	776.9	171.6	627.0	97.2	150.0
N	638.5	108.1	545.5	56.4	92.9
Blocks 5–8					
C	632.7	129.1	500.5	52.7	132.2
I	731.9	160.0	599.4	101.0	132.5
N	623.5	109.2	528.3	56.8	95.2
Blocks 9–12					
C	613.6	126.9	485.7	43.1	127.9
I	688.5	160.4	559.1	99.8	129.4
N	587.8	89.7	509.1	51.6	78.8

Note. C = congruent; I = incongruent; N = neutral.

knowledge that a larger $\hat{s}_{RT}$ is often associated with a larger M_{RT} because the floor effect exerted by the lower bound on RT is attenuated when M_{RT} increases. Hence, a shift in variance is often assumed to be of no theoretical interest.[4] In the present case, however, the increase in $\hat{s}_{RT}$ was not associated with an increase in M_{RT}. Hence, the conclusion that a congruent word has no influence on color naming is untenable, but it is difficult to formulate a simple theoretical interpretation from the two measures provided by the traditional analysis.

Figure 1 shows why a larger $\hat{s}_{RT}$ was observed in the congruent condition. The value of τ associated with the congruent condition was larger than that for the neutral condition. Furthermore, σ did not differ between the congruent and neutral conditions. Hence, the larger $\hat{s}_{RT}$ in the congruent condition was due entirely to the presence of a longer tail. Although the longer tail directly contributed to M_{RT}, its effect was balanced by facilitation for μ in the congruent condition. Hence, the lack of facilitation for M_{RT} in the congruent condition was due to a change in shape of the RT distribution and does not indicate that a congruent word has no influence on color naming.

The foregoing illustrates how a traditional RT analysis, especially when restricted to M_{RT}, can be misleading. A congruent color word actually caused both facilitation in the mean of the normal component, μ, and an increase in positive skew. When examined in terms of M_{RT}, the two effects canceled each other. Thus, a facilitation effect was *masked* by a change in the shape of the RT distribution.

Facilitation in the congruent condition has been an inconsistent finding in the Stroop literature. Cheesman and Merikle (1984) found strong facilitation in their version of the task. Dunbar and MacLeod (1984) found facilitation on one occasion (Experiment 2), but not on others (Experiments 1, 3, and 4). Dalrymple-Alford (1972) found facilitation when the neutral words were noncolor words (e.g., *joy*), but not when the neutral words were nonwords (i.e., a series of *X*s). The inconsistency among these findings may reflect the size of the facilitation effect: One would expect inconsistency across studies if the effect were very small. Our analyses suggest another possibility: The inconsistency may reflect changes in the shape of the RT distribution.

It is tempting to attribute particular processes to the parameters of the ex-Gaussian model. Hohle (1965) suggested that τ reflects decision processes and that both μ and σ reflect noise. In our experiment, however, μ and σ were systematically affected by presentation condition and, hence, do not reflect noise. Clearly, the present results do not support Hohle's attribution of parameters to cognitive processes. The nuisance interpretation for skew attributes μ and σ to the processes of interest and τ to nuisance processes. In our experiment, however, τ was systematically influenced by the experimental manipulations. We would not expect the appearance of a word to cause more lapses of attention (or eye blinks) than the appearance of a nonword. Hence, the present results do not support a nuisance interpretation of τ. In short, we have not found an acceptable attribution for the parameters of the ex-Gaussian. Although the ex-Gaussian model describes RT data successfully, it does so without the benefit of an underlying theory.

Should we be uncomfortable with the ad hoc nature of the ex-Gaussian model? In the best of all possible worlds, we would like to be able to offer substantive models that predict response latency in particular paradigms, and we would like each model's parameters to map onto the cognitive processes that underlie performance in the particular paradigms. We do not have such an interpretation for the ex-Gaussian's parameters: We treat the ex-Gaussian as a descriptive first-order account of response latency. Because we do not propose the ex-Gaussian model as a model of cognitive processes, we do not need a

[4] The general linear model underlying most parametric inferential testing assumes that the variance associated with data from each treatment level is equal. RT data usually violate the homogeneity of variance assumption. The traditional remedy is scale transformation. As noted, scale transformations cause a loss of information. The ex-Gaussian analysis provides a more generally effective method of redressing the violation of the homogeneity of variance assumption without a loss of information.

cognitive attribution for its parameters. Indeed, we doubt that any single model, especially one of such simplicity, could describe the cognitive processes in psychology's various paradigms.

The virtue of an ex-Gaussian analysis is that it provides a good description of the data. Because it provides a good description, it allows researchers to decide whether skew can be ignored and, thereby, helps to avoid errors of interpretation, errors that, we fear, are likely with the traditional analysis of M_{RT}.

References

Anderson, N. H. (1961). Scales and statistics: Parametric and nonparametric. *Psychological Bulletin, 58,* 305–316.

Cheesman, J., & Merikle, P. (1984). Priming with and without awareness. *Perception & Psychophysics, 36,* 387–395.

Conover, W. J. (1980). *Practical nonparametric statistics.* New York: Wiley.

Dalrymple-Alford, E. C. (1972). Associative facilitation and interference in the Stroop color-word task. *Perception & Psychophysics, 11,* 274–276.

Dunbar, K., & MacLeod, C. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 622–639.

Dyer, F. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition, 1,* 106–120.

Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 651–675.

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.

Heathcote, A. (1988). Screen control and timing routines for the IBM microcomputuer family using a high-level language. *Behavior Research Methods, Instruments, & Computers, 20,* 289–297.

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6,* 598–615.

Hockley, W. E., & Corballis, M. C. (1982). Tests of serial scanning in item recognition. *Canadian Journal of Psychology, 36,* 189–212.

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

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition, 16,* 400–410.

La Heij, W., Van der Heijden, A. H. C., & Schreuder, R. (1985). Semantic priming and Stroop-like interference in word naming tasks. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 62–80.

Luce, R. D. (1986). *Response times.* New York: Oxford University Press.

McGill, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 309–360). New York: Wiley.

Mead, R. (1988). *The design of experiments.* New York: Cambridge University Press.

Miller, J. (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 539–543.

Neuman, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz & A. Sanders (Eds.), *Cognition and motor processes* (pp. 255–294). Berlin, Federal Republic of Germany: Springer-Verlag.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical recipes: The art of scientific computing.* Cambridge, England: Cambridge University Press.

Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology, 29,* 727–743.

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

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

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 190–214.

Sternberg, S. (1964, October). *Estimating the distribution of additive reaction time components.* Paper presented at the meeting of the Psychonomic Society, Niagara Falls, Ontario, Canada.

Sternberg, S. (1969a). The discovery of processing stages: Extensions of Donder's method. In W. G. Koster (Ed.), *Attention and Performance II* (pp. 267–315). Amsterdam: North-Holland.

Sternberg, S. (1969b). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57,* 421–457.

Wainer, H. (1977). Speed vs reaction time as a measure of cognitive performance. *Memory & Cognition, 5,* 278–280.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology.* New York: Holt, Rinehart & Winston.

Received October 13, 1989
Revision received February 20, 1990
Accepted March 22, 1990 ■