

Effects of Truncation on Reaction Time Analysis

Rolf Ulrich and Jeff Miller

Many reaction time (RT) researchers truncate their data sets, excluding as spurious all RTs falling outside a prespecified range. Such truncation can introduce bias because extreme but valid RTs may be excluded. This article examines biasing effects of truncation under various assumptions about the underlying distributions of valid and spurious RTs. For the mean, median, standard deviation, and skewness of RT, truncation bias is larger than some often-studied experimental effects. Truncation can also seriously distort linear relations between RT and an independent variable, additive RT patterns in factorial designs, and hazard functions, but it has little effect on statistical power. The authors report a promising maximum likelihood procedure for estimating properties of an untruncated distribution from a truncated sample and present in an appendix a set of procedures to control for truncation biases when testing hypotheses.

Researchers who collect reaction times (RTs) usually obtain a certain percentage of observations that are spurious in the sense of being strongly influenced by processes other than those that the experimenter intends to measure. For example, a subject may occasionally anticipate the stimulus rather than processing it; a physical device may falsely report that a response was made; or a subject may be distracted from the task by an incidental visual, auditory, or somatosensory input or by an intrusion of task-irrelevant cognition (e.g., remembering a forgotten appointment). When spurious observations are extreme, of course, they can distort results (e.g., mean RT) rather seriously unless they are dealt with (cf. Barnett & Lewis, 1978).

In the course of data analysis, two basic strategies are available to minimize the influence of spurious observations on the outcome of an experiment, and these strategies are often referred to as *accommodation* and *identification* (Beckman & Cook, 1983, p. 123). The former strategy is to use what is known in the statistical literature as a *robust estimator*—a summary measure of RT that is relatively insensitive to spurious RTs (cf. Stuart & Ord, 1987). For example, some researchers use the median RT as a summary measure of location, as the median is relatively uninfluenced by spuriously fast or slow RTs. The latter strategy is to use

an *outlier elimination* method to exclude spurious RTs from the analysis. For example, one might truncate the data set to some a priori cutoffs (e.g., excluding all RTs outside the range of 200–2,000 ms). Alternatively, one might let the data suggest which RTs were spurious, excluding all observations greater than two or three standard deviations from a given mean.

Even a cursory examination of the RT literature reveals that different experimenters use different approaches to the problem of spurious RTs. Consider, for example, the 1992 volume of the *Journal of Experimental Psychology: Human Perception and Performance*, which contains 35 articles reporting studies of RT. A minority—13 articles—ignored the possibility of spurious RTs, simply analyzing the mean of all RTs (though often excluding errors);¹ in 5 articles, the median was used; and in 17 articles, an outlier detection procedure was used, rejecting from 0.3 to 6% of the RTs as spurious. Of the latter 17 articles, 9 eliminated observations using truncation (using eight different ranges of acceptable RTs), 5 eliminated observations more than a criterion number of standard deviations from the mean (four different criteria and two different definitions of the mean were used), and 3 used other techniques.

It is striking that 35 arbitrarily selected studies used such a variety of different computational procedures to obtain the central tendency of the RTs collected from a given subject in a given condition. One might naturally wonder why the experimenters used the method that they did, instead of one of the other methods. More important, one might wonder whether the results would have been different had one of the other methods been used, and one could question how experimenters ought to minimize the disruptive effects of spurious RTs.

The present article begins to address these questions systematically, through a detailed examination of the effects of truncating RT distributions. Our main goal is to describe the biasing effects of truncation on different kinds of RT analy-

Rolf Ulrich, University of Konstanz, Konstanz, Germany; Jeff Miller, Department of Psychology, University of California, San Diego.

This work was supported by the Deutsche Forschungsgemeinschaft (UL 88/2-1), Grant 13/92 from the University of Konstanz, and Grant PHS-40733 from the National Institute of Mental Health. This work was begun while Rolf Ulrich was a visiting scholar at the Center for Human Information Processing, University of California, San Diego.

We would like to thank Robert L. Greene, Andrew Heathcote, Geoffrey R. Loftus, R. Duncan Luce, James T. Townsend, Dirk Vorberg, and several anonymous reviewers for helpful comments.

Correspondence concerning this article should be addressed to Rolf Ulrich, Universität Konstanz, Fachgruppe Psychologie, Postfach 5560 D 29, 78434 Konstanz, Germany, or Jeff Miller, who is now at Department of Psychology, University of Otago, Post Office Box 56, Dunedin, New Zealand.

¹ The number 13 may be inflated, because we included in this category all studies that did not explicitly describe some other procedure for summarizing RT.

ses. RT experiments are typically designed to test predictions of explicit or implicit models, which have heretofore made predictions about full rather than truncated distributions of RT. Unfortunately, techniques for excluding outliers—truncation included—usually exclude a certain percentage of valid observations as well as spurious ones, because it is rarely possible to discriminate perfectly between spurious data points and valid but extreme observations. Indeed, most current RT models describe RT distributions as having an infinite tail in at least one direction, so it is theoretically quite reasonable to expect that truncation would eliminate some valid RTs in addition to some or all of the spurious ones. The question, then, is to what extent results are biased when extreme but valid RTs are excluded. Is the bias introduced by truncation so small that it can safely be ignored, or is it instead large enough to artifactually create or conceal discrepancies between models and data?

Obviously, bias could be a serious problem in experiments testing quantitative models, because it could perturb the data values that the quantitative model was designed to fit. Perhaps surprisingly, bias can also be devilish in experiments testing for qualitative effects, such as an experiment designed to find out whether a certain drug increases mean RT in a given task. Even if the size of the effect is not an issue, only its presence versus absence, there is the danger that bias may differ across conditions. For example, eliminating observations more than a criterion number of standard deviations from the mean introduces a bias that depends on both the sample size and the skewness of the RT distribution, and conditions differing in either one of these characteristics are differentially biased (Miller, 1991). Bias that depends on condition is extremely dangerous, even for qualitative comparisons, because it can artifactually produce or conceal differences between conditions.

We had three main reasons for our choice to focus on the truncation method. First, we wanted to examine one of the outlier exclusion methods rather than one of the robust estimators (e.g., median). Existing RT models typically make predictions about traditional summary statistics (e.g., mean or variance of RT), because such predictions require the fewest distributional assumptions (cf. Sternberg, 1969). Much stronger assumptions are needed to make predictions about robust estimators (e.g., that median RT should increase linearly across conditions), so these estimators are generally only useful in making qualitative comparisons among conditions, not for testing models. Moreover, no single robust estimation method can be used for all the different RT analyses in which a researcher might be interested, because one needs a different robust estimator for each type of qualitative comparison (e.g., central tendency or variability). After outlier exclusion, in contrast, it is possible to compute any desired summary of the remaining observed RTs (e.g., mean or variance), or even to examine their full probability distribution. This generality is particularly important as RT models become more sophisticated, because model tests may require simultaneous examination of several summary measures, full probability distributions of RT, or both (e.g., Meyer, Irwin, Osman, & Kounios, 1988; Miller, 1982; Roberts & Sternberg, 1992). Second, truncation is relatively tractable mathematically, because it does not depend on the RTs in the

sample (as does, e.g., exclusion of values more than a certain number of standard deviations from the mean). Third, and most important, informal examination of the literature indicates that truncation is quite common; it is nearly the standard procedure in some areas of RT research (cf. Jensen, 1987, p. 110). Even if another method turns out to be unequivocally superior, it will still be useful to understand truncation effects to interpret the previous studies that have used this method.

Although the present concerns may seem inappropriate for an article in a psychological journal, the statistics literature has not adequately addressed these issues as they apply to RT research. Statistical treatment of outliers may be traced back to the beginning of the 18th century (cf. Anscombe, 1960), but optimal, general techniques have not been identified, at least partly because the concept of an outlier is subjective and inherently somewhat ambiguous (cf. Beckman & Cook, 1983). The statistical literature contains many detailed comparisons of the dozens of alternative outlier exclusion methods that have been proposed (e.g., Andrews et al., 1972; for reviews, see Barnett & Lewis, 1984; Beckman & Cook, 1983; Hawkins, 1980; Lovie, 1986; Rousseeuw & Leroy, 1987), but their applicability to the RT domain is limited for two reasons. First, almost all the comparisons assume symmetric (i.e., normal) underlying distributions (cf. Beckman & Cook, 1983), but RT distributions are skewed (McCormack & Wright, 1964) and skewness is known to modulate the effects of outlier exclusion (e.g., Miller, 1988, 1991). These methods might be useful if RTs could be transformed to achieve normality, but the appropriate transformation is not known. There are a few methods appropriate for nonnormal distributions, but these require specific assumptions that are unlikely to be met in RT research. Moreover, these methods often require large sample sizes for effective use (Huber, 1981), and large samples are often unavailable in RT research, given that outlier rejection must be done separately for each subject and condition. Second, within the statistical literature, the emphasis has been on finding the most effective procedure for outlier exclusion, that is, the procedure that eliminates the most spurious observations and retains the most valid ones. For RT research, however, the question of bias is surely even more important, because experimenters must avoid statistical artifacts that create or conceal small yet theoretically important differences. Indeed, the results of Ratcliff (1979, pp. 456–458) suggest that such biasing effects may be quite substantial. He examined the effects of truncating observed RT distributions on the estimates of their higher moments. In these distributions, truncating at 5 s eliminated about 1% of the RTs, whereas truncating at 2 s eliminated 4%. The means were about 20–50 ms smaller with the 2-s truncation cutoff than with the 5-s cutoff, variances were about half, and the third and fourth moments were approximately an order of magnitude smaller. Thus, his study indicated that truncation can have a large impact on RT analyses. One of our goals in the present article, then, is to pursue Ratcliff's observations in a more theoretical fashion and a somewhat broader framework.

Data Analysis Under Truncation, Contamination, or Both

To evaluate truncation effects, we consider three distinct scenarios for data analysis. In the first, called Truncation, No Contamination, truncation eliminates all spurious RTs, so no spurious observations contaminate the analysis. However, truncation also eliminates some valid RTs because of the long tails on RT distributions. Under this scenario, then, the issue is whether the exclusion of some valid RTs introduces substantial biases into the results, and, if so, what those biases are. This scenario merits examination because (a) the researcher may succeed in eliminating all spurious RTs by truncation or (b) there may be no spurious RTs in the first place. Of course, researchers would not truncate if they knew there were no spurious RTs, but they could never be certain of this in practice and might therefore truncate "to be safe" or in imitation of common practice. Our analyses indicate that substantial biases can arise under this scenario.

In a second scenario, called Truncation Plus Contamination, truncation not only eliminates some valid RTs but also fails to eliminate some spurious ones, which therefore contaminate the analysis. This scenario is generally more realistic than the first, because not all spurious RTs necessarily have sufficiently extreme values to be truncated. Unfortunately, it is also more complicated than the first, because effects of contamination must be considered in addition to those of truncation. An optimist might hope that the contaminating RTs would roughly compensate for the valid RTs that are eliminated, leaving approximately unbiased results. Unfortunately, our results suggest that this view is far too optimistic in most cases.

Finally, for comparison purposes, we considered a baseline scenario, called No Truncation, Contamination, in which neither truncation nor any other form of outlier exclusion is used. This scenario should be evaluated formally, because it seems clear that truncation should not be used if it produces more bias than the spurious RTs it is intended to eliminate. Perhaps surprisingly, we find that this option is sometimes best (i.e., the treatment of truncation is worse than the disease of spurious observations).

Two other logically possible scenarios were not examined. One is the scenario of perfect truncation, which eliminates all spurious RTs and retains all valid ones. Because this procedure is clearly superior to any of the others and unattainable in practice, there is no need to consider it formally. The second is that in which truncation retains all valid RTs but fails to eliminate some spurious ones. As noted above, most current RT models predict that RT distributions have an infinite tail in at least one direction, at least theoretically, so this scenario is impossible to analyze with current RT models.

Overview of the Article

The remainder of this article is organized as follows. The next section presents the basic definitions and formulas used in our analyses of the three scenarios just outlined. The basic ideas of this section are well-known in the statistical literature, and derivations of the formulas may be found in standard texts such as Everitt and Hand (1981), Mood, Graybill,

and Boes (1974), and Parzen (1962). We make no attempt to present the most general versions of the definitions and formulas but instead give versions specifically tailored to the analyses conducted here.

After the formulas have been presented, the next section (Bias Effects of Truncation on Mean, Median, Standard Deviation, and Skewness) investigates the biasing effects of truncation on the four indicated summary measures of RT distributions. The following section (Effects of Truncation on Tests of Models) examines effects on two specific and very common types of model-based RT analyses. One subsection, *Effects of Truncation on Linear Increases in RT*, deals with effects on experiments investigating the relation between RT and a quantitative independent variable (e.g., number of items in a visual display or memory set and angle of rotation from upright). These experiments often evaluate the fit of linear functions relating RT to the independent variable, and it is of interest to determine how predictions of linearity may need to be modified when outliers are excluded. Similarly, the other subsection *Effects of Truncation on Additivity of Factor Effects* deals with effects in factorial experiments. In many experiments, additivity or interaction of the factor effects is of critical theoretical importance, and so it is of interest to see how truncation changes the predictions of models associated with these outcomes (cf. Townsend, 1992). In subsequent sections, we also consider the effects of truncation on hazard functions and statistical power, and we examine the ability of a maximum likelihood estimation procedure to counteract the deleterious effects of truncation.

Many researchers may be more interested in the practical implications of these analyses than in the analyses themselves. To assist these researchers, we present in Appendix A a set of specific recommendations on how to analyze RT when some truncation has occurred. These appear to us to be the most appropriate heuristics available, given the results of the analyses presented here.

Definitions, Notation, and Formulas

This section presents the basic formalisms underlying the analyses presented in the remainder of the article. The definitions and notation are necessary to all readers, but the formulas can be skipped by those who are only interested in qualitative summaries of the results.

Table 1 summarizes the notation. The left-most column shows the symbols for an observed RT and various properties of its distribution: the probability density function (PDF), the cumulative probability density function (CDF), the mean, the variance, and the hazard function.² The symbols in the other columns are described next. Throughout the text, we follow the usual convention of using boldface letters (e.g., T_o) to denote random variables.

² The hazard function of a distribution is considered only in the section *How Does Truncation Distort Hazard Functions?*, and unfamiliarity with this function will not cause difficulty in understanding the other sections.

Table 1
Symbols Used to Refer to Untruncated and Truncated Distributions of Observed RTs and to the Underlying Distributions of Valid and Spurious RTs Making Up the Observed RTs

Symbol for	RT distributions				
	Observable		Underlying		
	Untruncated	Truncated	Spuriously fast	Valid	Spuriously slow
RT	T_o	T_{*o}	T_f	T_v	T_s
PDF	$f_o(t)$	$f_{*o}(t)$	$f_f(t)$	$f_v(t)$	$f_s(t)$
CDF	$F_o(t)$	$F_{*o}(t)$	$F_f(t)$	$F_v(t)$	$F_s(t)$
Mean	$E[T_o]$	$E[T_{*o}]$	$E[T_f]$	$E[T_v]$	$E[T_s]$
Variance	$Var [T_o]$	$Var [T_{*o}]$	$Var [T_f]$	$Var [T_v]$	$Var [T_s]$
Hazard function	$h_o(t)$	$h_{*o}(t)$	$h_f(t)$	$h_v(t)$	$h_s(t)$
Mixture probability	—	—	g_f	$1-g_f-g_s$	g_s
Acceptance probability	a_o	1.0	a_f	a_v	a_s

Note. Symbols are used without subscripts [e.g., $f(t)$] in general formulas that are true for each of the distributions under consideration (e.g., Equations 1–6). For example, the symbol f in the main text could refer to either $f_o, f_f, f_v,$ or f_s , whereas the symbol f_* refers to one of their truncated versions $f_f, f_v,$ and f_s . The other symbols without subscripts (e.g., $a, F,$ and h) can be interpreted analogously. RT = reaction time; PDF = probability density function; CDF = cumulative probability density function.

Truncation, No Contamination

Definitions are as follows: Using *truncation*, the researcher chooses lower (L) and upper (U) cutoffs for RT and discards from the sample any observed RT that is less than L or greater than U .³ RTs between the cutoffs are referred to as *accepted RTs*, and the probability that an RT falls within the cutoffs is referred to as the *acceptance probability*.

For example, the solid curve in the upper panel of Figure 1 shows a probability distribution of observable RTs, and the vertical lines indicate truncation cutoffs that a researcher might adopt. The dashed line shows the resulting probability distribution after truncation with those cutoffs, which in this case eliminated the bottom 1.3% and the top 5% of the untruncated distribution. Clearly, truncation changes the RT distribution, except when the entire distribution lies within the truncation cutoffs, and may have various effects on its mean, variance, and so forth, depending on the shape of the distribution and the placement of the cutoffs. In this example, the mean and standard deviation of the untruncated distribution are 622 and 201, respectively, whereas the corresponding values for the truncated distribution are 593 and 132.

We make use of several well-known formulas, shown below, to compute properties of truncated distributions from those of their untruncated counterparts. Subscripts are omitted from the terms in the formulas [e.g., $f(t)$], because the formulas apply to all distributions (e.g., valid RTs and observed RTs). To simplify the notation, we use an asterisk to indicate all truncated distributions and their associated variables, moments, densities, and so forth. For example, the second column from the left in Table 1 illustrates the notation for RTs from the truncated distribution of observable RTs. For brevity, the means, PDFs, and so forth of such distributions are referred to as truncated means, truncated PDFs, and so on.

With lower and upper truncation cutoffs of L and U , respectively, the acceptance probability for a truncated density is

$$a = F(U) - F(L) \tag{1}$$

and the truncated density is given by

$$f_*(t) = \frac{f(t)}{a}, \tag{2}$$

with cumulative form

$$F_*(t) = \frac{F(t) - F(L)}{a}, \tag{3}$$

mean

$$E[T_*] = \frac{\int_L^U tf(t)dt}{a}, \tag{4}$$

and variance

$$Var[T_*] = \frac{\int_L^U t^2f(t)dt}{a} - E[T_*]^2. \tag{5}$$

Finally, the median of T_* is given by

$$Med[T_*] = F^{-1} \left[\frac{F(L) + F(U)}{2} \right], \tag{6}$$

where F^{-1} is the inverse of F (see Appendix B for proof).

³ Some RT researchers rerun truncated trials rather than simply throwing them away. This strategy, of course, does not eliminate the statistical effects of outlier exclusion, but merely serves to ensure the desired sample size in each condition.

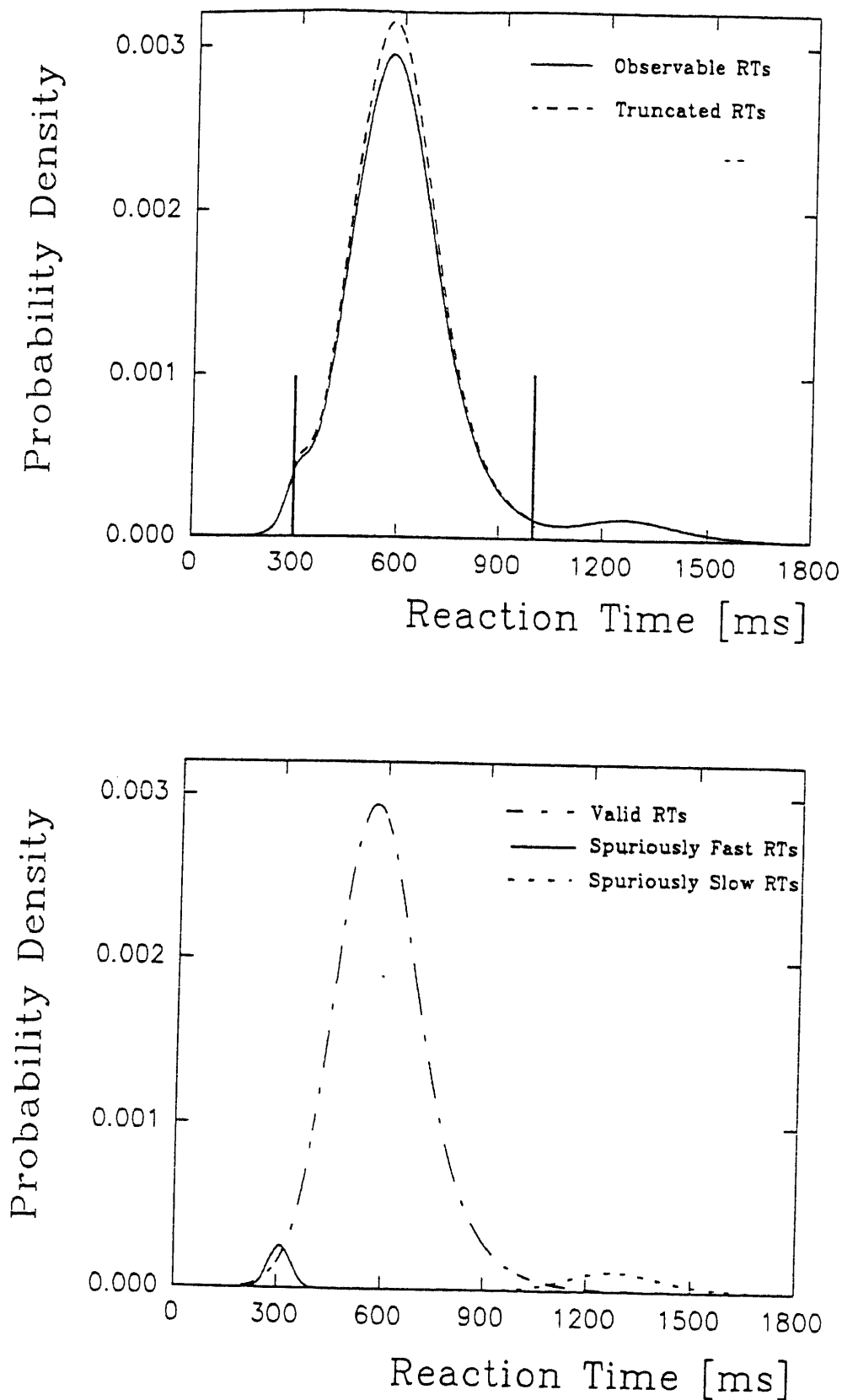


Figure 1. An example illustrating the effects of truncation and spurious observations. Upper panel: A distribution of observable reaction times (RTs, solid curve) and the truncated version of that distribution (dashed curve) obtained after truncation at the indicated cutoffs (vertical lines). Lower panel: The left-most, middle, and right-most curves show the relative frequencies of spuriously fast, valid, and spuriously slow RTs, respectively, contributing to the observable RT distribution (solid line) in the upper panel.

No Truncation, Contamination

As is standard within the statistical literature (e.g., Andrews et al., 1972), we used a mixture model to study the contaminating effects of spurious observations.

The definition of a mixture model is as follows: An observable RT distribution is a *mixture distribution* if the RTs come from different underlying probability distributions on different trials (cf. Luce, 1986; Meyer, Yantis, Osman, & Smith, 1985; Yantis, Meyer, & Smith, 1991).

In this article, we consider observed RT distributions that are mixtures of three underlying distributions: spuriously fast RTs, valid RTs, and spuriously slow RTs, as shown in the lower panel of Figure 1. The valid RTs come from trials on which the subject was performing the task as intended, and these RTs reflect the psychological processes under study. The spuriously fast and slow RTs come from trials on which the subject was not performing as intended, and was instead doing something that led to a relatively fast or slow response.⁴ Notation for these three types of distributions is given in the right side of Table 1. To express the fact that the observable RTs are a mixture of the spuriously fast, valid, and spuriously slow RTs, we write,

$$\mathbf{T}_o = \begin{cases} \mathbf{T}_f & \text{with probability } g_f \\ \mathbf{T}_v & \text{with probability } 1 - g_f - g_s, \\ \mathbf{T}_s & \text{with probability } g_s \end{cases} \quad (7)$$

which is more simply written as

$$\mathbf{T}_o = \text{mix}(\mathbf{T}_f, \mathbf{T}_v, \mathbf{T}_s). \quad (8)$$

Each distribution underlying the observable mixture (i.e., spuriously fast, valid, and spuriously slow) has an associated *mixture probability*, which is the probability that an observed RT comes from that distribution. As indicated in Table 1, the mixture probabilities g_f , $(1 - g_f - g_s)$, and g_s associated with these three underlying distributions must sum to 1.0, because each observed RT must be one of the three types. In Figure 1, the mixture probability associated with each underlying distribution is equal to the area under that distribution. As is visually suggested by the figure, the observable mixture distribution is most strongly determined by the underlying distribution with the highest probability.

The mixture distribution of observable RTs is uniquely determined by the three underlying distributions and their associated mixture probabilities, and many properties of the observable mixture distribution can be calculated directly. For example, the i th raw moment of the observable mixture distribution is given by

$$E[\mathbf{T}_o^i] = g_f \cdot E[\mathbf{T}_f^i] + (1 - g_f - g_s) \cdot E[\mathbf{T}_v^i] + g_s \cdot E[\mathbf{T}_s^i], \quad (9)$$

from which the mean and variance may be computed as

$$E[\mathbf{T}_o] = g_f \cdot E[\mathbf{T}_f] + (1 - g_f - g_s) \cdot E[\mathbf{T}_v] + g_s \cdot E[\mathbf{T}_s] \quad (10)$$

and

$$\begin{aligned} \text{Var}[\mathbf{T}_o] = & g_f \cdot E[\mathbf{T}_f]^2 + (1 - g_f - g_s) \cdot E[\mathbf{T}_v]^2 + g_s \cdot E[\mathbf{T}_s]^2 \\ & - \{g_f \cdot E[\mathbf{T}_f] + (1 - g_f - g_s) \cdot E[\mathbf{T}_v] + g_s \cdot E[\mathbf{T}_s]\}^2 \\ & + (1 - g_f - g_s) \cdot \text{Var}[\mathbf{T}_v] + g_f \cdot \text{Var}[\mathbf{T}_f] + g_s \cdot \text{Var}[\mathbf{T}_s]. \end{aligned} \quad (11)$$

We also make use of the PDF and CDF of the observable mixture distribution, given by the following formulas:

$$f_o(t) = g_f \cdot f_f(t) + (1 - g_f - g_s) \cdot f_v(t) + g_s \cdot f_s(t), \quad (12)$$

and

$$F_o(t) = g_f \cdot F_f(t) + (1 - g_f - g_s) \cdot F_v(t) + g_s \cdot F_s(t). \quad (13)$$

Truncation Plus Contamination

In principle, there is nothing conceptually special about the case of truncated RT distributions contaminated with some spurious observations. A truncated mixture distribution is the same as a mixture of the truncated versions of its underlying distributions, so one can either apply the formulas for truncation to mixture distributions or apply the mixture formulas to truncated versions of the underlying distributions, with identical results. In practice, however, the formulas become rather cumbersome, as shown next. We again use the asterisk to indicate truncated versions of the distributions underlying the mixture, so, for example, \mathbf{T}_{*v} is an RT from the truncated distribution of valid RTs.

With lower and upper truncation cutoffs of L and U , respectively, the PDF of the observable mixture distribution of RTs is

$$f_{*o}(t) = \frac{g_f \cdot f_f(t) + (1 - g_f - g_s) \cdot f_v(t) + g_s \cdot f_s(t)}{g_f \cdot a_f + (1 - g_f - g_s) \cdot a_v + g_s \cdot a_s} \quad (14)$$

for $L \leq t \leq U$, and zero elsewhere. The associated CDF is

$$F_{*o}(t) = \frac{g_f \cdot [F_f(t) - F_f(L)] + (1 - g_f - g_s) \cdot [F_v(t) - F_v(L)] + g_s \cdot [F_s(t) - F_s(L)]}{g_f \cdot a_f + (1 - g_f - g_s) \cdot a_v + g_s \cdot a_s}. \quad (15)$$

Integrating the PDF, the i th raw moment (from which mean, variance, and skewness may be computed) of the truncated distribution of observable RTs is

$$E[\mathbf{T}_{*o}^i] = \frac{g_f \cdot \int_L^U t^i f_f(t) dt + (1 - g_f - g_s) \cdot \int_L^U t^i f_v(t) dt + g_s \cdot \int_L^U t^i f_s(t) dt}{g_f \cdot a_f + (1 - g_f - g_s) \cdot a_v + g_s \cdot a_s}. \quad (16)$$

⁴ Despite the fact that we consider only three underlying distributions, our analyses do not require the assumption that all RTs from a given underlying distribution are generated with the subject in the same state. The underlying distribution of valid RTs may itself be a mixture distribution, as has sometimes been assumed (e.g., Luce, 1986; Meyer et al., 1985), and so may the underlying distributions of spurious RTs. The practical implication of this is that, whenever our analyses do not assume specific shapes for the underlying distributions, the conclusions are valid even if there are multiple types of spuriously fast, valid, and spuriously slow RTs.

Bias Effects of Truncation on Mean, Median, Standard Deviation, and Skewness

In this section we consider how four common summary measures of RT distributions are influenced by truncation. Initially, we investigate the behavior of these summary measures under the scenario of Truncation, No Contamination to see how model predictions change when valid but extreme RTs are eliminated. Subsequently, we consider the more realistic scenario in which some spurious RTs remain after truncation (i.e., Truncation Plus Contamination).

It is clear on intuitive grounds that truncation can influence summary measures of a distribution. For example, if one eliminates responses slower than a certain cutoff, the mean or median of the truncated distribution will certainly be smaller than the mean or median of the original (untruncated) distribution, although the change in the median will generally be smaller than that in the mean. Similarly, the standard deviation of a distribution decreases when observations from either one or both tails are excluded, and skewness also decreases under these conditions, at least if approximately equal proportions of observations are excluded from each end of the distribution.

Although the directions of these biases are easily understood, the sizes are not intuitively clear. Because researchers derive predictions of RT models from assumptions about untruncated RT distributions (e.g., Luce, 1986; Townsend & Ashby, 1983), it is important to ascertain the size of the effects of truncation on the various summary measures. Basically, the issue is whether truncation biases are so small that they can safely be ignored, or so large that they must be considered when deriving and testing predictions of RT models.

Unfortunately, it is impossible to say exactly how large truncation effects are in general, because these effects depend on the exact circumstances. Obviously, the effect of truncation on a given parameter must depend a great deal on the shape of the RT distribution and on the percentage of observations truncated from each tail of the distribution. The goal of this section, therefore, is to sharpen our intuitions by investigating the size of truncation effects in a range of plausible cases.

We examine the effects of truncation on three underlying RT distributions: the special Erlangian, the lognormal, and the ex-Gaussian. Like empirical RT distributions, these distributions are all skewed, with long tails extending into the range of slower responses, and all have been found to give particularly good fits to empirical RT distributions (Heathcote, Popiel, & Mewhort, 1991; Hockley, 1984; Hohle, 1965; Luce, 1986; Ratcliff & Murdock, 1976; Woodworth & Schlosberg, 1954).

The biases due to truncation can be evaluated by comparing the properties of the truncated distribution with those of the untruncated distribution. It can be seen, however, that the former (e.g., Equations 4, 5, and 6) are too general to evaluate without specific distributional assumptions, so we proceed to examine some particular cases. As the results clearly depend on the truncation values L and U , we examined a range of such values.

The Special Erlangian Distribution

A suitable balance between mathematical tractability and correspondence to observations is offered by the special Erlangian distribution, a special case of the gamma distribution in which the shape parameter m is restricted to integer values. This distribution corresponds to the sum of m independent random variables having a common exponential distribution. The special Erlangian distribution is a popular tool in psychology, having been used in modeling a variety of time-dependent processes (e.g., Luce, 1986; McGill, 1963; McGill & Gibbon, 1965; Ratcliff & Murdock, 1976; Townsend & Ashby, 1983; Ulrich & Wing, 1991).

The PDF of the special Erlangian is given for $t > 0$ by

$$f(t) = \frac{\lambda e^{-\lambda t} (\lambda t)^{m-1}}{(m-1)!}, \quad (17)$$

where λ is a positive real, m a positive integer, and t denotes the time.⁵ The parameter λ is only affected by the unit of time and hence is called the *scale parameter* of the special Erlangian. In contrast, the parameter m is unaffected by the unit of time but controls the shape of the special Erlangian, so it is called the *shape parameter*. With $m = 1$, the distribution is simply an exponential, with a probability density that is maximum at $t = 0$ and decreases monotonically as t increases. For $m > 1$, the distribution is somewhat bell-shaped, with probability density first increasing then decreasing as t increases. For $m = 2$ the distribution is highly skewed, with the mode and median equal to approximately 50% and 84% of the mean, respectively. As m increases, the shape of the special Erlangian distribution rapidly approaches that of the normal distribution, with a mode and median of approximately 90% and 97% of the mean for $m = 10$. The mean and variance are $E[\mathbf{T}] = (m/\lambda)$ and $\text{Var}[\mathbf{T}] = (m/\lambda^2)$, respectively, and the CDF is given by

$$F(t) = 1 - \sum_{r=0}^{m-1} \frac{e^{-\lambda t} (\lambda t)^r}{r!}. \quad (18)$$

When the untruncated distribution is a special Erlangian, the expected raw moments of \mathbf{T}_* are easily calculated using Proposition 1: If \mathbf{T} follows a special Erlangian distribution with parameters m and λ , then the i th raw moment of \mathbf{T}_* is given by

$$E[\mathbf{T}_*^i] = \frac{(m+i-1)!}{(m-1)!} \lambda^{-i} \frac{G(U) - G(L)}{F(U) - F(L)}, \quad (19)$$

where L and U are the lower and upper truncation points,

⁵ It may seem that the special Erlangian is a poor model for RTs, even with $m > 1$, because the density is positive for all values of t greater than 0. This problem is easily avoided by a generalization of Equation 17, in which t is replaced by $t - r_0$, where $r_0 > 0$ represents an irreducible minimum RT; that is, the response can never occur before r_0 units of time have elapsed. The consequences of this generalization are trivial in most cases; the nontrivial cases are considered further on.

respectively. Function F is the CDF of a special Erlangian with parameters λ and m , and function G is the CDF of a special Erlangian with scale parameter λ and shape parameter $(m + i)$. (Proof of Proposition 1 is given in Appendix B.)

From Equation 19 one can easily compute central moments as needed. For example, the variance of \mathbf{T}_* is

$$\text{Var}[\mathbf{T}_*] = E[\mathbf{T}_*^2] - E[\mathbf{T}_*]^2. \quad (20)$$

We defined skewness itself as

$$\text{Skew}[\mathbf{T}_*] \equiv \sqrt[3]{E[(\mathbf{T}_* - E[\mathbf{T}_*])^3]}. \quad (21)$$

Although this is not the most conventional measure of skewness, it provides a skewness measurement with the same units as the three other summary measures (i.e., mean, median, and standard deviation) that we consider.

Using Proposition 1, we investigated the influence of truncation on special Erlangian distributions with $m = 2, \dots, 35$. For computational purposes, the scale parameter λ was adjusted in such a way that the expected mean of \mathbf{T} was always equal to 150 ms, although, as described below, the final results are independent of the value of the scale parameter λ .

For each value of the shape parameter m , the bias of the mean, median, standard deviation, and skewness were computed as a function of the percentages p and q eliminated from the lower and upper tails of f , respectively. These percentages were varied orthogonally from 0% to 16%. The truncation cutoffs L and U corresponding to p and q , respectively, were determined with a numerical search algorithm using Equation 18. These cutoffs were then inserted into Equation 19 to obtain the first, second, and third raw moments, from which the first three central moments (mean, variance, and skewness) of \mathbf{T}_* were computed. The median of \mathbf{T}_* was computed on the basis of Equations 6 and 18.

To summarize the computations, we computed a measure of the *relative bias* introduced by truncation for each summary measure (i.e., mean, median, standard deviation, skewness) as a function of p and q . For example, the relative bias, δ , of the mean $E[\mathbf{T}_*]$ was defined as

$$\delta(p, q) \equiv \frac{E[\mathbf{T}_*] - E[\mathbf{T}]}{E[\mathbf{T}]} \cdot 100\%. \quad (22)$$

Note that a negative (positive) value of δ indicates that the parameter of \mathbf{T}_* underestimates (overestimates) the corresponding parameter of \mathbf{T} . The relative biases of the median, standard deviation, and skewness were defined analogously.

The upper panels of Figures 2–5 summarize the results of the bias computations, showing the relative bias of each summary measure as a function of p and q . The panels on the right side of each figure show the results for $m = 2$ and the panels on the left, $m = 10$. (Note that the ordinates of these four figures have different scales.)

Mean. As expected, the mean of \mathbf{T}_* overestimates (underestimates) the mean of \mathbf{T} if the distribution f is truncated from below (above). Underestimation is also obtained if f is truncated from below and above to the same extent, $p = q$,

because of the skew in f , and this underestimation increases with p . The underestimation diminishes as the shape parameter m increases, because f becomes more symmetrical. It is important to note that even a small value of q can introduce a meaningful bias; with $m = 2$, for example, a bias of -5.1% is observed if the upper 2% of f is truncated (i.e., $p = 0\%$ and $q = 2\%$). It is interesting to note that p and q exert nearly additive effects on δ , as reflected in the near-parallel lines.

Median. In general, p and q affect the median less than the mean, precisely because the median is less affected by skewness in the first place. In general, p and q produce nearly linear and additive effects, and, as was true for the mean, the bias of the median decreases as f becomes more symmetrical. Unlike the mean, truncating f by the same amount from below and above introduces no bias at all into the median.

Standard deviation. As is intuitively evident, $\text{SD}[\mathbf{T}_*]$ underestimates the true standard deviation whenever truncation is used. In general, bias is larger if f is truncated at its upper tail than its lower tail, and even a small value of q produces a meaningful bias. The effect of q on δ is slightly stronger for smaller than for larger values of m , whereas the reverse is true for p .

Skewness. For $m = 2$, skewness is nearly unaffected if f is truncated from below, but it is strongly affected if f is truncated from above. For $m = 10$, the truncation percentages p and q produce a bizarre interaction: The effect of q is much larger when p is small than when p is large.

Irreducible minimum RT and relative bias. It should be noted that, although the relative biases of the mean and median are independent of the units of measurement of RT, they would change if a positive constant r_0 were added to the random variable \mathbf{T} . The constant r_0 may be conceived as the *irreducible minimum RT* (Woodworth & Schlosberg, 1954), reflecting some relatively invariant sensory or motor time. A rather conservative estimate of r_0 is 100 ms (cf. Luce, 1986, p. 59–62). The constant r_0 would increase the denominator but not the numerator of Equation 22, thereby decreasing the relative biases of both mean and median. In the above bias computations, we assumed $r_0 = 0$, and hence the analyses yielded the maximum possible relative biases of the mean and median.

It is easy to use the above computations even when $r_0 > 0$. For example, the relative $\delta(p, q)$ can be used to compute the untruncated mean from the truncated mean by rearranging the definition of Equation 22, namely

$$E[\mathbf{T}] = \frac{E[\mathbf{T}_*]}{1 + \delta(p, q)/100}. \quad (23)$$

To correct for $r_0 > 0$, Equation 23 should be rewritten as

$$E[\mathbf{T}] = r_0 + \frac{E[\mathbf{T}_*] - r_0}{1 + \delta(p, q)/100}. \quad (24)$$

As relative biases of medians were also computed assuming $r_0 = 0$, Equation 24 should also be used in that case, replacing means with medians. Relative biases of standard deviations and skewness were computed without this assumption, however, so a comparable adjustment is not needed and an equation analogous to Equation 23 can be used to correct these summary measures. (The same consideration applies to the

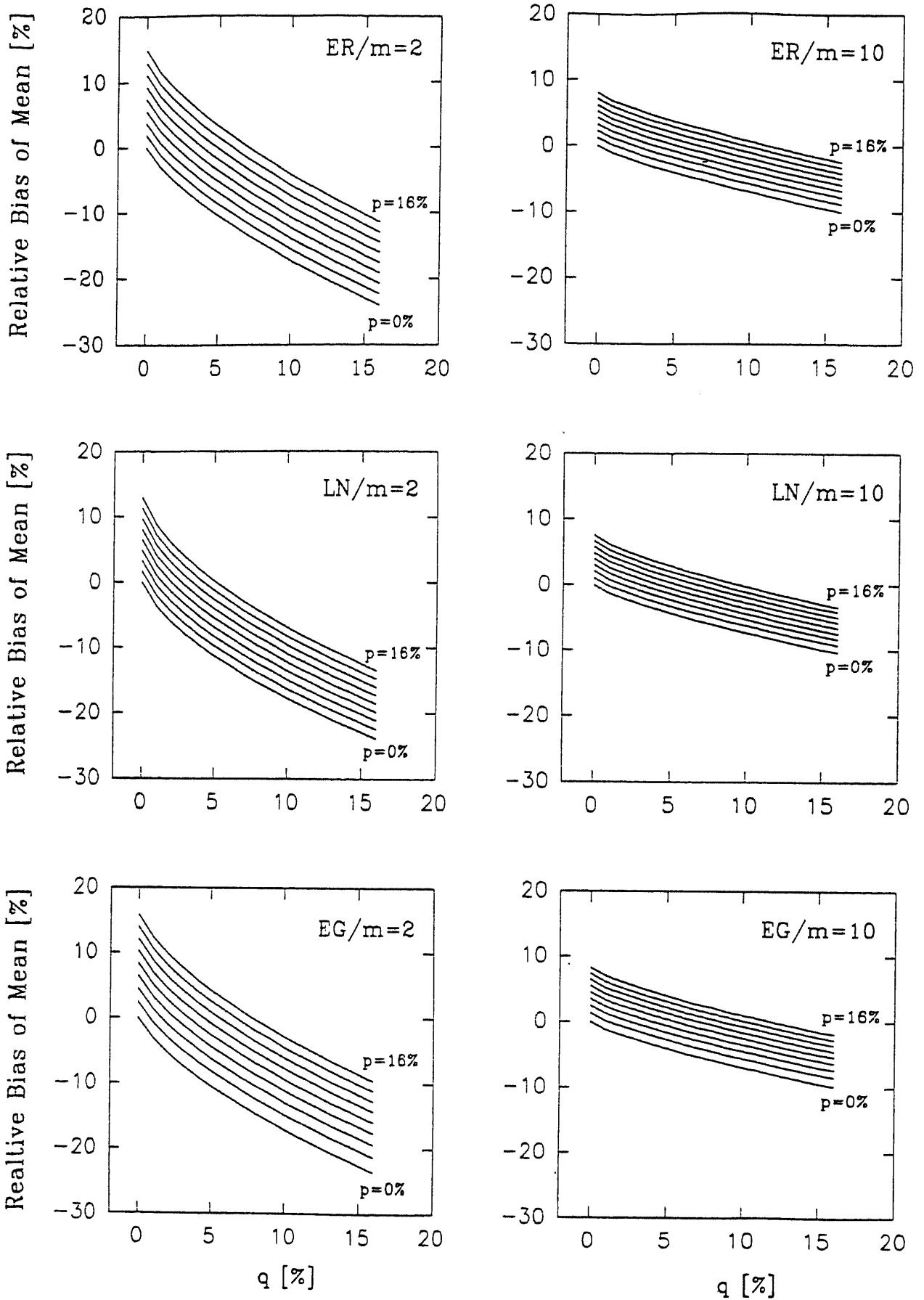


Figure 2. Relative bias of mean as a function of p (percentage of observations truncated from lower tail of f) and q (percentage truncated from upper tail of f). Within each panel, each line denotes a different value of $p = 0, 2, 4, 6, 8, 10, 12, 14, 16$, but only the most extreme values (i.e., $p = 0\%$ and $p = 16\%$) are indicated because the effects of p are monotonic from $p = 0$ to $p = 16$. ER = Erlangian distribution; LN = lognormal distribution; EG = ex-Gaussian distribution.

Table 2

Constants A, B, C, and D for the Best Fitting Power Function Approximation $\hat{\delta}(p, q)$ of Mean Reaction Time as a Function of m and of Theoretical Distribution

m	Erlangian distribution				Lognormal distribution				Ex-Gaussian distribution			
	A	B	C	D	A	B	C	D	A	B	C	D
2	1.015	.943	-3.045	.762	.921	.918	-3.730	.689	1.371	.862	-3.176	.739
3	.951	.925	-2.335	.775	.851	.913	-2.809	.716	1.230	.855	-2.467	.746
4	.894	.914	-1.942	.783	.797	.909	-2.300	.732	1.114	.853	-2.063	.751
5	.844	.906	-1.688	.788	.754	.905	-1.974	.744	1.025	.852	-1.799	.756
6	.801	.900	-1.508	.793	.718	.901	-1.745	.752	.956	.851	-1.608	.760
7	.764	.895	-1.372	.796	.687	.898	-1.574	.759	.898	.850	-1.464	.763
8	.732	.891	-1.265	.799	.660	.896	-1.440	.764	.849	.850	-1.349	.766
9	.704	.888	-1.178	.801	.637	.894	-1.333	.769	.808	.849	-1.257	.768
10	.679	.886	-1.706	.803	.616	.892	-1.245	.773	.772	.849	-1.178	.771
11	.656	.884	-1.045	.804	.597	.890	-1.170	.776	.742	.849	-1.112	.773
12	.636	.882	-.993	.806	.580	.888	-1.107	.779	.714	.849	-1.056	.775
13	.617	.880	-.947	.807	.565	.887	-1.052	.782	.690	.848	-1.007	.776
14	.600	.878	-.907	.808	.551	.886	-1.004	.784	.668	.848	-.963	.779
15	.585	.877	-.871	.809	.538	.884	-.961	.786	.648	.848	-.926	.780
20	.523	.872	-.738	.813	.485	.880	-.804	.794	.570	.848	-.782	.786
25	.477	.869	-.650	.816	.446	.876	-.702	.799	.513	.847	-.683	.791
30	.443	.866	-.586	.818	.415	.873	-.629	.803	.473	.847	-.616	.795
35	.415	.864	-.538	.820	.391	.871	-.574	.806	.439	.847	-.563	.797

Note. The correlation coefficient of each approximation is better than 0.9997, 0.9996, and 0.9998 for the special Erlangian, lognormal, and the ex-Gaussian distributions, respectively. The mean absolute deviation between approximated and true bias is .04%, .05%, and .04% for the special Erlangian, lognormal, and ex-Gaussian distributions, respectively.

lognormal and ex-Gaussian distributions discussed in subsequent sections.)

A useful approximation for relative bias. Appendix A describes procedures that can be used to counteract some of the effects of the substantial truncation biases evident in Figures 2-5. Because these procedures require many relative bias computations, it is convenient to have a simple function that can be used to compute approximate relative bias, to avoid the laborious computations needed to obtain the values shown in Figures 2-5. Fortunately, we found that the power function

$$\hat{\delta}(p, q) = A \cdot p^B + C \cdot q^D \quad (25)$$

provides an excellent approximation of $\delta(p, q)$ for the mean, median, and standard deviation shown in Figures 2-4, although not for skewness (Figure 5). The fitted power coefficient B (D) is equal to 1 if p (q) has a linear effect on δ , less than 1 if the effect is sublinear, and larger than 1 if the effect is superlinear. Tables 2, 3, and 4 provide the numerical values of A , B , C , and D for the best-fitting approximations to the relative biases of the mean, standard deviation, and median, respectively, for each value of the shape parameter m . (The tables also include results for the lognormal and ex-Gaussian distributions, which are discussed next.) These values are needed for the procedures presented in Appendix A.⁶

Unfortunately, Equation 25 cannot be used to approximate the relative bias of skewness. However, one may use $(E[\mathbf{T}_*] - \text{Med}[\mathbf{T}_*])/SD[\mathbf{T}_*]$ as an alternative measure to assess truncation effects on skewness. This alternative measure lies between -1 and +1 (cf. Stuart & Ord, 1987, p. 116), although

values between 0 (i.e., symmetrical distribution) and 0.31 (i.e., exponential distribution) are to be expected for RTs.

The Lognormal Distribution

Another useful model of RT distributions is the so-called lognormal distribution, which has given good fits in several RT analyses (e.g., Bree, 1975; Hockley, 1984; Ratcliff & Murdock, 1976; Woodworth & Schlosberg, 1954) and which could be generated by any of several plausible stochastic latency mechanisms (Ulrich & Miller, in press).

The lognormal distribution is usually defined by saying that the time \mathbf{T} follows a lognormal distribution if the logarithm $\ln(\mathbf{T})$ of time \mathbf{T} is normally distributed with mean $E[\ln(\mathbf{T})] = \mu$ and standard deviation $SD[\ln(\mathbf{T})] = \sigma$. Therefore, \mathbf{T} is necessarily a positive random variable. The lognormal distribution is usually considered as a possible model in statistics whenever a random variable with a positive skewness is needed (e.g., Crow & Shimizu, 1988).

⁶ To ensure that the approximations would be adequate across the range $0 < p < 10\%$ and $0 < q < 10\%$, we computed values of A-D that minimized the sum of squared errors of the approximations for p and q values in this range. As an example, consider the computation of the relative bias of the mean for the special Erlangian with $m = 2$, $p = 4\%$, and $q = 8\%$. On the basis of Table 2 one computes the approximation $\hat{\delta}(4, 8) = 1.015 \cdot 4^{.943} - 3.045 \cdot 8^{.762} = -11.1$. For practical work, this is clearly a satisfactory approximation to the true relative bias of $\delta(4, 8) = -11.2$.

Table 3

Constants A, B, C, and D for the Best Fitting Approximation $\hat{\delta}(p, q)$ of Standard Deviation as a Function of m and of Theoretical Distribution

m	Erlangian distribution				Lognormal distribution				Ex-Gaussian distribution			
	A	B	C	D	A	B	C	D	A	B	C	D
2	-0.320	1.081	-8.932	.588	-0.148	1.177	-14.695	.486	-0.991	.829	-9.686	.562
3	-0.626	.958	-7.757	.603	-0.345	1.025	-12.045	.519	-1.388	.792	-8.673	.566
4	-0.859	.904	-7.080	.613	-0.517	.959	-10.544	.540	-1.624	.775	-8.043	.570
5	-1.040	.873	-6.629	.620	-0.662	.922	-9.565	.555	-1.788	.765	-7.605	.574
6	-1.186	.852	-6.303	.626	-0.786	.897	-8.869	.567	-1.909	.759	-7.248	.578
7	-1.306	.837	-6.054	.631	-0.893	.878	-8.345	.576	-2.001	.754	-6.986	.581
8	-1.407	.826	-5.855	.634	-0.986	.864	-7.934	.584	-2.070	.751	-6.772	.584
9	-1.493	.817	-5.693	.638	-1.069	.853	-7.601	.590	-2.141	.747	-6.584	.588
10	-1.569	.809	-5.557	.640	-1.143	.844	-7.325	.595	-2.195	.744	-6.421	.590
11	-1.635	.803	-5.441	.643	-1.209	.837	-7.093	.600	-2.247	.742	-6.272	.593
12	-1.694	.798	-5.340	.645	-1.269	.830	-6.893	.604	-2.284	.740	-6.159	.595
13	-1.747	.793	-5.252	.647	-1.324	.824	-6.720	.608	-2.336	.736	-6.034	.599
14	-1.794	.789	-5.174	.649	-1.374	.819	-6.568	.611	-2.355	.737	-5.948	.600
15	-1.838	.786	-5.104	.650	-1.420	.815	-6.433	.614	-2.381	.736	-5.872	.601
20	-2.009	.773	-4.842	.657	-1.608	.798	-5.932	.626	-2.494	.731	-5.536	.610
25	-2.130	.765	-4.666	.661	-1.745	.788	-5.604	.634	-2.573	.728	-5.296	.616
30	-2.221	.759	-4.538	.665	-1.853	.780	-5.369	.640	-2.636	.725	-5.109	.622
35	-2.293	.754	-4.439	.667	-1.939	.774	-5.191	.644	-2.675	.724	-4.988	.626

Note. The correlation coefficient of each approximation is better than 0.9986, 0.9988, and 0.9998 for the special Erlangian, lognormal, and the ex-Gaussian distributions, respectively. The mean absolute deviation between approximated and true bias is .37%, .45%, and .44% for the special Erlangian, lognormal, and ex-Gaussian distributions, respectively.

The PDF of the Lognormal is unimodal and for $t > 0$ given by

$$f(t) = \frac{1}{\sqrt{2\pi\sigma t}} \exp\left[-\frac{(\ln t - \mu)^2}{2\sigma^2}\right] \quad (26)$$

and the corresponding CDF is

$$F(t) = \Phi\left[\frac{\ln t - \mu}{\sigma}\right], \quad (27)$$

where the function Φ denotes the CDF of the Standard Normal distribution. Its mean and variance are

$$E[\mathbf{T}] = \exp\left[\mu + \frac{\sigma^2}{2}\right] \quad (28)$$

and

$$\text{Var}[\mathbf{T}] = \exp(2\mu + \sigma^2) \times [\exp(\sigma^2) - 1], \quad (29)$$

respectively. Finally, the quantile t_p as a function of p is given by

$$t_p = \exp[\mu + \sigma \cdot \Phi^{-1}(p)], \quad (30)$$

where function Φ^{-1} denotes the inverse CDF of a standard normal distribution (see Shimizu & Crow, 1988, pp. 9-10).

The parameter $\sigma > 0$ determines the shape of the lognormal distribution. For small values of σ (smaller than 0.2) the lognormal is almost symmetrical and hence difficult to distinguish from the normal distribution (Kotz, 1973). Skewness increases with σ .

The raw moments of \mathbf{T}_* for a truncated lognormal distribution can be computed on the basis of the corresponding

untruncated distribution, as Proposition 2 shows: If \mathbf{T} has a lognormal distribution with shape parameter σ and scale parameter μ , then the i th raw moment of \mathbf{T}_* is given by

$$E[\mathbf{T}_*^i] = \exp\left[i\mu + \frac{i^2\sigma^2}{2}\right] \times \frac{\Phi\left[\frac{\ln U - i\sigma^2 - \mu}{\sigma}\right] - \Phi\left[\frac{\ln L - i\sigma^2 - \mu}{\sigma}\right]}{\Phi\left[\frac{\ln U - \mu}{\sigma}\right] - \Phi\left[\frac{\ln L - \mu}{\sigma}\right]}, \quad (31)$$

where L is the lower and U the upper truncation point. (Proof of Proposition 2 is given in Appendix B.)

On the basis of Proposition 2, we performed a bias analysis similar to the one already conducted with the special Erlangian distribution. For ease of comparison the values of the shape parameter σ were chosen so that the coefficients of variation for the lognormal distribution would match those of the special Erlangian conditions already studied. The coefficient of variation for the special Erlangian distribution is given by

$$\frac{SD[\mathbf{T}]}{E[\mathbf{T}]} = \frac{1}{\sqrt{m}} \quad (32)$$

and the one for the lognormal is given by

$$\frac{SD[\mathbf{T}]}{E[\mathbf{T}]} = \sqrt{\exp(\sigma^2) - 1} \quad (33)$$

By equating Equations 32 and 33, one finds $\sigma =$

Table 4

Constants A, B, C, and D for the Best Fitting Approximation $\hat{\delta}(p, q)$ of Median as a Function of m and of Theoretical Distribution

m	Erlangian distribution				Lognormal distribution				Ex-Gaussian distribution			
	A	B	C	D	A	B	C	D	A	B	C	D
2	.946	1.009	-.955	.993	.793	1.011	-.803	.992	.822	1.007	-.829	.995
3	.755	1.007	-.761	.995	.668	1.009	-.675	.993	.667	1.005	-.671	.997
4	.647	1.006	-.651	.996	.589	1.008	-.595	.994	.580	1.004	-.582	.997
5	.575	1.005	-.578	.996	.533	1.007	-.537	.995	.521	1.004	-.523	.998
6	.522	1.005	-.525	.997	.490	1.007	-.494	.995	.478	1.004	-.479	.998
7	.482	1.005	-.484	.997	.456	1.006	-.459	.995	.444	1.003	-.445	.998
8	.450	1.004	-.452	.997	.429	1.006	-.431	.996	.416	1.003	-.417	.999
9	.423	1.004	-.425	.997	.405	1.006	-.408	.996	.393	1.003	-.395	.999
10	.401	1.004	-.403	.998	.386	1.005	-.388	.996	.374	1.003	-.375	.999
11	.382	1.004	-.383	.998	.369	1.005	-.371	.996	.358	1.003	-.359	.999
12	.365	1.004	-.367	.998	.354	1.005	-.355	.997	.343	1.003	-.344	.999
13	.351	1.004	-.352	.998	.340	1.005	-.342	.997	.330	1.003	-.331	.999
14	.338	1.003	-.339	.998	.328	1.005	-.330	.997	.319	1.003	-.320	.999
15	.326	1.003	-.327	.998	.318	1.005	-.319	.997	.309	1.002	-.309	.999
20	.282	1.003	-.283	.999	.276	1.004	-.277	.997	.269	1.002	-.270	.999
25	.252	1.003	-.252	.999	.248	1.004	-.249	.998	.241	1.002	-.241	.999
30	.230	1.003	-.230	.999	.227	1.003	-.227	.998	.221	1.002	-.221	1.000
35	.212	1.002	-.213	.999	.210	1.003	-.211	.998	.205	1.002	-.205	1.000

Note. The correlation coefficient of each approximation is better than 0.9999, 0.9998, and 0.9999 for the special Erlangian, lognormal, and the ex-Gaussian distributions, respectively. The mean absolute deviation between approximated and true bias is .01%, .02%, and .01% for the special Erlangian, lognormal, and ex-Gaussian distributions, respectively.

$\sqrt{\ln(1/m + 1)}$. For example, the corresponding shape values for $m = 2$ and $m = 10$ are $\sigma = .64$ and $\sigma = .31$, respectively. For each value of σ , the scale parameter μ was chosen to hold $E[T]$ constant at 150.

Figures 2-5 show the results of the bias analyses for σ equal to .64 and .31, the values matching coefficients of variation with special Erlangians having $m = 2$ and $m = 10$. It can be seen that the magnitudes of the biases are similar to those obtained with the special Erlangians. Bias approximations for $m = 2, \dots, 35$ are provided by Tables 2, 3, and 4 for the mean, median, and standard deviation of the lognormal, respectively.

The Ex-Gaussian Distribution

The third and last distribution included in our analysis is the distribution of the sum of independent normal and exponential random variables, termed the ex-Gaussian by Burbeck and Luce (1982). This distribution is mathematically less tractable than either the lognormal or the special Erlangian, but it has nonetheless played an important role in RT modeling (cf. Luce, 1986). The distribution was first considered in detail by Hohle (1965), who reported remarkably good fits of the ex-Gaussian to empirical RT distributions of individual subjects. Since then, the ex-Gaussian has often been used to describe RT distributions in an economical way (e.g., Heathcote et al., 1991; Hockley, 1984; Ratcliff, 1978; Ratcliff & Murdock, 1976).

An ex-Gaussian random variable T may be defined as the sum of two stochastically independent random components: (a) an exponentially distributed component with mean⁷ τ and (b) a normally distributed or Gaussian component with mean

μ and standard deviation σ . Thus, the three parameters μ , σ , and τ completely specify the shape of the ex-Gaussian and allow for its rather large flexibility in shape. The ex-Gaussian distribution has the theoretical disadvantage that it allows negative RTs, because the Gaussian component can in principle be an arbitrarily large negative number. In practical use this theoretical difficulty is unimportant, however, because the probability of a negative RT is very nearly zero as long as μ is relatively large compared to σ . For example, with $\mu = 300$, $\sigma = 50$, and $\tau = 100$, this probability is less than 1 in 1,000,000,000.

The PDF of the ex-Gaussian is computed by convoluting a normal with an exponential PDF and is given by

$$f(t) = \frac{1}{\tau} \exp \left[\frac{\mu - t}{\tau} + \frac{\sigma^2}{2\tau^2} \right] \cdot \Phi \left[\frac{t - \mu}{\sigma} - \frac{\sigma}{\tau} \right]. \quad (34)$$

The parameter μ determines the location of the ex-Gaussian on the time axis without affecting its shape, so μ is a pure location or displacement parameter. Both σ and τ affect the shape of the ex-Gaussian. For relatively large values of σ , the distribution is bell-shaped, but for relatively small values of σ it is J-shaped, like an exponential distribution displaced by μ units to the right along the time axis. Roughly speaking, μ and σ affect the leading edge of the distribution (the minimum of T), and τ determines the thickness of the distribution's right tail and thus its positive skew.

⁷ Note that the relationship between the rate λ and the mean τ of an exponential random variable is $\tau = 1/\lambda$.

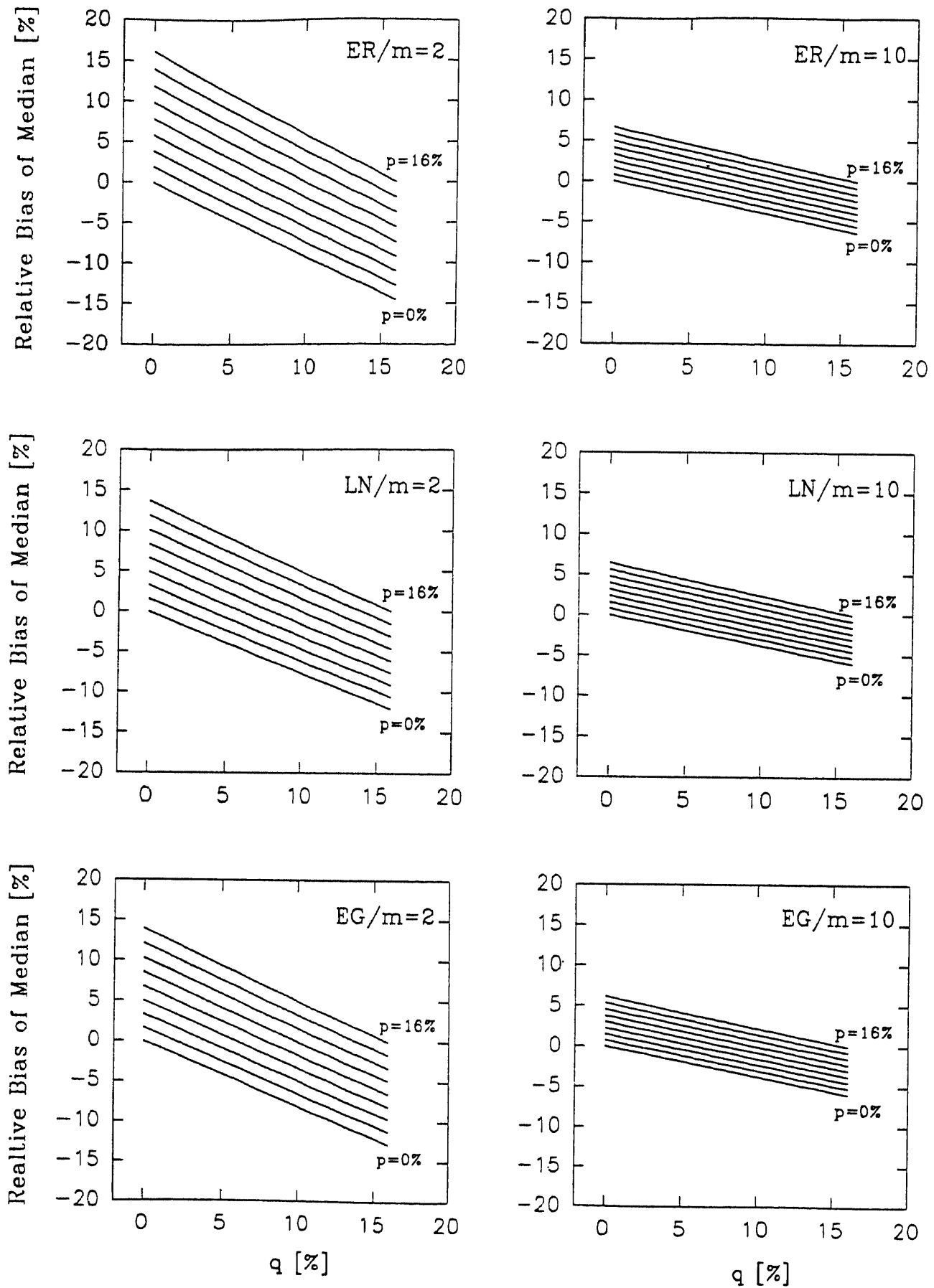


Figure 3. Relative bias of median as a function of p (percentage of observations truncated from lower tail of f) and q (percentage truncated from upper tail of f). Within each panel, each line denotes a different value of $p = 0, 2, 4, 6, 8, 10, 12, 14, 16$, but only the most extreme values (i.e., $p = 0\%$ and $p = 16\%$) are indicated because the effects of p are monotonic from $p = 0$ to $p = 16$. ER = Erlangian distribution; LN = lognormal distribution; EG = ex-Gaussian distribution.

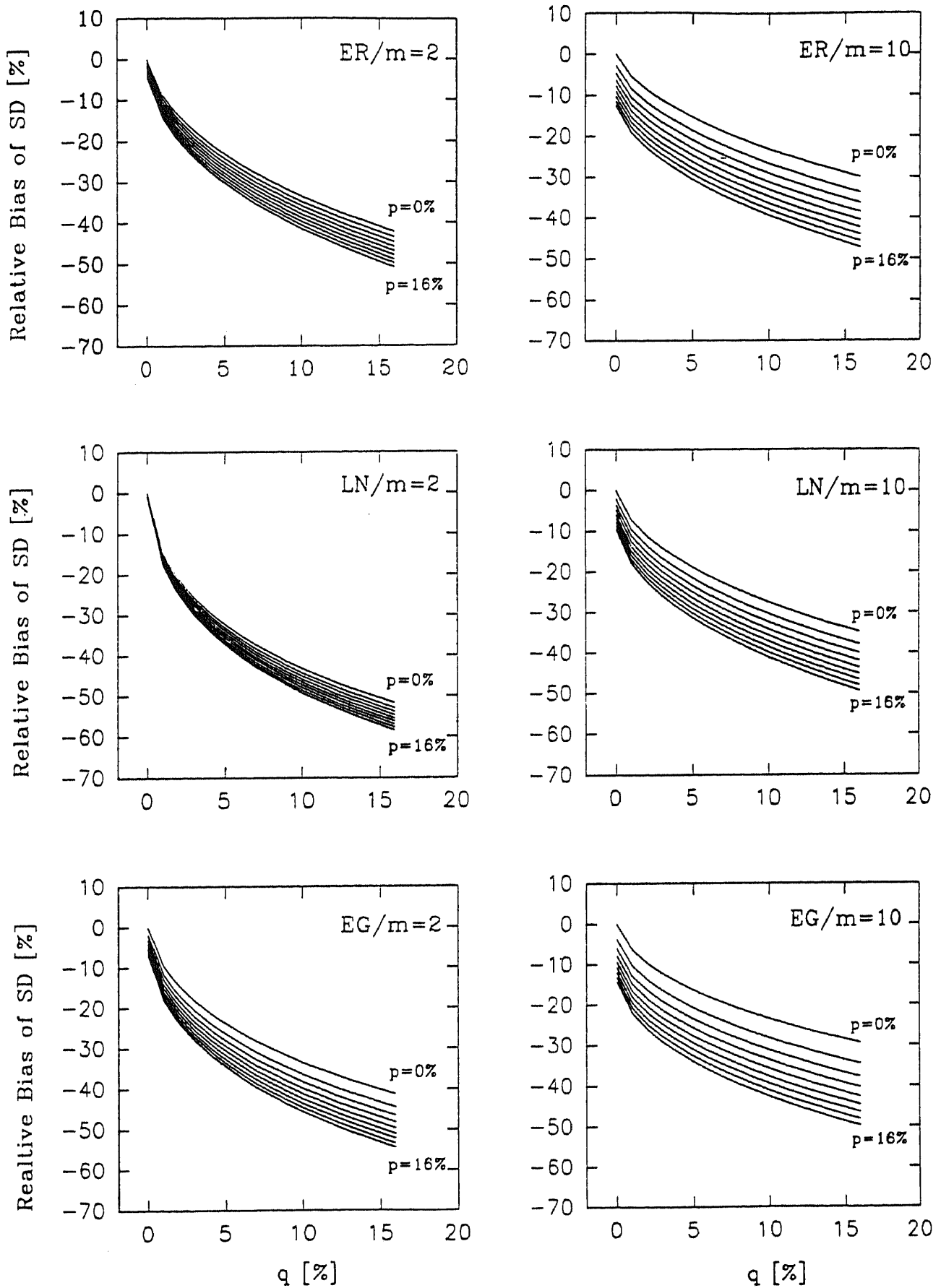


Figure 4. Relative bias of standard deviation as a function of p (percentage of observations truncated from lower tail of f) and q (percentage truncated from upper tail of f). Within each panel, each line denotes a different value of $p = 0, 2, 4, 6, 8, 10, 12, 14, 16$, but only the most extreme values (i.e., $p = 0\%$ and $p = 16\%$) are indicated because the effects of p are monotonic from $p = 0$ to $p = 16$. ER = Erlangian distribution; LN = lognormal distribution; EG = ex-Gaussian distribution.

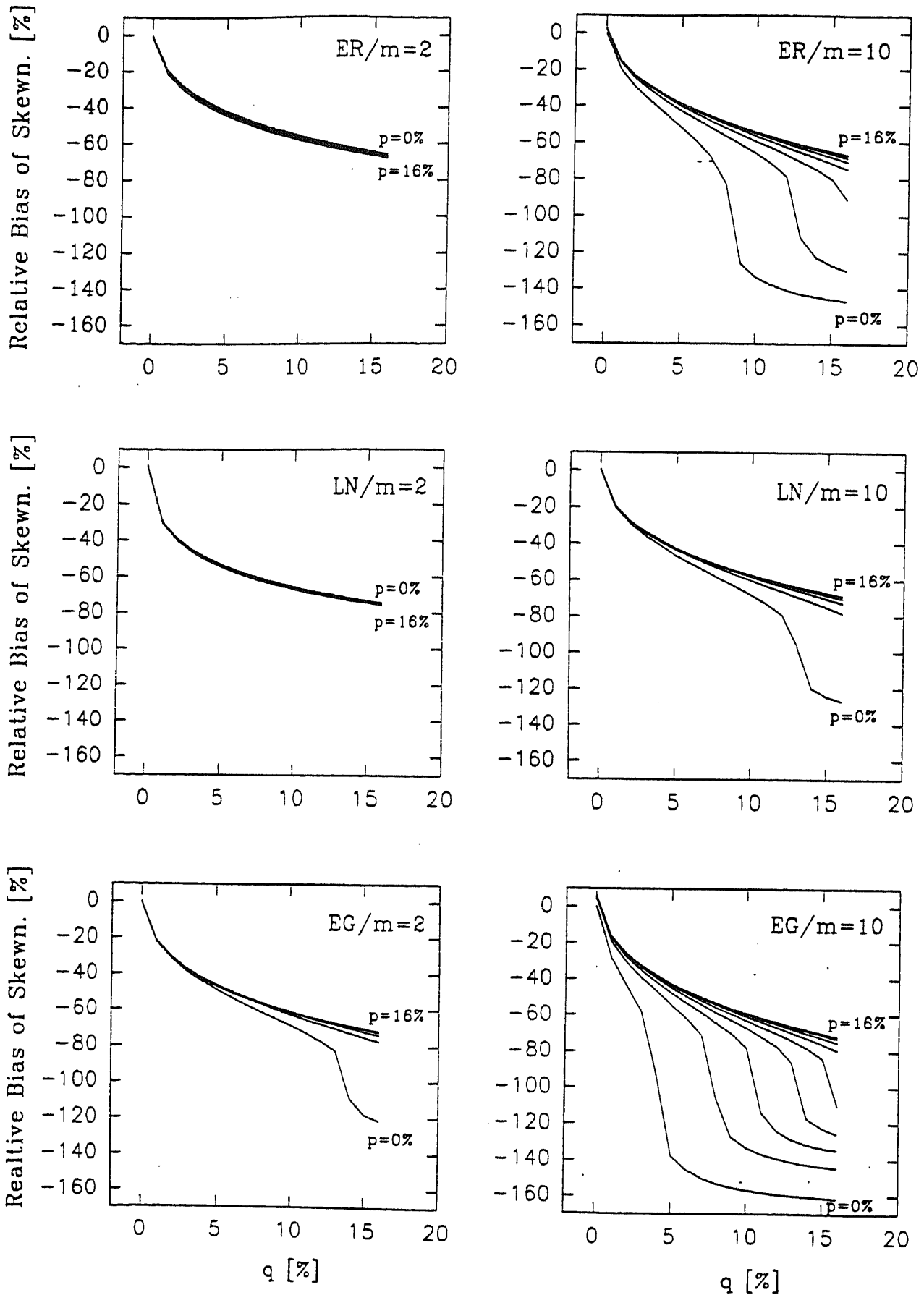


Figure 5. Relative bias of skewness as a function of p (percentage of observations truncated from lower tail of f) and q (percentage truncated from upper tail of f). Within each panel, each line denotes a different value of $p = 0, 2, 4, 6, 8, 10, 12, 14, 16$, but only the most extreme values (i.e., $p = 0\%$ and $p = 16\%$) are indicated because the effects of p are monotonic from $p = 0$ to $p = 16$. ER = Erlangian distribution; LN = lognormal distribution; EG = ex-Gaussian distribution.

Integrating Equation 34 yields the CDF of the ex-Gaussian,

$$F(t) = \Phi\left[\frac{t - \mu}{\sigma}\right] - \exp\left[\frac{\mu - t}{\tau} + \frac{\sigma^2}{2\tau^2}\right] \cdot \Phi\left[\frac{t - \mu}{\sigma} - \frac{\sigma}{\tau}\right]. \quad (35)$$

The mean and the variance of the ex-Gaussian are

$$E[T] = \mu + \tau \quad (36)$$

and

$$\text{Var}[T] = \sigma^2 + \tau^2, \quad (37)$$

respectively. The skewness of the ex-Gaussian is determined solely by the exponential component and is given by

$$\text{Skew}[T] = 2^{1/3}\tau. \quad (38)$$

The *i*th raw moment of a doubly truncated ex-Gaussian can be expressed as

$$E[T^i | L \leq T \leq U] = \frac{\frac{1}{\tau} \exp\left[\frac{\mu}{\tau} + \frac{\sigma^2}{2\tau^2}\right] \int_L^U t^i e^{-t/\tau} \Phi\left[\frac{t - \mu}{\sigma} - \frac{\sigma}{\tau}\right] dt}{F(L) - F(U)}, \quad (39)$$

where the function *F* is given by Equation 35. Romberg's numerical integration method (see Press, Flannery, Teukolsky, & Vetterling, 1986, p. 114–115) was used to evaluate Equation 39.

For ease of comparison, the ex-Gaussian was matched with the special Erlangian for various values of shape parameters *m*. It can be shown that an ex-Gaussian with parameters

$$\mu = \frac{m - m^{1/3}}{\lambda}, \quad (40)$$

$$\sigma = \frac{\sqrt{m - m^{2/3}}}{\lambda}, \quad (41)$$

and

$$\tau = \frac{m^{1/3}}{\lambda} \quad (42)$$

matches the mean, standard deviation, and skewness of a special Erlangian with parameters *m* and λ . For example, an ex-Gaussian with $\mu = 56$, $\sigma = 48$, and $\tau = 94$ matches the mean, standard deviation, and the skewness of a special Erlangian with *m* = 2 and $\lambda = 1/75$. Because the special Erlangian becomes more symmetrical with larger *m*, the ratio

$$\frac{\tau}{\sigma} = \frac{m^{1/3}}{\sqrt{m - m^{2/3}}} = \frac{1}{\sqrt{m^{1/3} - 1}} \quad (43)$$

decreases with *m*, that is, the contribution of the ex-Gaussian's exponential component diminishes with increas-

ing *m*. For each value of *m*, the scale parameter λ was chosen such that $E[T]$ was constant at 150, so all the matched ex-Gaussians in these bias computations had a mean of 150 ms.

Figures 2–5 show the results of the bias analyses for distributions matched with Erlangians having *ms* equal to 2 and 10. The results are very similar to those obtained with the Erlangian and lognormal. Tables 2–4 provide bias approximations for the ex-Gaussian in the range of *ms* = 2–35. This corresponds to values of τ/σ ranging from 1.96 to 0.66. Because some studies (Heathcote et al., 1991; Hockley, 1984; Hohle, 1965; Ratcliff & Murdock, 1976) have reported more extreme values of τ/σ —between 3 and 15—additional computations were performed to examine this range. The results of these additional computations are summarized in Table 5.

Summary of Truncation Bias

The main results depicted in Figures 2–5 clearly indicate that the effects of truncation are too large to ignore. Truncation can easily alter mean and median RT by 10% or more, so its effects are as large as those of many common experimental manipulations. The effects on standard deviation and skewness are considerably larger still, indicating that truncation would likely have extremely serious consequences on distributional analyses. The high degree of consistency of truncation effects across distributional types (i.e., Erlangian, lognormal, and ex-Gaussian) strongly suggests that truncation biases will be approximately this large regardless of the exact distribution of observable RTs, and not just a problem associated with certain theoretical distributions.

Bias Effects With Contaminated RTs

The preceding analyses investigated truncation biases under the assumption that all spurious observations (i.e., un-

Table 5
Constants of the Best Fitting Power Function
Approximations $\hat{\delta}(p, q)$ for Mean, Standard Deviation,
and Median of ex-Gaussians for Extreme τ to σ Ratios

Statistic and <i>m</i>	τ/σ	Constants			
		A	B	C	D
<i>M</i>					
1.01339	15	1.167	.934	-4.837	.733
1.02098	12	1.186	.930	-4.820	.733
1.03750	9	1.220	.923	-4.768	.733
1.37174	3	1.417	.877	-4.018	.734
<i>SD</i>					
1.01339	15	.359	-0.422	-12.338	.542
1.02098	12	-.105	.986	-11.616	.562
1.03750	9	-.075	1.237	-11.654	.560
1.37174	3	-.518	.909	-10.768	.560
<i>Mdn.</i>					
1.01339	15	1.409	1.013	-1.430	.989
1.02098	12	1.396	1.013	-1.417	.989
1.03750	9	1.369	1.013	-1.390	.989
1.37174	3	1.040	1.011	-1.053	.992

Note. The correlation coefficient of each fit is better than *r* = 0.999. The mean absolute deviations are .23, .57, and .12 for the mean, standard deviation, and median, respectively.

usually fast or slow RTs produced by aberrant processes) were eliminated by truncation. In practice, however, truncation does not eliminate all spurious observations, so T_{*o} is contaminated by a certain number of data values that ought to be classified as outliers but are not. Our purpose in this section is to investigate bias effects in this more realistic situation.

As discussed earlier, the contamination of true RTs with spurious observations is best described by the mixture distribution given in Equation 8, and the properties of such distributions were reviewed earlier (e.g., Equations 9–13). We conducted extensive computations using Equation 9 to investigate the biasing effects of truncation when the observed RTs are a mixture distribution contaminated by various percentages of spuriously fast and slow observations. As before, the relative bias δ was computed for each summary measure (mean, median, standard deviation, and skewness). For example, the relative bias, δ , of $SD[T_{*o}]$ is now defined as

$$\delta(p, q) \equiv \frac{SD[T_{*o}] - SD[T]}{SD[T]} \quad (44)$$

The definition of δ for the remaining summary measures is analogous.

In all computations, ex-Gaussian distributions were used for $f_v(t)$, $f_s(t)$, and $f_f(t)$, with parameter combinations ($\mu_v = 500$, $\sigma_v = 100$, $\tau_v = 100$), ($\mu_s = 1200$, $\sigma_s = 110$, $\tau_s = 100$), and ($\mu_f = 300$, $\sigma_f = 30$, $\tau_f = 10$), respectively. The choice of the ex-Gaussian as the model for spurious times is somewhat arbitrary, but we could find no useful data on the RT distributions generated by spurious processes. In the end, we selected the ex-Gaussian because its Gaussian and exponential components correspond intuitively to plausible sorts of spurious processes. A Gaussian component could be produced by a number of independent, additive sources of delay, and an exponential delay could be produced by a distraction away from the task. In any case, the ex-Gaussian, Erlangian, and lognormal distributions can mimic one another very well, so it probably does not matter much which of them is used as the model for outliers.

The values of 0, .025, .05, and .1 were used in a factorial combination for the mixture probabilities g_f and g_s . The parameters for spuriously fast and slow RTs were chosen to maximize the effects of spurious observations while preserving (approximately) the usual unimodal character of the distribution of RT. To satisfy these criteria, we shifted the distribution of spuriously fast RTs to the left until a small second peak was apparent in the PDF of the mixture. Analogously, we shifted $f_s(t)$ to the right until a small peak emerged on the right tail of the mixture PDF. Thus, we attempted to maximally separate $f_f(t)$ and $f_s(t)$ under the constraint that the resulting mixture distribution would have secondary modes too small to be detectable with reasonable sample sizes. Figure 6 illustrates the resulting PDF of T_o for mixture probability combinations (g_f, g_s) of (0, 0), (0, .05), (.05, 0), and (.05, .05). The resulting means of these mixtures, computed with Equation 10, are 600, 635, 586, and 621 ms, respectively, corresponding to relative biases of 0, 5.8, -2.4, and

3.4%. The corresponding relative biases of the standard deviations are 0, 47.3, 7.3, and 54.0%. These values clearly demonstrate how contamination may seriously distort the parameters of a valid RT distribution.

For each of the 16 combinations of g_f and g_s , relative bias was computed for each summary measure and for each of the 49 combinations of upper and lower truncation probabilities (p, q). The resulting 784 bias values for each summary measure were analyzed as a $4 \times 4 \times 7 \times 7$ factorial design, and the four main effects accounted for 96.4, 91.7, 99.75, and 91.0% of the variance in the bias values for the mean, standard deviation, median, and skewness, respectively. Note that these computations also include the special case of contaminated but untruncated RTs (i.e., $p = q = 0$).

Because the main effects were responsible for most of the variance, Table 6 shows only the marginal means of the bias values for each summary measure. Each marginal mean was obtained by averaging all bias values at the indicated factor level, averaging across levels of the other factors. Qualitatively, each factor has exactly the effects that would be predicted. As seen previously (e.g., Figures 2–5), mean, median, and skewness increase with the proportion truncated from the lower tail (p) and decrease with the proportion truncated from the upper tail (q). Standard deviation also decreases as more observations are truncated from either tail, but the effect is much larger with truncation from the upper tail than with truncation from the lower tail. Variation of the proportion of contaminating observations indicates that mean, median, and skewness increase with the proportion of spuriously slow observations (g_s) and decrease with the proportion of spuriously fast ones (g_f). The effects on mean and median are reassuringly small. Standard deviation increases with the proportion of spurious observations of either type, but the effect is much larger for spuriously slow observations than spuriously fast ones.

Because the main effects accounted for most of the variance in the factorial analyses, results for the full set of 784 conditions can be estimated adequately from the values in Table 6. Specifically, the estimate is the sum of the marginal mean biases for the four parameter values of interest minus three times the overall mean bias, which can be obtained by averaging any row. For example, from the table, one would estimate the bias in skewness to be $4.5 + 29.4 - 1.3 - 17.6 - 3 \times 1.52 = 10.4$ for the condition, with $p = .04$, $q = .02$, $g_f = .05$, and $g_s = .025$, where 4.5, 29.4, -1.3, and -17.6 are the marginal mean biases for $p = .04$, $q = .02$, $g_f = .05$, and $g_s = .025$, respectively, and 1.52 is the overall mean skewness obtained by averaging any of the rows for skewness. This estimate agrees reasonably well with the actual value of 6.3, at least relative to the full range (-200 to 100) of values for bias in skewness.

Besides the main effects shown in Table 6, the only other sizable source of variance was the $q \times g_s$ interaction, which accounted for 3.34, 7.89, 0.1, and 4.29% of the variance in the four summary measures. The form of this interaction was quite simple: Variation in g_s had less effect on bias as q increased. This is to be expected, because increasing q tends to exclude the spurious observations that cause the main effect of g_s . The analogous interaction was present for $p \times g_f$ but its effects on the four measures were much smaller (less

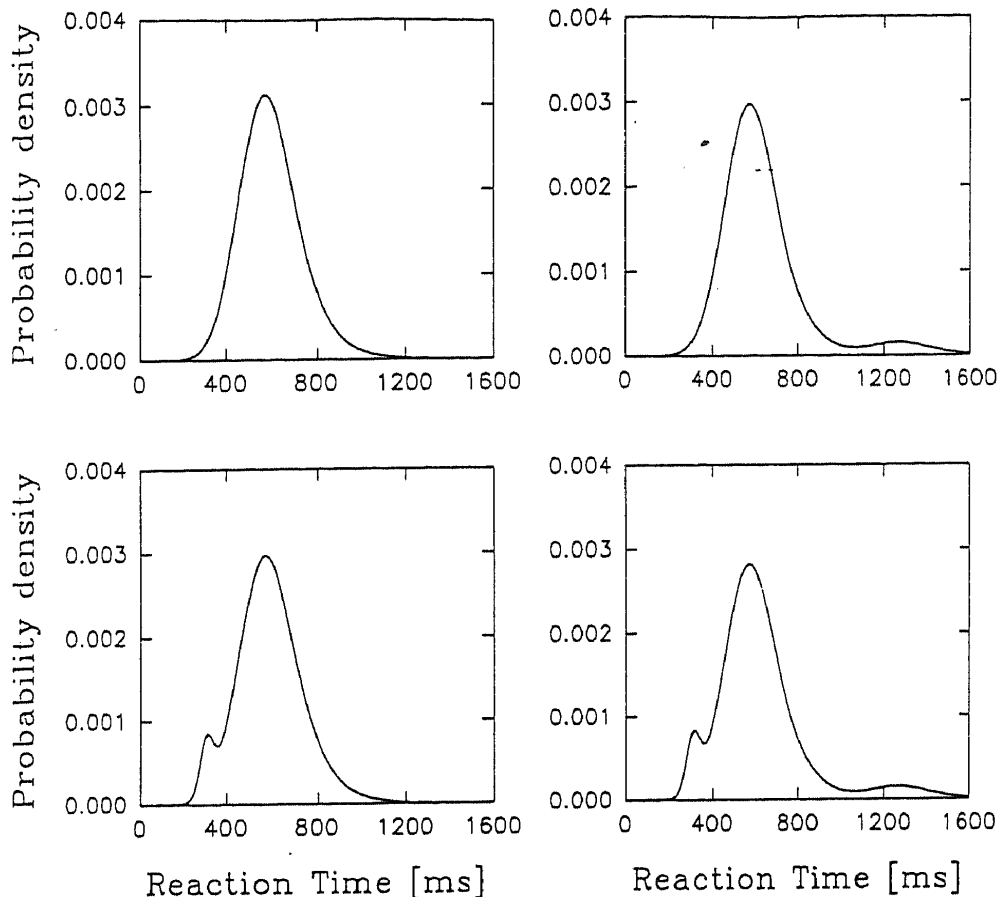


Figure 6. Mixture distributions for four combinations of mixture probabilities (g_f, g_s). Left upper panel: (0.00, 0.00). Right upper panel: (0.00, 0.05). Left lower panel: (0.05, 0.00). Right lower panel: (0.05, 0.05). (See text for more details.)

than 0.2% of the variance), because the effect of g_f was much smaller in the first place.

Bias Effects: Summary and Conclusions

The major point established by these analyses is that truncation can indeed have nonnegligible effects on summary measures estimated from RT distributions. The most commonly estimated summary measure, the mean, could easily be off by 4%–5% with truncation of as few as 2% of the true RTs. This translates into a bias effect of approximately 25 to 50 ms for tasks with mean RTs in the range of 500–1,000 ms, and this bias effect is larger than some well-studied experimental effects. In view of the noticeable effects of truncation bias on mean RT, it is important to examine the effects of such bias on particular sorts of experimental comparisons among means, and this is the topic of the next section.

In general, the median suffers rather less bias than the mean (generally around one fourth as much). Unfortunately, it will not suffice to adopt this as the universal measure of central tendency, because this statistic has an inherent bias that varies with sample size (Miller, 1988) and because many RT models make predictions about mean RT but not about median RT (e.g., Sternberg, 1969). Not surprisingly, standard deviation and skewness estimates suffer much greater truncation bias effects than the mean and median. The importance

of these large biases, of course, will increase as more RT models become sophisticated enough to make predictions about these summary measures in addition to their predictions about central tendency.

The situation is more complicated when spuriously fast or slow RTs contaminate the RT distribution used in the analysis, because these contaminating observations could counteract the effects of truncation fairly effectively, especially when only mean RT is considered. For example, in the present computations the presence of 10% spuriously slow observations increased the mean by just about as much as truncating 16% of the distribution decreased it, leaving a nearly unbiased estimate of the mean under these conditions. In practice, however, it is obviously very risky to count on such counteracting effects to yield unbiased estimates, unless one has extremely good information about the probabilities and distributions of spurious RTs. Furthermore, spurious observations cause more severe problems in the estimation of higher moments, so model tests using higher distributional moments or full distributions will be quite sensitive to such observations.

Effects of Truncation on Tests of Models

This section examines the effects of truncation on certain common tests of RT models. Such models generally make

Table 6
Relative Bias of Mean, Standard Deviation, Median,
and Skewness as a Function of $p\%$, $q\%$, g_f , and g_s

Parameter value	<i>M</i>	<i>SD</i>	<i>Mdn.</i>	<i>Skewness</i>
$p\% =$				
0	-0.9	18.5	-1.3	-12.5
.5	-0.6	17.4	-1.2	-8.5
1	-0.3	16.5	-1.0	-5.8
2	0.2	14.9	-0.7	-1.7
4	1.2	12.2	-0.1	4.5
8	3.1	7.9	1.1	13.4
16	6.5	2.0	3.6	21.2
$q\% =$				
0	5.3	39.9	1.4	80.1
.5	4.5	33.0	1.2	62.2
1	3.9	28.0	1.1	50.9
2	2.8	19.5	0.8	29.4
4	0.8	6.5	0.2	-7.9
8	-2.1	-10.5	-1.0	-70.5
16	-6.2	-27.0	-3.5	-133.7
$g_f =$				
0	3.3	7.3	1.3	11.3
.025	2.2	10.4	0.6	4.3
.05	1.0	13.6	-0.1	-1.3
.1	-1.4	19.8	-1.7	-8.2
$g_s =$				
0	-2.6	-12.9	-1.3	-64.6
.025	-.07	0.9	-0.6	-17.6
.05	1.6	16.6	0.2	18.4
.1	6.8	46.5	1.8	69.9

Note. Only the marginal means of relative bias are presented because these factors had approximately additive effects on bias. $p\%$ = proportion truncated from lower tail; $q\%$ = proportion truncated from upper tail; g_f = proportion of spuriously fast observations; g_s = proportion of spuriously slow observations.

predictions concerning the means of full, untruncated RT distributions. Thus, if the models are tested by comparing their predictions against the observed means of truncated RT distributions, it is possible that truncation will alter the results, either by introducing apparent violations of the model or by concealing actual violations. The question, then, is whether truncation effects are so small that they can safely be ignored, as is commonly done.

It is clearly impossible to provide a universal answer to this question, because the effect of truncation depends to some degree on the type of model prediction being tested. It is possible to examine certain common types of tests, however, and this may improve our intuition about the extent of truncation problems in general. Specifically, this section presents an analysis of the effect of truncation bias in two hypothesis-testing paradigms that have played particularly important roles in RT research. The first paradigm examines the question of whether RT is a linear function of an independent variable, and the second concerns additivity of factor effects in factorial experiments. Within each paradigm, we consider the three scenarios discussed in the introduction: Truncation, No Contamination, No Truncation, Contamination, and Truncation Plus Contamination.

Possible Distributions of Spurious RTs

Because these hypothesis-testing paradigms involve more than one experimental condition, it is necessary to make some assumptions about how the spurious RTs vary across conditions. We assumed, first, that the probabilities and distributions of spuriously fast RTs were the same in all conditions. If subjects were making more very fast responses (e.g., fast guesses) in one condition than another, then these responses should probably be regarded as part of a strategic adaptation to the condition, and not as spurious observations.

We also assumed that the probabilities of spuriously slow RTs were independent of condition. In principle, this assumption is not very attractive, because conditions with longer RTs provide more opportunity for a distracting event (e.g., a sneeze) to occur. In typical RT experiments, though, conditions differ by at most 10%–15% of the mean RT, so the extra opportunity for distracting events should cause only small violations of this assumption.

With respect to the latencies of the spuriously slow RTs, we considered two possibilities. One, called *condition independent*, is that these RTs have the same distribution in all conditions, as they might if the subject used a special processing strategy, one insensitive to conditions, on all such trials. For this model the observed RTs in condition j are given by the mixture model

$$\mathbf{T}_{oj} = \text{mix}(\mathbf{T}_f, \mathbf{T}_j, \mathbf{T}_s), \quad (45)$$

where \mathbf{T}_j denotes valid RTs under condition j .

The other possibility, called *delayed startup*, is that the spuriously slow RTs show the same effects of experimental conditions as the valid trials but are merely slower (and possibly more variable) overall owing to the addition of a random delay \mathbf{D} on spurious trials. For example, processing might start \mathbf{D} ms after stimulus onset because of inattention and then proceed normally once this delay is over. Assuming that the distribution of \mathbf{D} does not depend on experimental condition, this set of assumptions yields the alternative mixture model

$$\mathbf{T}_{oj} = \text{mix}(\mathbf{T}_f, \mathbf{T}_j, \mathbf{D} + \mathbf{T}_j), \quad (46)$$

in which \mathbf{T}_s is replaced by the sum $\mathbf{D} + \mathbf{T}_j$.

Effects of Truncation on Linear Increases in RT

Many researchers have tested models predicting a linear increase in the mean (and sometimes variance) of RT as a function of an independent variable such as positive set size in memory scanning tasks, display size in visual search tasks, logarithm of the number of alternatives in choice RT tasks, or angle of stimulus orientation in mental rotation tasks. In the basic linear model (e.g., Sternberg, 1969), an RT in condition j , \mathbf{T}_j , can be represented by the sum

$$\mathbf{T}_j = \sum_{k=1}^j \mathbf{Z}_k + \mathbf{R}, \quad (47)$$

where $\mathbf{Z}_1, \dots, \mathbf{Z}_j$ represent random variables with the common mean m and \mathbf{R} is a residual component with mean c .

In many experiments of this type, observed RT increases nearly linearly as a function of the independent variable. There are, however, sometimes hints that the function is slightly negatively accelerated. For example, Briggs (1974) reanalyzed a number of published studies using the memory scanning task (e.g., Sternberg, 1967). Across a total of 145 studies, he found that a negatively accelerated RT function provided a better fit than a linear RT function in 90 out of 155 cases. Indeed, the curvilinear relationship between mean RT and set size appears to be quite a common observation, judging from Baddeley's (1990) comment that "his [Sternberg's] results have proved to be replicable, although not everyone manages to get such elegant straight lines" (pp. 277-278). It seems possible that minor differences in truncation, contamination, or both might influence deviations from linearity, and thereby account for some of the discrepancies among results. In the next three parts of this section, then, we consider the three scenarios of truncation, contamination, or both in this paradigm, to see whether any of these might contribute substantially to deviations from linearity.

Truncation, No Contamination. Intuitively, it seems clear that truncation of the RT distributions might account for some of the negative acceleration seen in functions relating RT to an independent variable. With exclusion of RTs longer than some fixed value (e.g., 2 s), it is clear that the percentage of slow but valid RTs truncated from a condition would increase with the true mean of the condition. As shown earlier (e.g., Figure 2), this means that, as true mean RT increases, the mean of the truncated distribution increasingly underestimates the mean of the true distribution. If this underestimation is a linear effect, it would simply decrease the slope of the function relating RT to the independent variable. If underestimation increases more than linearly with the mean of the true distribution, however, then it could be responsible for the negative acceleration often observed in truncated mean RT functions, even if the true mean RT function is linear.

To see whether negatively accelerated RT functions might be an artifact of outlier exclusion, we carried out a series of computations with the *ex*-Gaussian distribution, using parameter estimates of μ , σ , and τ reported by Hockley (1984). These estimates were obtained from measurements of RT as a function of positive set size, j , in a memory scanning task, and they appear to be particularly useful for the present purposes because they were based on large numbers of observations and because very few trials were omitted from the estimation procedure.

As the *ex*-Gaussian distribution denotes the sum of a Gaussian component and an exponential component, any increase in mean RT with j may reflect an influence of j on μ , τ , or both. Specifically,

$$E[\mathbf{T}_j] = \mu_j + \tau_j, \quad (48)$$

where the index j denotes positive set size. In memory scanning experiments, for example, the increase in mean RT seems mainly to reflect an increase of τ (Hockley, 1984; Hockley & Corballis, 1982). In visual search tasks, on the other hand, the increase can be due to either μ or τ , depending on the arrangement of the stimuli in the display. With linear stimulus arrays the increase is usually due to μ (Hockley,

1984), whereas with rectangular or circular displays large shifts in τ are observed (Heathcote & Mewhort, 1993).

For positive responses in a memory search task, Hockley (1984) found that the increase in mean RT with j was well described by these equations:

$$\mu_j = 4.2 \cdot j + 430.8 \quad (49)$$

$$\tau_j = 43.2 \cdot j + 29.6 \quad (50)$$

$$\sigma_j \approx 70. \quad (51)$$

Without truncation, Equations 48, 49, and 50 combine to yield the overall RT function

$$E[\mathbf{T}_j] = 47.4 \cdot j + 460.4. \quad (52)$$

To investigate the effect of truncation on these linear RT functions, a lower cutoff L_1 was computed so that it truncated a specified proportion p_1 from the lower tail of PDF f_1 in the condition with $j = 1$. Analogously, an upper cutoff U_6 truncated a proportion of q_6 from the upper tail of f_6 in the condition with $j = 6$. The values (0, .005, .01, .02, .04, .08) were used for p_1 and q_6 , and were varied orthogonally. For each combination (p_1, q_6), the truncated mean $E[\mathbf{T}_{*j}] \equiv E[\mathbf{T}_j | L_1 \leq \mathbf{T}_j \leq U_6]$ was computed for all six conditions ($j = 1, \dots, 6$).

The distorting effects of truncation on linearity were summarized using two measures. First, a second-order polynomial

$$P(j) = c_0 + c_1 \cdot j + c_2 \cdot j^2 \quad (53)$$

was fitted by the method of least squares to the set of six truncated means obtained under each truncation combination (p_1, q_6). This second-order polynomial provided a virtually perfect fit to the obtained RT functions, accounting in each case for more than 99.99% of the variance. The coefficient c_2 in Equation 53 provides a convenient index of the nonlinear distortion introduced by truncation. To the extent that c_2 is negative (positive), then the RT function through the truncated means is negatively (positively) accelerated.

A second index of nonlinearity was the difference between the increments in mean RT produced by the second and sixth items in the positive set. Let $\Delta_j \equiv E[\mathbf{T}_{*j}] - E[\mathbf{T}_{*j-1}]$, for $j = 2, \dots, 6$, be the difference between each pair of adjacent truncated means. In the case of linear RT functions, these differences are all equal. With a negatively (positively) accelerated function, however, these differences will decrease (increase) as j increases. Thus, another easily interpretable index of deviation from linearity is

$$\Delta' \equiv \Delta_6 - \Delta_2. \quad (54)$$

Like coefficient c_2 , the index Δ' is negative (positive) if the RT function is negatively (positively) accelerated, and its absolute value, $|\Delta'|$, provides information about the size of the nonlinearity.

Table 7 shows the nonlinearity indices c_2 and Δ' for the functions relating mean RT to j for each combination of truncation proportions (p_1, q_6). Truncation at the upper end

clearly produces a noticeable negative acceleration in the RT function. For example, consider the difference Δ of two adjacent truncated means. For $p_1 = 0$ and $q_6 = .02$, the obtained values of Δ_j were 47, 47, 44, 40, and 36 ms for $j = 2, \dots, 6$, respectively. Clearly, these values display a meaningful negative acceleration; this negative acceleration is also apparent in the summary indices Δ' and c_2 .

Surprisingly, linearity of mean RTs is nearly uninfluenced if the distributions are truncated from below. As an example, for the combination $p_1 = .02$ and ($q_6 = 0$) the corresponding Δ_j s are 46, 47, 47, 47, and 47 ms, respectively, indicating an extremely slight positive acceleration. Clearly, however, the influence of q_6 dominates, so a negatively accelerated distortion is the most likely case.

The distortion of linearity produced by truncation from above has the additional consequences that the true slope m will be underestimated and the true intercept c will be overestimated. For example, if 8% of RTs are truncated from above, m decreases from 47 to 35 ms and c increases from 460 to 483 ms.

These computations using parameter estimates from Hockley (1984) reveal the size of truncation effects when set size has approximately 90% of its effect on the exponential component and 10% on the normal component (cf. Equations 49 and 50), parameters for which the skewness as well as the mean of RT increases with set size. Intuitively, it seems that truncation would produce less distortion of linearity if skewness increased less—or even decreased—with set size. To check on the influence of skewness we conducted additional sets of computations comparable to those shown in Table 7, varying the dependence of μ and τ on j . When the set size effect was split equally between μ and τ , for example, the distortion was about two thirds the size of that shown in Table 7. Distorting effects were small—only about 30%–40% as large as those in the table—when 90% of the set size effect was in μ and 10% was in τ , but in this case the skewness of the RT distributions was unrealistically small, with respect not only to the parameter estimates given by Hockley (1984) but also to those reported by other investigators (e.g., Heathcote et al., 1991; Ratcliff, 1979).

Overall, then, negatively accelerated mean RT functions must be interpreted cautiously when truncation is used, because of the possibility that the true function is linear and the observed negative acceleration is an artifact of truncation. It is difficult to estimate the effects of truncation on most previously reported studies, because few researchers specify their truncation cutoffs or report the percentage of observations excluded. Clearly, then, one implication of these results is that future researchers should report these values. In certain cases, of course, the true function may actually be negatively accelerated, as is sometimes predicted on theoretical grounds (e.g., Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; Townsend & Ashby, 1983). In the memory search task, for example, Jones and Anderson (1982) found that RT increased linearly with set size early in practice and that the function became negatively accelerated later in practice. The truncation artifact should decrease with practice, so it appears likely that the

Table 7

Effects of Truncation on Linear Relation of RT to Set Size j Across Six Conditions Having ex -Gaussian Distributions

p_1 (%)	Effect measure	q_6 (%)					
		0.0	0.5	1.0	2.0	4.0	8.0
0.0	Δ'	0	-4	-7	-11	-17	-23
	c_2	0.0	-0.5	-0.9	-1.4	-2.1	-3.0
0.5	Δ'	0	-5	-7	-11	-17	-23
	c_2	0.1	-0.6	-0.9	-1.4	-2.2	-3.0
1.0	Δ'	-0	-4	-7	-12	-18	-22
	c_2	-0.0	-0.5	-0.9	-1.5	-2.3	-2.9
2.0	Δ'	1	-4	-6	-11	-16	-22
	c_2	0.1	-0.5	-0.8	-1.4	-2.1	-2.6
4.0	Δ'	1	-3	-6	-10	-16	-22
	c_2	0.1	-0.3	-0.8	-1.4	-2.1	-2.9
8.0	Δ'	1	-3	-7	-10	-15	-21
	c_2	0.2	-0.4	-0.8	-1.2	-2.0	-2.9

Note. For each combination of $p_1\%$ (proportion truncated from lower tail of f_1) and $q_6\%$ (proportion truncated from upper tail of f_6), the tabled values are $\Delta' = \Delta_6 - \Delta_2$ and c_2 ; —both are measures of deviation from linearity (compare with text). The value of Δ' is rounded to the nearest millisecond, and c_2 is rounded to the nearest ms/10.

true function really was negatively accelerated after practice in this experiment.⁸

Calculations parallel to those reported earlier were also carried out for variances of RT, to examine truncation effects on that summary measure, and the overall results were quite straightforward. Note that, even without truncation, Equations 49–51 yield positively accelerated rather than linear functions relating the variance of RT to set size (i.e., $c_2 \approx .12 \times c_1$ in a second-order polynomial analogous to Equation 53). It was nonetheless possible to assess the effects of truncation by seeing how it altered the shape of the function obtained in the untruncated case. Truncation from below decreased total variance by as much as 15%–20%, but had little effect on the shape of the function. Truncation from above, on the other hand, changed the shape quite dramatically. The positive acceleration was virtually eliminated with $q = 2\%$, and a nontrivial negative acceleration was present with $q = 8\%$ (i.e., $c_2 \approx -.10 \times c_1$). Thus, the function relating variance of RT to a quantitative independent variable can be affected quite drastically by truncation.

No Truncation, Contamination. Next, we consider the situation in which spurious RTs are present but truncation is not used. Surprisingly, our analyses indicate that model predictions are unaffected by the presence of spurious RTs, whether spuriously slow RTs are generated in accordance with the condition-independent model or the delayed-startup model.

For the condition-independent mixture model, the combination of Equations 45 and 47 implies the following overall

⁸ We are indebted to Andrew Heathcote for pointing out this example.

RT function:

$$E[\text{mix}(\mathbf{T}_f, \mathbf{T}_j, \mathbf{T}_s)] = g_f \cdot E[\mathbf{T}_f] + g_s \cdot E[\mathbf{T}_s] \\ + (1 - g_f - g_s) \cdot (m \cdot j + c) = c_o + m_o \cdot j. \quad (55)$$

Thus, the mean RT function remains linear in the presence of contamination, but its intercept and slope shift to $c_o = g_f \cdot E[\mathbf{T}_f] + g_s \cdot E[\mathbf{T}_s] + (1 - g_f - g_s) \cdot c$ and $m_o = (1 - g_f - g_s) \cdot m$, respectively. Similarly, the alternative delayed-startup mixture model given by Equation 46 implies the RT function

$$E[\text{mix}(\mathbf{T}_f, \mathbf{T}_j, \mathbf{D} + \mathbf{T}_j)] = g_f \cdot E[\mathbf{T}_f] + g_s \cdot E[\mathbf{D}] \\ + (1 - g_f) \cdot (m \cdot j + c) = c_o + m_o \cdot j, \quad (56)$$

with intercept and slope shifting to $c_o = g_f \cdot E[\mathbf{T}_f] + g_s \cdot E[\mathbf{D}] + (1 - g_f) \cdot c$ and $m_o = (1 - g_f) \cdot m$, respectively. In the absence of truncation, then, contamination does not distort underlying linearity of mean RT under either mixture model, although the slope of the linear function can be greatly reduced if the proportion of contaminating observations is high.

The variance of RT can also be used to test serial models (e.g., Schneider & Shiffrin, 1977; Sternberg, 1969; Townsend & Ashby, 1983), so it is of further interest to examine the effects of contamination on this summary measure. Again, we proceed from Equation 47 and invoke the common assumptions that $\mathbf{Z}_1, \dots, \mathbf{Z}_j, \mathbf{R}$ are pairwise uncorrelated and that $\text{Var}[\mathbf{Z}_k] = \text{Var}[\mathbf{Z}]$ for $k = 1, \dots, j$. Given these assumptions, the true variance function, $\text{Var}[\mathbf{T}_j] = j \cdot \text{Var}[\mathbf{Z}] + \text{Var}[\mathbf{R}]$, increases linearly with j .

Unfortunately, the linear increase in variance can be totally destroyed by contamination. Using Equation 11, one obtains for the condition-independent mixture model

$$\text{Var}[\text{mix}(\mathbf{T}_f, \mathbf{T}_j, \mathbf{T}_s)] = g_f \cdot (E[\mathbf{T}_f]^2 + \text{Var}[\mathbf{T}_f]) + g_s \cdot (E[\mathbf{T}_s]^2 \\ + \text{Var}[\mathbf{T}_s]) - \{E[\mathbf{T}_f] \cdot g_f + (E[\mathbf{R}] + j \cdot E[\mathbf{Z}])(1 - g_f - g_s) \\ + E[\mathbf{T}_s] \cdot g_s\}^2 + (1 - g_f - g_s) \cdot \{E[\mathbf{R}] + j \cdot E[\mathbf{Z}]\}^2 + \text{Var}[\mathbf{R}] \\ + j \cdot \text{Var}[\mathbf{Z}]. \quad (57)$$

Thus, under these conditions $\text{Var}[\mathbf{T}_{oj}]$ is a quadratic function of j rather than a linear function, and sample computations reveal that the form of Equation 57 strongly depends on g_f and g_s . To give a numerical example, assume that $E[\mathbf{T}_f] = 310$, $\text{Var}[\mathbf{T}_f] = 36.6^2$, $E[\mathbf{T}_s] = 1,300$, $\text{Var}[\mathbf{T}_s] = 148.7^2$, $E[\mathbf{Z}] = 47.4$, $\text{Var}[\mathbf{Z}] = 43.2^2$, $E[\mathbf{R}] = 460.4$, and $\text{Var}[\mathbf{R}] = 100^2$. (The last four values are estimates derived from the study of Hockley, 1984; the remaining ones correspond to the parameter values used above for spurious RTs.) For $g_f = .05$ and $g_s = .025$, one obtains

$$\text{Var}[\mathbf{T}_{oj}] = 28,442.5 + 545.1 \cdot j + 155.9 \cdot j^2 \quad (58)$$

and for $g_f = .025$ and $g_s = .05$

$$\text{Var}[\mathbf{T}_{oj}] = 44,740.2 - 1,625.3 \cdot j + 155.9 \cdot j^2. \quad (59)$$

Both numerical examples clearly demonstrate that g_f and g_s perturb the true underlying linearity ($10,000 + 1,866.4 \cdot j$). Nevertheless, Equation 58 represents an almost perfect linear increase in $\text{Var}[\mathbf{T}_{oj}]$ for $j = 1, \dots, 6$ ($r^2 \approx .98$). However, Equation 59 represents a strong decrease in variance for $j =$

$1, \dots, 6$. Rearranging Equation 57 yields the weight of the j^2 term, which is given by $E[\mathbf{Z}]^2 \cdot (g_f + g_s)(1 - g_f - g_s)$. Hence, the linear distortion of the true underlying variance function depends only on the percentage of spurious RTs but not, surprisingly, on the distributions of spurious RTs.

Similar variance perturbations are obtained for the alternative delayed-startup mixture model Equation 46. For this model one obtains

$$\text{Var}[\text{mix}(\mathbf{T}_f, \mathbf{T}_j, \mathbf{T}_j + \mathbf{D})] = g_f \cdot (E[\mathbf{T}_f]^2 + \text{Var}[\mathbf{T}_f]) \\ + g_s \cdot \{E[\mathbf{D}] + E[\mathbf{R}] + j \cdot E[\mathbf{Z}]\}^2 + \text{Var}[\mathbf{D}] + \text{Var}[\mathbf{R}] \\ + j \cdot \text{Var}[\mathbf{Z}] - \{E[\mathbf{T}_f] \cdot g_f + (E[\mathbf{R}] + j \cdot E[\mathbf{Z}])(1 - g_f) \\ + E[\mathbf{D}] \cdot g_s\}^2 + (1 - g_f - g_s) \cdot \{E[\mathbf{R}] + j \cdot E[\mathbf{Z}]\}^2 + \text{Var}[\mathbf{R}] \\ + j \cdot \text{Var}[\mathbf{Z}]. \quad (60)$$

Again, $\text{Var}[\mathbf{T}_{oj}]$ is a quadratic function of j rather than a linear function and again Equation 60 depends strongly on g_f and g_s . To give a numerical example, assume the same parameter values that were used for the condition-independent mixture model. For purposes of comparison we assume $E[\mathbf{D}] = 1,300$ and $\text{Var}[\mathbf{D}] = 148.7^2$; note that these values were also chosen for $E[\mathbf{T}_s]$ and $\text{Var}[\mathbf{T}_s]$, respectively, in the previous computation. Under these assumptions, one computes for $g_f = .05$ and $g_s = .025$

$$\text{Var}[\mathbf{T}_{oj}] = 52,876.8 + 2,604.2 \cdot j + 106.7 \cdot j^2 \quad (61)$$

and for $g_f = .025$ and $g_s = .05$

$$\text{Var}[\mathbf{T}_{oj}] = 92,204.2 + 2,321.1 \cdot j + 54.8 \cdot j^2. \quad (62)$$

Rearranging Equation 61 yields the general weight, $E[\mathbf{Z}] \cdot g_f(1 - g_f)$, for the j^2 term. Interestingly, this weight depends neither on \mathbf{D} nor on g_s .

In summary, on the one hand contamination does not perturb underlying linearity of mean RT functions, although it does cause underestimation of the slopes of these functions. On the other hand, contamination does perturb—possibly quite seriously—linear increases in the variance of RT. In general, the distortion of a truly linear variance function depends only on the percentage but not on the distributions of spurious RTs.

Truncation Plus Contamination. The combined effects of truncation and contamination on linear RT functions were investigated by including spuriously fast and slow observations into the same distributions of valid RTs examined in Table 7. Truncation cutoffs were adjusted to give p_1 and q_6 values of .005, .01, .02, .04, and .08, varied orthogonally, as in Table 7. For each combination of p_1 and q_6 , the distortion of linearity was examined for 144 different mixture models defined by a factorial design based on four values of g_f (0, .025, .05, .1), the same four values of g_s , three values of a_f (.1, .2, .3), and the same three values of a_s . The distribution of spuriously fast RTs was assumed to be ex-Gaussian with parameters $\mu = u_f$, $\sigma = 30$, and $\tau = 10$, where u_f was adjusted to give the desired probability of acceptance a_f (given fixed values of σ and τ , u_f is uniquely determined by the selected values of a_f and the lower truncation cutoff dictated by p_1). The distribution of spuriously slow RTs was assumed to be ex-Gaussian with parameters $\mu = u_s$, $\sigma = 110$, and $\tau = 100$,

where u_s was similarly adjusted to give the desired acceptance probability a_s for the predetermined upper truncation cutoff. Note that the proportion of accepted RTs that are spurious can be computed from these probabilities. For example, if $a_s = .1$ and $g_s = .05$, then $.1 \times .05 = .005$ is the proportion of accepted RTs from the spuriously slow distribution.

Table 8 shows obtained distortions of linearity, measured with Δ' , for a subset of the conditions examined. Results are shown for each value of q_6 , because, as with uncontaminated truncated distributions (Table 7), the extent of truncation from above had a large effect on the amount of negative acceleration. Table 8 shows results only for extreme values of the other parameters (g_f , g_s , a_f , a_s , and p_1), however, because these had small effects on negative acceleration. If the results are compared with those of Table 7, it seems clear that truncation still produces substantial negative acceleration of linear RT functions even when reasonable amounts of contamination are present. Thus, contamination provides no protection against the biasing effects of truncation.

Effects of Truncation on Additivity of Factor Effects

Another very common type of hypothesis testing in RT research is to check for the additivity or interaction of factor effects on the mean (and possibly variance) of RT in factorial experiments. This paradigm is common because many models of specific tasks and processes predict additivity or interaction of a certain pair of factors (e.g., Bundesen, Larsen, & Farrell, 1981; Cowan & Barron, 1987; Dixon & Just, 1986; Egeth & Dagenbach, 1991; Lambert & Hockey, 1986; Miller, 1983; Murphy & Eriksen, 1987; Pashler, 1984a, 1984b).

In addition, factorial experiments derive special importance in RT research because patterns of factor additivity and interaction have implications about the underlying structure of information-processing architectures. Originally, Sternberg (1969) showed that serial-stage models often predict additive effects in factorial experiments, and he developed the Additive Factor Method (AFM) based on this type of prediction. When the assumptions of this method are met, interactive (or additive) factor effects on RT can be interpreted as evidence that the factors do (or do not) affect a common stage of mental processing. This potentially powerful tool has not only been used widely (cf. Massaro & Cowan, 1993) but also continues to be developed (e.g., Roberts & Sternberg, 1992). More recently, researchers have considered the predictions of nonserial RT models with respect to factor additivity and interaction (see Townsend, 1992, for a review). For example, many parallel processing models predict underadditive interactions of factors affecting concurrent stages (cf. Miller, 1993; Sternberg, 1969, p. 288; Townsend, 1984; Townsend & Ashby, 1983, chapter 12). RT models based on program evaluation and review technique (PERT) networks, on the other hand, can predict factor additivity, underadditivity, or overadditivity, depending on the precise configuration of the network and the stages affected by the experimental factors (Schweickert, 1978; Schweickert & Townsend, 1989; Townsend & Schweickert, 1985, 1989). Finally, McClelland (1979) offered an explanation of factor additivity and interactions based on a continuous information-processing model known as the cascade model.

Thus, patterns of additivity and interaction provide an important source of evidence for testing between alternative mental architectures.

Because patterns of additivity and interaction are so important, it is necessary to consider the possible biasing effects of truncation, contamination, or both in factorial designs. The next three parts of this section do so for our standard scenarios, restricting consideration to the case of 2×2 designs to simplify the analysis. We consider only experiments with truly additive mean RTs; the complementary effects (i.e., biases when true interactions are present) should be of approximately the same magnitude.

Some notation is necessary. Let the random variable T_{ab} denote RTs under the (a, b) combination of factor levels, with $a = 1, 2$ and $b = 1, 2$, indicating the levels of Factors A and B, respectively. Thus, $E[T_{ab}]$ is the true mean of valid RTs in the condition with Factor A at level a and Factor B at level b . Without loss of generality, we label conditions so that $E[T_{11}] < E[T_{12}]$ and $E[T_{11}] < E[T_{21}]$. As noted above, for these analyses, we assume that T_{ab} conforms to an additive RT model, so that

$$T_{ab} = Z_a + Z_b + R, \quad (63)$$

where the components Z_a and Z_b are selectively influenced by Factors A and B, respectively.

Truncation, No Contamination. Suppose that, for each condition of the factorial design, the experimenter excludes observed RTs outside of the interval $[L, U]$. Now, even if two factors actually produce additive effects on mean RT, additivity may not be present for the truncated mean RTs estimated by the experimenter, $E[T_{*ab}] \equiv E[T_{ab} | L < T_{ab} < U]$, $a, b = 1, 2$.

To examine the effect of truncation on tests of additivity, we define the *distortion of additivity index* (D) as

$$D \equiv \frac{D_2 - D_1}{D_1 + D_2} \cdot 100\%, \quad (64)$$

where $D_1 = E[T_{*12}] - E[T_{*11}]$ and $D_2 = E[T_{*22}] - E[T_{*21}]$. Intuitively, D_1 is the influence of Factor B at Level 1 of Factor A and D_2 is the influence of Factor B at Level 2 of Factor A. Thus, $D_2 - D_1$ measures the change in B's effect across the two levels of A. D is simply this difference normalized by the sum $D_1 + D_2$ to create the reference values $D = 100\%$ and $D = -100\%$ in case of $D_1 = 0$ and $D_2 = 0$, respectively. Thus, $D < 0$ and $D > 0$ correspond to underadditive and overadditive distortions, respectively. For example, if the truncated means for T_{*11} , T_{*12} , T_{*21} , and T_{*22} are 300, 340, 330, and 350 ms, respectively, then there is an underadditive distortion of $D = -33.3\%$. As desired, D is independent of both the scale and the absolute size of the means.

We examined distortions of additivity produced by truncation of ex-Gaussian RT distributions in 2×2 factorial designs with a variety of different parameter values. To do this, we first had to choose values of the parameters μ , σ , and τ for each cell of the design so that the resulting true means exhibited additivity. Hence, a total of 12 parameters had to be specified for each factorial design.

To ensure that we used a range of fairly realistic parameter estimates, we started with a set of RTs observed from 16

Table 8
Effects of Truncation and Contamination on Linear Relation of Reaction Time to Set Size j Across Six Conditions Having the Same ex-Gaussian Distributions as in Table 7

g_f	g_s	a_f	a_s	p_1 (%)	Value of q_6 (%)				
					.5	1	2	4	8
0	.1	—	.1	.5	-4.3	-7.7	-11.2	-16.5	-22.5
0	.1	—	.3	.5	-4.1	-6.8	-10.7	-15.9	-21.7
.1	0	.1	—	.5	-4.4	-7.2	-11.3	-16.7	-22.7
.1	0	.3	—	.5	-4.3	-7.1	-11.2	-16.5	-22.5
.1	.1	.1	.1	.5	-4.3	-7.6	-11.0	-16.3	-22.3
.1	.1	.1	.3	.5	-4.0	-6.7	-10.6	-15.7	-21.5
.1	.1	.3	.1	.5	-4.2	-7.5	-10.8	-16.1	-22.0
.1	.1	.3	.3	.5	-4.0	-6.6	-10.4	-15.5	-21.2
0	.1	—	.1	8.	-2.4	-4.7	-9.4	-14.8	-20.8
0	.1	—	.3	8.	-1.4	-4.3	-8.4	-11.7	-19.6
.1	0	.1	—	8.	-2.8	-5.6	-9.9	-15.3	-21.2
.1	0	.3	—	8.	-2.8	-5.5	-9.6	-15.0	-20.8
.1	.1	.1	.1	8.	-2.3	-4.5	-9.2	-14.6	-20.5
.1	.1	.1	.3	8.	-1.2	-4.1	-8.1	-11.3	-19.3
.1	.1	.3	.1	8.	-2.2	-4.4	-9.1	-14.4	-20.2
.1	.1	.3	.3	8.	-1.2	-4.0	-8.0	-11.1	-19.0

Note. Tabled values are $\Delta' = \Delta_6 - \Delta_2$, shown as a function of $q_6\%$ (percentage truncated from upper tail of f_6), g_f (proportion of spuriously fast trials), g_s (proportion of spuriously slow trials), a_f (proportion of spuriously fast trials escaping truncation), a_s (proportion of spuriously slow trials escaping truncation), and $p_1\%$ (percentage truncated from lower tail of f_1). Note that the value of a_f is irrelevant when $g_f = 0$, and similarly for a_s and g_s .

subjects tested in three different 2×2 factorial experiments.⁹ We obtained maximum likelihood estimates of μ , σ , and τ separately for each subject and cell, and the ranges of obtained parameter estimates are summarized in Table 9.

Given these 16 sets of 12 parameter estimates, we proceeded to examine the distorting effects of truncation as follows. First, to ensure that all the distortion was due to truncation, we adjusted each subject's estimates of μ for the four cells so that additivity of means was perfect for the untruncated distributions; the adjustments were small for every subject. Next we obtained for each subject the lower cutoffs L_p truncating $p = 0, .5, 1, 2, 4$, or 8% from the lower tail of density f_{11} , and upper cutoffs U_q truncating the same proportions from the upper tail of density f_{22} . Using all possible pairs of these lower and upper cutoffs for a given subject, we computed the distortion of additivity D resulting from that combination of lower and upper truncation cutoffs, resulting in the different conditions (p_{11}, q_{22}) of truncation shown in Table 10 being examined separately for each subject. The distortion from additivity with the appropriately truncated distributions was computed for each subject, and the values in Table 10 show the means, minima, and maxima (across subjects) of the distortions thus computed.

The results are not entirely clear-cut. The largest mean distortion was -5.7% , which is barely large enough to be detected in typical RT experiments. On the other hand, the minimum and maximum distortion effects were fairly severe, showing 10% – 15% distortions toward either under- or overadditive interactions. Thus, although the average distortion effects were relatively small across these three experiments, it is certainly possible that considerably larger

distortion effects could be obtained in other experiments with reasonable parameter combinations. Examination of the distortions obtained for individual subjects indicated that distortions tended to be most severe when (a) there was more overlap of RT distributions in different cells of the design and (b) factors influenced τ rather than μ .

We also examined the effects of truncation on the pattern of variances across the four cells of the designs. With the parameter values summarized in Table 9, variances were not generally additive in the first place; distortions of additivity D , computed in terms of effects on variances rather than effects on means, ranged over approximately $\pm 150\%$. Nonetheless, we looked at how truncation altered the original D values. As in the paradigm examining linearity of RT, truncation from below had relatively little effect on the pattern of variances. Truncation from above, on the other hand, changed variance patterns fairly dramatically in the direction of underadditivity, typically reducing D by 50% – 200% .

No Truncation, Contamination. If spurious RTs are present but truncation is not used, additive models are impervious to contamination effects in mean RT but not in the variance of RT, as were underlying linear models. For the

⁹ We thank Saul Sternberg and Seth Roberts for providing these data, which are those of the *Detection, Identification, $n_a = 2$* , and *Identification, $n_a = 8$* experiments discussed by Roberts and Sternberg (1992). The numbers of RTs per subject per cell were approximately 95, 175, and 250 in the three experiments. Following Roberts and Sternberg, we omitted from all analyses one subject who showed virtually no effect of one experimental factor.

Table 9
Mean, Minimum (min), and Maximum (max) of Estimated Parameter Values, μ , σ , and τ , for 16 Subjects in Each Cell of Three 2×2 Experiments

Factor B level	Parameter	Factor A level					
		1			2		
		<i>M</i>	Min	Max	<i>M</i>	Min	Max
Level 1	μ	271.1	171.7	408.9	313.6	202.0	550.0
	σ	23.9	10.0	56.6	25.4	8.5	59.8
	τ	14.4	5.6	31.6	25.0	11.6	44.1
Level 2	μ	293.3	178.3	448.4	335.8	198.2	568.4
	σ	24.9	8.2	62.7	28.0	6.6	60.3
	τ	22.9	5.6	53.0	33.2	14.1	73.7

condition-independent mixture model, the mean contaminated RT is

$$E[\text{mix}(\mathbf{T}_f, \mathbf{T}_{ab}, \mathbf{T}_s)] = g_f \cdot E[\mathbf{T}_f] + g_s \cdot E[\mathbf{T}_s] + (1 - g_f - g_s) \cdot E[\mathbf{R}] + (1 - g_f - g_s) \cdot E[\mathbf{Z}_a] + (1 - g_f - g_s) \cdot E[\mathbf{Z}_b]. \quad (65)$$

Thus, additivity of factor effects (i.e., on \mathbf{Z}_a and \mathbf{Z}_b) will be preserved as long as contamination is invariant across conditions, although the main effect of each factor will be attenuated by a factor of $(1 - g_f - g_s)$. A similar result is obtained for the delayed-startup mixture model

$$E[\text{mix}(\mathbf{T}_f, \mathbf{T}_{ab}, \mathbf{T}_s)] = g_f \cdot E[\mathbf{T}_f] + g_s \cdot E[\mathbf{D}] + (1 - g_f) \cdot E[\mathbf{R}] + (1 - g_f) \cdot E[\mathbf{Z}_a] + (1 - g_f) \cdot E[\mathbf{Z}_b]. \quad (66)$$

Again, additivity of factor effects is preserved. Only spuriously fast RTs attenuate the main effect of each factor.

As was found with linearity, however, additive factor effects on RT variance are destroyed by contamination. From Equation 63 and the common assumption that \mathbf{Z}_a , \mathbf{Z}_b , and \mathbf{R} are pairwise uncorrelated, the true variance function $\text{Var}[\mathbf{T}_{ab}] = \text{Var}[\mathbf{Z}_a] + \text{Var}[\mathbf{Z}_b] + \text{Var}[\mathbf{R}]$ shows additive effects on variance.

The elimination of variance additivity by contamination is most easily shown for the condition-independent mixture model in Equation 45. Given that the true RTs of the mixture conform to the additive model in Equation 63, the variance in each experimental condition can be computed from Equation 11:

$$\begin{aligned} \text{Var}[\text{mix}(\mathbf{T}_f, \mathbf{T}_s, \mathbf{T}_{ab})] &= g_f \cdot E[\mathbf{T}_f]^2 + (1 - g_f - g_s) \cdot (E[\mathbf{Z}_a] \\ &+ E[\mathbf{Z}_b] + E[\mathbf{R}])^2 + E[\mathbf{T}_s] \cdot g_s - \{E[\mathbf{T}_f] \cdot g_f + (E[\mathbf{Z}_a] \\ &+ E[\mathbf{Z}_b] + E[\mathbf{R}]) \cdot (1 - g_f - g_s) + E[\mathbf{T}_s] \cdot g_s\}^2 \\ &+ g_f \cdot \text{Var}[\mathbf{T}_f] + g_s \cdot \text{Var}[\mathbf{T}_s] + (1 - g_f - g_s) \cdot (\text{Var}[\mathbf{Z}_a] \\ &+ \text{Var}[\mathbf{Z}_b] + \text{Var}[\mathbf{R}]). \quad (67) \end{aligned}$$

Rearranging Equation 67 yields the nonadditive term

$$2(1 - g_f - g_s)(g_f + g_s) \cdot E[\mathbf{Z}_a] \cdot E[\mathbf{Z}_b], \quad (68)$$

which destroys the true additivity of factor effects on vari-

ance, producing overadditivity instead, except in the degenerate case where $E[\mathbf{Z}_a]$ and $E[\mathbf{Z}_b]$ do not vary across conditions of the factorial design. It is interesting to note that the multiplier of $E[\mathbf{Z}_a] \cdot E[\mathbf{Z}_b]$ in the nonadditive term, and hence the distortion of additivity, depends only on the percentages of spurious RTs and not on their distributional properties.

Similar variance perturbations are obtained for the delayed-startup mixture model. With this model, the distortion term is $2(1 - g_f)g_f E[\mathbf{Z}_a] \cdot E[\mathbf{Z}_b]$, which depends only on the percentage of spuriously fast RTs. In principle, the different distortions of variance additivity produced by the delayed-onset and condition-independent mixture models could be useful in testing between these contamination models.

Truncation Plus Contamination. The combined effects of truncation and contamination on additive patterns of RT were investigated by including spuriously fast, slow, or both types of RTs into the same distributions of valid RTs examined in Table 10. The method used was similar to that used in the previous examination of the combined effects on linear RT functions. For the parameter values estimated for each of the 16 subjects included in Table 10, truncation cutoffs were adjusted to give p_{11} and q_{22} values of .005, .01, .02, .04, and .08, varied orthogonally, as in the earlier table. For each subject and combination of p_{11} and q_{22} , the distortion of additivity was examined for 144 different mixture models defined by a factorial design based on four values of g_f (.0, .025, .05, .1), the same four values of g_s , three values of a_f (.1, .2, .3), and the same three values of a_s . The distribution of spuriously fast RTs was assumed to be ex-Gaussian with parameters $\mu = u_f$, $\sigma = 30$, and $\tau = 10$, where u_f was adjusted to give the desired probability of acceptance a_f (given fixed values of σ and τ , u_f is uniquely determined by the selected values of a_f and the lower truncation cutoff dictated by p_{11}). The distribution of spuriously slow RTs was assumed to be ex-Gaussian with parameters $\mu = u_s$, $\sigma = 110$, and $\tau = 100$, where u_s was similarly adjusted to give the desired acceptance probability a_s for the predetermined upper truncation cutoff.

Table 11 shows the obtained mean, minimum, and maximum (across 16 subjects) distortions of additivity D , for a subset of the conditions examined. The table shows results for three values of q_{22} , because, as with uncontaminated truncated distributions (Table 10), the extent of truncation from

Table 10
Mean, Minimum, and Maximum Distortion of Additivity, D, as a Function of the Proportions (p, q) Truncated From the Lower Tail of the Fastest Condition and the Upper Tail of the Slowest Condition

p (%)	Distortion measure	q (%)					
		0.0	0.5	1.0	2.0	4.0	8.0
0.0	M	0.2	-0.6	-1.1	-2.0	-3.6	-5.7
	Minimum	-0.2	-2.4	-2.9	-5.2	-9.9	-14.8
	Maximum	2.4	2.0	1.8	1.5	1.5	1.1
0.5	M	0.7	0.1	-0.7	-1.4	-2.9	-5.0
	Minimum	-0.4	-1.0	-2.2	-4.1	-7.3	-12.1
	Maximum	3.9	4.4	1.5	1.7	1.7	1.3
1.0	M	1.1	0.4	-0.1	-1.1	-2.6	-4.7
	Minimum	-0.7	-2.2	-3.4	-4.9	-7.4	-11.5
	Maximum	10.4	8.7	8.3	5.9	3.4	1.5
2.0	M	1.1	0.4	-0.2	-1.1	-2.6	-4.8
	Minimum	-0.5	-1.8	-2.8	-4.5	-6.9	-10.5
	Maximum	5.1	3.7	2.4	2.1	2.1	1.7
4.0	M	1.9	1.1	0.6	-0.4	-1.9	-4.1
	Minimum	-0.5	-2.0	-3.1	-4.8	-7.2	-9.9
	Maximum	8.4	6.8	5.6	3.6	2.6	2.2
8.0	M	3.0	2.3	1.8	0.8	-0.7	-2.9
	Minimum	-0.9	-2.5	-3.6	-5.4	-7.9	-10.7
	Maximum	12.8	11.3	10.1	8.2	5.1	3.0

Note. p = proportion truncated from lower tail of fastest condition; q = proportion truncated from upper tail of slowest condition.

above had a large effect on the distortion of additivity. The table shows results only for extreme values of the other parameters (g_f , g_s , a_f , a_s , and p_{11}), however, because these had smaller and monotonic effects on D .

The results indicate that these reasonable amounts of contamination have truly negligible effects on the distortion of additivity that is caused by truncation. For any given values of p_{11} and q_{22} , virtually identical distortions are obtained regardless of the percentages and acceptance probabilities of fast and slow spurious observations. Indeed, these distortions are nearly identical to those obtained with the same values of p_{11} and q_{22} with no truncation (cf. Table 10).

Summary: Truncation, Contamination, and Model Testing

Somewhat paradoxically, our analyses strongly suggest that researchers should ignore the problem of spurious observations when testing the predictions of linear or additive models for mean RT, and that it may be quite important that truncation not be used to exclude outliers. Under very realistic assumptions about the true distributions of RTs, truncation can clearly perturb the results enough to muddy the waters with respect to either of these types of model tests. Truncation produces particularly large biases in tests of linearity, and it appears extremely likely that truncation would alter the results enough to produce invalid conclusions. The biases introduced by truncation are approximately the same whether or not spuriously fast and slow RTs are present in the data set, so there is no reason to believe that residual effects of untruncated spurious observations provide any protection from truncation bias. Perhaps

most important, there is no major advantage to be gained from truncation in the first place. If the mean RTs on valid trials follow linear or additive patterns, then, under fairly general and reasonable assumptions about the nature of the spurious RTs, it is likely that they will still do so even if some responses that are spuriously fast and slow contaminate the data set. Thus, the attempt to remove such spurious observations by truncation is likely to cause more harm than the spurious observations themselves, which only introduce noise, not bias.

Unfortunately, researchers attempting to test predictions about RT variance face a much more difficult situation, because in this case there are potentially very harmful effects, both of including spurious observations in the analysis and of using truncation to eliminate them. At this point, the only recommendation we can make for researchers interested in evaluating such predictions is to take extreme caution that the experimental procedure minimizes the number of spurious observations.

Additional Considerations

How Does Truncation Distort Hazard Functions?

Hazard functions with uncontaminated distributions. The hazard function of an RT distribution can be represented as a plot, as a function of time, of the probability that the response will occur in the next small unit of time given that it has not already occurred. Hazard functions reveal properties of RT distributions that are very difficult to detect in standard PDF plots (see Luce, 1986, and the references therein). In fact, although many theoretically different PDFs of RTs look very similar in shape (usually bell shaped and

Table 11

Mean, Minimum (min), and Maximum (max) Distortion of Additivity, D , as a Function of the Proportions Truncated From the Lower Tail of the Fastest Condition and the Upper Tail of the Slowest Condition, g_f , g_s , a_f and a_s

Proportion					Value of q_{22} (%)								
					.5			2			8		
g_f	g_s	a_f	a_s	p_{11} (%)	M	Min	Max	M	Min	Max	M	Min	Max
0	.1		.1	.5	0.0	-1.0	2.9	-1.4	-4.1	2.0	-5.2	-12.0	1.4
0	.1		.3	.5	0.1	-0.9	3.3	-1.4	-4.0	2.0	-5.1	-11.9	1.3
.1	0	.1		.5	0.0	-1.0	2.8	-1.5	-4.2	2.0	-5.3	-12.1	1.3
.1	0	.3		.5	0.0	-1.0	2.8	-1.5	-4.2	2.0	-5.4	-12.3	1.3
.1	.1	.1	.1	.5	0.0	-1.0	2.8	-1.4	-4.1	2.0	-5.2	-12.1	1.3
.1	.1	.1	.3	.5	0.1	-0.9	3.3	-1.4	-4.0	2.0	-5.2	-12.0	1.3
.1	.1	.3	.1	.5	0.0	-1.0	2.8	-1.5	-4.1	2.0	-5.4	-12.3	1.3
.1	.1	.3	.3	.5	0.1	-0.9	3.3	-1.4	-4.1	2.0	-5.3	-12.2	1.2
0	.1		.1	8.	2.5	-2.4	11.7	1.0	-5.3	8.5	-2.7	-10.6	3.0
0	.1		.3	8.	2.8	-2.1	12.3	1.2	-5.0	9.1	-2.5	-10.3	3.1
.1	0	.1		8.	2.3	-2.5	11.3	0.9	-5.3	8.2	-2.9	-10.8	3.0
.1	0	.3		8.	2.4	-2.4	11.3	0.9	-5.3	8.2	-2.9	-10.9	3.0
.1	.1	.1	.1	8.	2.5	-2.3	11.7	1.0	-5.2	8.6	-2.7	-10.6	3.0
.1	.1	.1	.3	8.	2.9	-2.1	12.4	1.3	-4.9	9.2	-2.5	-10.4	3.1
.1	.1	.3	.1	8.	2.6	-2.3	11.7	1.1	-5.2	8.5	-2.8	-10.8	3.0
.1	.1	.3	.3	8.	3.0	-2.0	12.4	1.4	-4.9	9.1	-2.6	-10.5	3.1

Note. Note that the value of a_f is irrelevant when $g_f = 0$, and similarly for a_s and g_s . p and q = proportions truncated from lower tail of fastest condition and upper tail of the slowest condition, respectively; g_f = proportion of spuriously fast trials; g_s = proportion of spuriously slow trials, a_f = proportion of spuriously fast trials escaping truncation; a_s = proportion of spuriously slow trials escaping truncation.

skewed to the right), the corresponding shapes of their hazard functions may differ quite remarkably (see examples provided by Luce, 1986, p. 18) and thus provide additional information. Unfortunately, hazard functions must be interpreted cautiously when the distribution of RTs has been truncated, because truncation can drastically alter the shape of the hazard function, as is illustrated in this section.

The hazard function h of a random variable \mathbf{T} is defined as

$$h(t) = \frac{f(t)}{1 - F(t)}, \quad (69)$$

where functions F and f denote the CDF and the PDF of \mathbf{T} . Hence, h characterizes the instantaneous probability of occurrence of a response at $\mathbf{T} = t$, conditioned on the lack of a response by that time.

To illustrate the fact that quite similar PDFs may have rather different h s, we note that h is an increasing function of t for the special Erlangian distribution, a constant function for the exponential distribution, a first increasing and then decreasing function for the lognormal, and an increasing function for the ex-Gaussian (see Barlow & Proschan, 1965; pp. 14-15; Luce, 1986, pp. 507-511).

If the PDF of \mathbf{T} is truncated, then the associated hazard function is given by the following (see Elandt-Johnson & Johnson, 1980, pp. 54-55):

$$h(t|L \leq \mathbf{T} \leq U) = \frac{f(t)}{F(U) - F(t)} \quad (70)$$

$$= h(t) \cdot \frac{1 - F(t)}{F(U) - F(t)}. \quad (71)$$

The shape of the true hazard function is not altered by truncation of scores from the bottom tail of the distribution, as is obvious from the fact that the hazard function for the truncated distribution does not depend on L .

However, truncation of the upper tail can considerably distort the true shape. Consider, for example, the exponential distribution. The constant value of its hazard function is equal to its rate λ , $h(t) = \lambda$, but the hazard function of the truncated distribution equals

$$h(t|\mathbf{T} \leq U) = \lambda \cdot \{1 - \exp[\lambda \cdot (t - U)]\}^{-1}. \quad (72)$$

This function increases with t , and in fact the increase is quite rapid as t nears the upper cutoff U . Thus, truncation can produce substantial shape distortions of true hazard functions, especially near the upper cutoff. Clearly, hazard functions will have limited usefulness in RT research when RT distributions have to be truncated from above, because the shape of hazard functions is not robust against this truncation.

Hazard functions with contaminated distributions. When some contamination by outliers remains even after truncation of the RT distributions, the spurious RTs and truncation will both act to distort the shape of the true hazard function. The exact nature of the combined shape distortion is not clear, but an optimist might hope that these effects would fortuitously counteract each other, thereby allowing interpretation of observed hazard functions with contaminated, truncated RT distributions, even though uncontaminated truncated distributions yield seriously distorted hazard functions. As might be expected, our investigations indicate that this optimism is ill-founded.

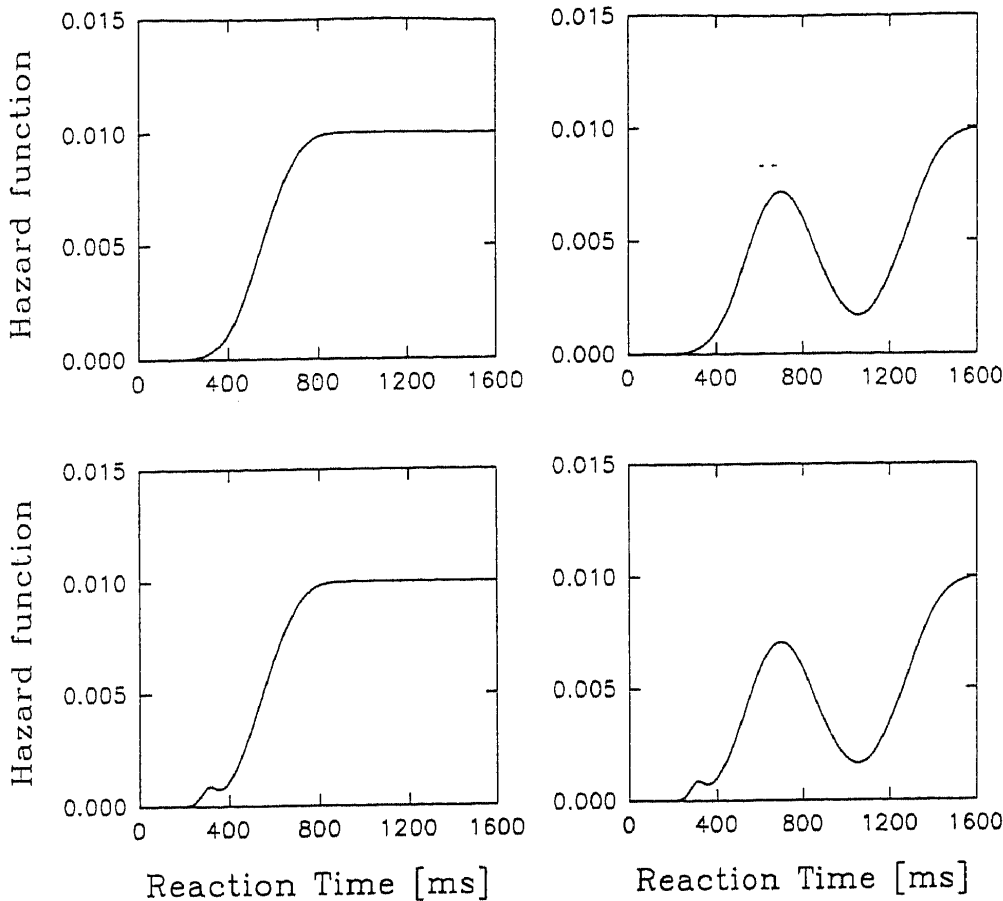


Figure 7. Hazard functions for four combinations of mixture probabilities (g_f, g_s). Left upper panel: (0.00, 0.00). Right upper panel: (0.00, 0.05). Left lower panel: (0.05, 0.00). Right lower panel: (0.05, 0.05). The corresponding probability density functions are shown in Figure 6. (See text for more details.)

The hazard function of the mixture distribution given by Equation 8 is

$$h_o(t) = \frac{g_f h_f(t) S_f(t) + (1 - g_s - g_f) h(t) S_v(t) + g_s h_s(t) S_s(t)}{g_f S_f(t) + (1 - g_s - g_f) S_v(t) + g_s S_s(t)}, \quad (73)$$

where $S_i(t) = 1 - F_i(t)$ for $i = f, s, v$ (see Bain, 1978, p. 117–118). Note that h_o can be interpreted as the weighted average of h_f, h_v , and h_s .

We investigated the distorting effects of spuriously fast and slow RTs on the hazard function h of true RTs, using Equation 73 based on the mixture model $T_o = \text{mix}(T_f, T_v, T_s)$. Figure 7 shows h_o for four different mixtures of the distributions shown in Figure 6, with mixture probability combinations (g_f, g_s) of (0, 0), (.05, 0), (0, .05), and (.05, .05). It can be seen that the true h , shown in (0, 0), increases to an asymptotic value of $\lambda = .01$.¹⁰ As is to be expected, spuriously fast RTs produce only slight distortions of the true hazard function. Spuriously slow RTs, however, dramatically change the asymptotic property of the hazard function, as can be seen in the panels with (0, .05) and (.05, .05). The effect of spuriously slow RTs is so large, in fact, that it creates doubt about whether anything can be inferred from observed hazard functions, except in experiments where great pains are taken to ensure that such RTs never occur.

Clearly, the combination of truncation and spuriously slow RTs does not magically solve the problem, as is illustrated by Figure 8, which shows hazard functions for the same distributions truncated at $t = 1,400$. (Similar results are obtained with other truncation points.) In the present reasonable case, at least, it appears that the distorting effects of truncation and spuriously slow RTs reinforce each other rather than compensating, so the observed hazard functions have shapes nothing like the true, underlying ones.

How Is Statistical Power Affected by Truncation?

Power with uncontaminated distributions. It is well known that the power of a statistical test increases with sample size. In particular, a given difference in population mean RTs of two experimental conditions is more likely to yield a statistically significant difference between sample means with large samples than with small ones. As truncation decreases sample sizes, it appears that a given mean RT difference should be more difficult to detect with truncation than without it. We investigated this question to see whether truncation seriously decreases statistical power.

¹⁰ See Appendix C for a discussion of models in which hazard functions approach asymptotic values.

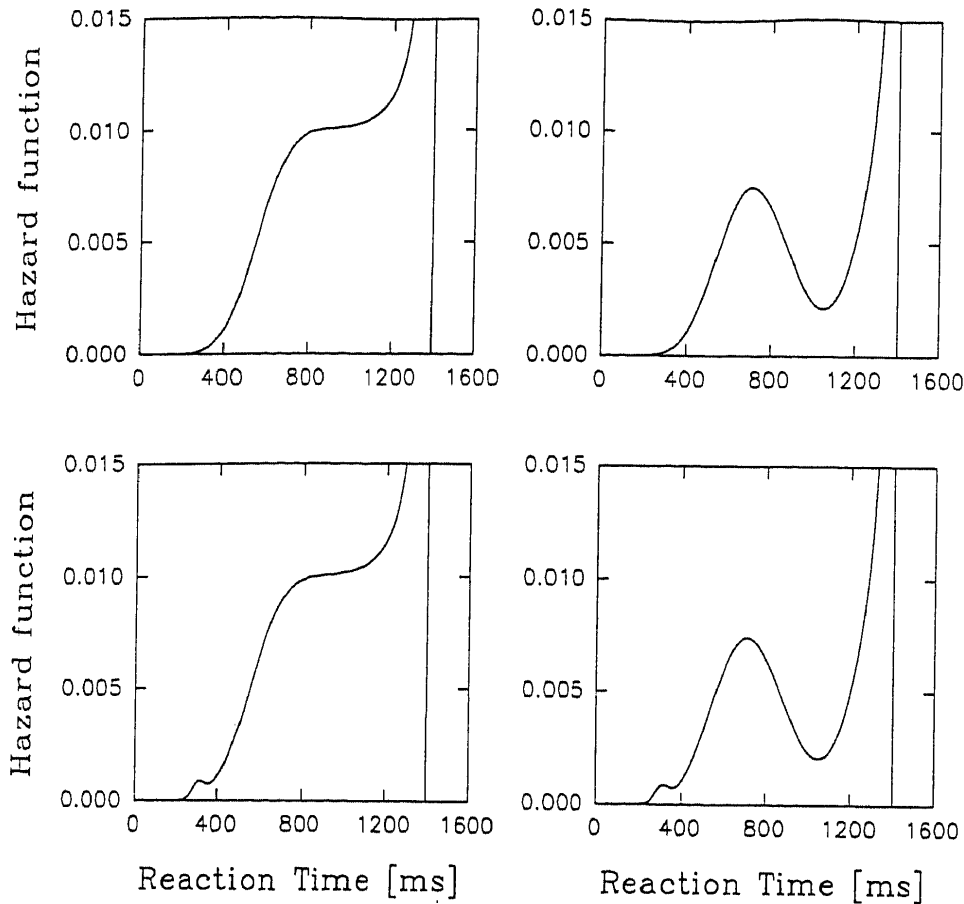


Figure 8. Hazard functions of truncated probability density functions for four combinations of mixture probabilities (g_f, g_s). The truncation point in each panel is 1,400. Left upper panel: (0.00, 0.00). Right upper panel: (0.00, 0.05). Left lower panel: (0.05, 0.00). Right lower panel: (0.05, 0.05). The corresponding untruncated probability density functions are shown in Figure 6. (See text for more details.)

The investigations were carried out using Monte Carlo simulations, and they were limited to the case of experiments comparing two conditions with different mean RTs. In a first set of simulations we examined truncation effects in the limiting case without spurious observations (i.e., when researchers truncate because they incorrectly suspect that there may be outliers). In a second set of simulations described next, we examined truncation effects when spurious observations were actually present.

As the basis for each simulation in the first set, we assumed two untruncated PDFs, f_x and f_y , with the mean $E[Y]$ of f_y being larger than the mean $E[X]$ of f_x . The simulations implemented a scenario in which two researchers analyzed the same data. Researcher A used untruncated samples in the analysis, whereas Researcher B used truncated samples. Thus, independent random samples X_1, \dots, X_N and Y_1, \dots, Y_N were drawn from f_x and f_y , respectively. Each researcher tested the null hypothesis $H_0: E[Y] = E[X]$ against the alternative hypothesis $H_1: E[Y] > E[X]$, using a z test:

$$Z = \frac{M_y - M_x}{\sqrt{\frac{SD[Y]^2 + SD[X]^2}{N}}} \quad (74)$$

where M_x and M_y are the sample means and $SD[X]$ and $SD[Y]$ are the population¹¹ standard deviations of the distributions f_x and f_y , respectively.

For any given set of sample values, Researcher A rejected H_0 in favor of H_1 if the statistic Z was larger than a given threshold value (e.g., 1.65). Researcher B analyzed the same samples X_1, \dots, X_N and Y_1, \dots, Y_N used by Researcher A, and he also tested H_0 against H_1 . However, unlike Researcher A, Researcher B included in the z test only those sample values within a prespecified interval (e.g., between 100 ms and 2,000 ms). Clearly, Researcher B tended to have smaller

¹¹ In a real experiment, the values of $SD[X]$ and $SD[Y]$ are of course unknown, and they are therefore estimated by the corresponding sample values. In this case, the random variable Z only approaches a standard normal distribution for large sample sizes, and a t test is often used instead of a z test. Nonetheless, we used the z test for our simulations, because this test allowed us to more easily equate the power of the test across the various untruncated conditions; for sample sizes larger than 30, which we used, the z and t tests are nearly identical anyway. Furthermore, our simulation results were nearly uninfluenced by whether population standard deviations or corresponding sample estimates were used for computing the Z statistics, so we only report the results with sample standard deviations involved in computing Z .

samples for the z test than Researcher A, and having a smaller sample tends to reduce the power of B's test relative to the power of A's. In addition, the difference in means of the truncated distributions $E[\mathbf{Y}_*] - E[\mathbf{X}_*]$ may well be smaller than the difference in means of the untruncated distributions $E[\mathbf{Y}] - E[\mathbf{X}]$, which would further lower B's power. On the other hand, the standard deviations $SD[\mathbf{X}_*]$ and $SD[\mathbf{Y}_*]$ of the truncated populations are smaller than those of the untruncated populations. Because the power of Equation 74 increases as population standard deviations decrease, this could partly compensate for the power loss produced by the reduction in sample size and mean difference. Hence, the question arises: How much is the power of Researcher B's statistical test (with truncation) lowered relative to the power of Researcher A's test (without truncation)?

We ran Monte Carlo simulations to investigate this question under a number of conditions. The main dependent variable of the simulations was the statistical power for Researcher B's analysis (with truncation). The main features of these simulations were as follows:

1. Independent samples of size N were drawn from f_x and f_y , which were Erlangian distributions with shape parameters m_x and $m_y = m_x + 1$ and rates $\lambda_x = \lambda_y$ (additional simulations using ex-Gaussian distributions are reported in the next section). Across simulations the parameter m_x was set to 2, 4, or 6, and the rates were adjusted so that the mean of f_x was always equal to 150 ms. The sample size N in each simulation was adjusted so that a power $1 - \beta = .8$ of the z test was obtained in the untruncated case.¹²

2. The lower truncation point L was set at a value that would truncate one of six proportions from the corresponding tail of f_x : $p_x = 0, .005, .01, .02, .04, \text{ or } .08$. Analogously, the upper truncation point U was set to cut off the same proportions from the upper tail of f_y , and p_x and q_y were varied orthogonally. The percentages truncated from the other two tails (i.e., top of f_x and bottom of f_y) were uniquely determined by the truncation points U and L .

3. Estimates of the power of the truncated analysis were obtained for each combination of levels of the independent variables. Each estimate was based on 10,000 repetitions of the following series of steps: (a) Two random samples of size N , $\mathbf{X}_1, \dots, \mathbf{X}_N$ and $\mathbf{Y}_1, \dots, \mathbf{Y}_N$, were drawn from the untruncated PDFs f_x and f_y , respectively. (b) The \mathbf{Z} statistic was computed using all of the observations, and it was checked to see whether the event $\{\mathbf{Z} > 1.65\}$ had occurred. This enabled us to estimate the statistical power for the untruncated analysis. Because the probability of this event could be computed from the assumptions, this served as a check on the calculations. (c) Observations outside of the interval $[L, U]$ were excluded, and the \mathbf{Z}_* statistic was computed for the reduced sample. This value was checked to see whether the event $\{\mathbf{Z}_* > 1.65\}$ had occurred. This permitted us to estimate the statistical power for the truncated analysis.

Table 12 summarizes the estimated power for truncated PDFs of the special Erlangian as a function of p_x , q_y , and m_x . It can be observed that the power of the z test is clearly affected by excluding extreme values. However, this effect is asymmetrical: Power is reduced much less if extreme values from the upper tails of the RT distributions are discarded than if values from the lower tails are discarded.

Table 12
Estimated Power as a Function of m_x , p_x %, and q_y %

p_x (%)	m_x	q_y (%)					
		0.0	0.5	1.0	2.0	4.0	8.0
0.0	2	.81	.80	.80	.81	.79	.78
	4	.81	.80	.80	.79	.78	.76
	6	.80	.80	.80	.78	.77	.75
0.5	2	.80	.80	.80	.79	.79	.76
	4	.79	.79	.78	.78	.76	.74
	6	.79	.78	.78	.77	.74	.73
1.0	2	.80	.79	.79	.79	.78	.77
	4	.79	.79	.78	.77	.76	.74
	6	.79	.76	.77	.76	.74	.71
2.0	2	.78	.78	.78	.77	.76	.75
	4	.77	.77	.76	.75	.74	.71
	6	.76	.76	.75	.73	.72	.70
4.0	2	.75	.76	.76	.74	.73	.73
	4	.74	.74	.75	.72	.71	.67
	6	.74	.73	.72	.71	.69	.67
8.0	2	.72	.73	.72	.70	.69	.66
	4	.70	.70	.69	.68	.65	.61
	6	.69	.68	.68	.67	.64	.59

Note. m_x = shape parameter of f_x ; q_y = proportion truncated from upper tail of f_y ; p_x = proportion truncated from the lower tail of f_x .

The exclusion of extremely large observations reduces variance more than the exclusion of extremely small observations (because of the skew), and power increases as variance decreases. In fact, there is almost no power loss if extreme values from the upper tails are discarded, especially for the more skewed distributions (i.e., smaller values of m_x and m_y).

Power with contaminated distributions. It is also appropriate to consider the case in which some spurious observations contaminate the RT distributions even after truncation. Thus, we conducted a second set of simulations in which observations were taken from mixture distributions of valid RTs, spuriously fast RTs, and spuriously slow RTs (Equation 8). We assumed that identical processes would produce spurious RTs in the two experimental conditions, so both the proportions and the distributions of spuriously fast and slow RTs were equated across the two conditions.

As in the previous simulations examining power, there were two PDFs of valid RTs, denoted by f_x and f_y , corresponding to the two different experimental conditions, and

¹² Given that the Central Limit Theorem holds approximately for sample means M_x and M_y , the relation between sample size N and power $(1 - \beta)$ is $N = \{(\text{Var}[\mathbf{X}] + \text{Var}[\mathbf{Y}] / (E[\mathbf{X}] - E[\mathbf{Y}])^2) \times [\Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta)]^2$, where function Φ^{-1} denotes the inverse CDF of the standard normal distribution and α the Type I error. For example, with $\alpha = .05$, $\beta = .10$, $SD[\mathbf{X}] = 50$, $SD[\mathbf{Y}] = 80$, $E[\mathbf{X}] = 400$, $E[\mathbf{Y}] = 440$ one obtains

$$\begin{aligned}
 N &= \frac{50^2 + 80^2}{(400 - 440)^2} \times [\Phi^{-1}(.95) + \Phi^{-1}(.90)]^2 \\
 &= 5.56 \times [1.65 + 1.28]^2 \\
 &= 47.7 \\
 &\approx 48.
 \end{aligned}$$

Thus a sample size of $N = 48$ is expected to yield the desired power of $\text{Pr}\{\mathbf{Z} > z_{.05} | H_1\} = 1 - \beta \approx .90$ in this example.

Table 13
*Estimated Power P_A for Untruncated Samples
 as a Function of g_s and g_f*

g_f (%)	Measure	g_s (%)			
		0.0	2.5	5.0	10.0
0.0	P_A	.80	.63	.51	.38
	$\max[P_B]$.80	.79	.77	.72
	p_x	.00	.00	.00	.00
	q_y	.00	.04	.08	.10
2.5	P_A	.75	.58	.49	.34
	$\max[P_B]$.77	.74	.72	.67
	p_x	.02	.04	.02	.04
	q_y	.00	.04	.08	.16
5.0	P_A	.70	.55	.46	.32
	$\max[P_B]$.74	.72	.69	.64
	p_x	.04	.04	.08	.08
	q_y	.00	.04	.04	.16
10.0	P_A	.61	.47	.38	.28
	$\max[P_B]$.70	.67	.65	.58
	p_x	.08	.08	.08	.08
	q_y	.00	.04	.08	.16

Note. $\max[P_B]$ denotes the maximal power for each combination of g_s and g_f under an optimal selection of the values p_x and q_y . P_A = power for untruncated samples; g_s = proportion of spuriously slow trials; g_f = proportion of spuriously fast trials; p_x = proportion truncated from lower tail of f_x ; q_y = proportion truncated from upper tail of f_y .

the mean of f_y was larger than the mean of f_x . Thus, the observed RTs in the two conditions were given by the mixtures f_{ox} and f_{oy} . Again, Researcher A used untruncated samples and Researcher B used truncated samples, and both evaluated the null hypothesis $H_0: E[Y] = E[X]$ against the alternative hypothesis $H_1: E[Y] > E[X]$ using a z test. As before, the dependent variable of the simulation was the statistical power for each researcher's analysis.

The parametric assumptions and independent variables of this set of simulations were as follows:

1. In all simulations, ex-Gaussians were used for the distributions of valid RTs f_x , f_y , and for the distributions of spuriously fast and slow RTs f_f and f_s , with parameter settings of ($\mu_x = 500$, $\sigma_x = 100$, $\tau_x = 100$), ($\mu_y = 545$, $\sigma_y = 105$, $\tau_y = 105$), ($\mu_f = 300$, $\sigma_f = 30$, $\tau_f = 10$), and ($\mu_s = 1200$, $\sigma_s = 110$, $\tau_s = 100$), respectively. With these parameters the true mean and standard deviation are 600 and 141 for f_x and 650 and 148 for f_y . A sample size of 104 is needed to detect this mean difference with a power probability of .8 in the untruncated and uncontaminated case. This basic sample size was kept constant across all simulations, although the effective sample size of Researcher B was always smaller, depending on the extent of truncation.

2. The truncation points L and U were set to values that cut off proportions p_x and q_y from the lower tail of f_{ox} and the upper tail of f_{oy} , respectively, and p_x and q_y were varied factorially over the values 0, .01, .02, .04, .08, and .16.

3. The mixture probabilities g_f and g_s were also varied factorially, using values of 0, .025, .05, and .1.

4. Estimates of power were obtained for the 576 combinations of the independent variables p_x , q_y , g_f , and g_s . Each estimate was based on 10,000 repetitions of the same three steps described for the first simulation set.

Table 13 summarizes the main results of the simulation. For each combination (g_f , g_s), the table presents the estimated power for Researcher A, P_A , who used contaminated samples untreated by truncation. P_A is attenuated very strongly as the amount of contamination increases, consistent with the intuition on which truncation is based in the first place. For example, it is attenuated by more than 20% for the relatively moderate values $g_f = .025$ and $g_s = .025$. As expected, P_A is more attenuated by increases in g_s than by increases in g_f . The table also presents $\max[P_B]$, which is the maximum power of Researcher B under each combination (g_f , g_s), using optimal values of p_x and q_y , which are also shown in the table.¹³ For example, with $g_f = .025$ and $g_s = .1$, Researcher B's power would be maximal for $p_x = .04$ and $q_y = .16$, namely, $\max[P_B] = .67$. Note that this maximal value is almost twice as large as Researcher A's power ($P_A = .34$). Thus, it is clear from the simulation results that appropriate truncation cutoffs can remove almost all of the negative effect of contamination on statistical power.

Summary of Truncation Effects

Truncation of RT distributions seems to be one of the most common methods of outlier exclusion, presumably because it is conceptually simple and can easily be implemented in computer-controlled experiments. For example, trials on which the measured RT does not fall within a preselected interval are sometimes excluded and repeated later in the same experimental block, thereby assuring a "complete" data set.

The results presented thus far, however, make it clear that various nontrivial distortions can occur when RT distributions are truncated from below, above, or both. We considered both the case in which all spurious observations are successfully eliminated by truncation and the case in which some spurious observations fall within the truncation bounds and therefore contaminate the distribution used for analysis. First, we examined truncation-induced bias in the mean, median, standard deviation, and skewness of RT. This analysis was performed for three important RT distributions: the special Erlangian, the lognormal, and the ex-Gaussian. Interestingly, effects of truncation were much the same for all three. In each case, bias depended on both the skewness of the underlying RT distribution and the amount truncated from above and below. Truncation tended to have a small to moderate biasing effect on the mean and median, but it had rather larger biasing effects on the standard deviation and skewness. With a lot of luck, spurious observations that fall within the truncation boundaries could counteract the effects of truncation on the mean or median, but this appears virtually impossible for the standard deviation and skewness. As discussed further in Appendix A, the tables presented here can be used to estimate bias effects in actual data.

¹³ Note that $\max[P_B]$ was determined on the basis of the 16 combinations of the (p_x , q_y) used in the simulation. Thus, $\max[P_B]$ is expected to be somewhat smaller than the true maximum power.

Second, we examined the distorting effects of truncation on linear RT functions, and the results indicate that meaningful distortions can result under fairly ordinary conditions. Distortions usually make truly linear mean RT functions appear to be negatively accelerated, consistent with a number of empirical findings, and would also tend to make positively accelerated mean RT functions appear more linear. Interestingly, under several plausible models, contamination by spurious observations does not disturb the linearity of a mean RT function, although it can change the slope and intercept. In analyses examining RT variance rather than mean, truncation and contamination both destroy underlying linear functions.

Third, we examined the effects of truncation on RT in factorial experiments. Truncation produced small to moderate underadditive distortions on mean RT, at least within the conditions examined in our analysis. As with linear functions, contamination by spurious observations had no effect on the additivity or interaction of factor effects on mean RT, but additivity of RT variances was seriously perturbed by both truncation and contamination.

Fourth, we analyzed the effects of truncation on hazard functions. These can be substantial, with truncation producing large qualitative changes in a hazard function's shape. Contamination by spurious observations also produces major qualitative changes, and it does not appear that contamination and truncation are likely to counteract each other's effects.

Fifth, we evaluated changes in statistical power as a function of truncation. These changes were remarkably small. Contamination by spurious observations, however, did seriously reduce power, but appropriate truncation could counteract this reduction very effectively.

Given the sizable distortions that can occur when RT distributions are truncated, we next describe a maximum likelihood method for counteracting the effects of truncation and recovering information about the untruncated distribution of RTs from a truncated sample. This method appears quite general, and should therefore be considered seriously whenever truncation is used, even when spurious observations contaminate the data set.

Compensating for Truncation Effects

For certain types of analyses, researchers will find it highly desirable to eliminate spurious observations from the data set, because of these observations' pernicious effects on the analysis being conducted. As just seen, for example, spurious observations can greatly reduce experimental power, and truncation can combat their effects fairly effectively. In addition, spurious observations seriously disturb a model's predictions about the mean and variance of RT, as discussed earlier, so it may be necessary to remove these observations before testing the model's predictions. When truncation is used to remove spurious observations, however, it is necessary to try to correct the analysis to remove the biasing effects that truncation has been found to produce.

Fortunately, if the family of the RT distribution can be specified (e.g., ex-Gaussian), it may be possible to remove

some or all of the truncation bias using a modified maximum likelihood approach. In this section, we show how researchers can estimate the true parameters of the underlying untruncated PDF f from the truncated sample and can then use these parameter estimates to infer (approximately) unbiased statistics of the full untruncated PDF. In addition, simulations indicate that this compensation procedure works surprisingly well even when the family of the RT distribution is unknown and occasional spurious observations occur either inside, outside, or inside and outside the truncation interval.

Intuitively, the likelihood of a data set corresponds to the overall probability of observing that data set. It is computed by assuming a particular underlying model, determining the likelihood of each individual observation given that model, and (with independent observations) finding the product of all the individual likelihoods. For example, with a data set of continuous, independent RTs (t_1, t_2, \dots, t_N) from the ex-Gaussian distribution, the likelihood function is

$$L(t_1, \dots, t_N | \mu, \sigma, \tau) = \left\{ \prod_{i=1}^N f(t_i | \mu, \sigma, \tau) \right\}, \quad (75)$$

where f is the ex-Gaussian probability density function and μ, σ, τ are particular values of its three parameters. Maximum likelihood estimates are simply numerical values of the parameters (in this case μ, σ , and τ) that maximize Equation 75 for a given set t_1, \dots, t_N of observed RTs; in many practical situations, these values can be obtained by numerical search algorithms.

In general, when fast and slow spurious RTs may be present in the data set, the likelihood function for a set of observed RTs is

$$L(t_1, \dots, t_N | \Theta) = \left\{ \prod_{i=1}^N f_o(t_i | \Theta) \right\}, \quad (76)$$

where f_o is the probability density of a mixture, given by Equation 12, and Θ is the set of parameters on which the mixture depends (i.e., g_f, g_s , and all the parameters of the distributions f_f, f_v , and f_s). Unfortunately, to find parameter estimates that maximize this likelihood, researchers would have to specify a particular distributional family (e.g., ex-Gaussian) for the fast and slow spurious RTs, and in practice there is never enough information about the spurious RTs to be able to do this with confidence. Thus, it is not practical to try to maximize the likelihood of the full data set.

Nonetheless, we thought a simplified version of this approach, one that did not require any assumptions about the distributions of spurious RTs, might be useful. We actually examined three different simplifications; we report first and in most detail the one which worked the best. At the end, the others are mentioned briefly.

One simplified approach is to use a likelihood function assuming that there are no spurious observations (i.e., $g_f = g_s = 0$) and still protect against biasing effects of spurious observations (in case the assumption is wrong) by truncating extreme RTs. Theoretically, this approach appears somewhat

self-contradictory, but it could still be useful in practice because the biasing effects of truncating valid RTs could approximately offset the biasing effects of the spurious RTs that are being ignored. In addition, we show later that, when g_f and g_s are not too far from zero, this approach provides an approximation to a more realistic, yet more difficult to minimize, likelihood function given by Equation 80.

Let $f(t|\Theta)$ be the untruncated PDF, which depends on the parameters $\Theta = (\theta_1, \dots, \theta_k)$, let t_1, \dots, t_{N_A} be the observed RTs in the acceptance interval $[L, U]$, and assume that there are N_L observed RTs less than L and N_U observed RTs greater than U (i.e., $N = N_L + N_A + N_U$ observations in total). Under these assumptions, the likelihood function $L(t_1, \dots, t_{N_A}, N_L, N_U|\Theta)$ for the data set is given by (Kendall & Stuart, 1967, Vol. 2, Equation 32.37)

$$L(t_1, \dots, t_{N_A}, N_L, N_U|\Theta) = \{F(L|\Theta)\}^{N_L} \times \left\{ \prod_{i=1}^{N_A} f(t_i|\Theta) \right\} \times \{1 - F(U|\Theta)\}^{N_U}. \quad (77)$$

The parameter values that maximize the likelihood function L , say $\hat{\Theta}$, are the maximum likelihood estimates of parameters Θ , and numerical methods are available for discovering these values (e.g., Press et al., 1986, chapter 10). Once the maximum likelihood estimates have been obtained, researchers can recover estimates of the desired statistics (e.g., mean, median, or variance) of the untruncated PDF f . Assuming the ex-Gaussian distribution, for example, if $\hat{\mu}$ and $\hat{\tau}$ are the maximum likelihood estimates for a given subject and condition, then the maximum likelihood estimate of $E[T]$ is simply $\hat{\mu} + \hat{\tau}$. Analogously, any other statistic can be estimated for the untruncated distribution, given estimates of the parameters of the full distribution. In an actual experiment, then, the researcher could obtain such estimates for each subject and condition, and analyze these estimates to test the hypotheses of interest.

We conducted nine groups of simulations to evaluate this strategy of compensating for truncation effects in studies of mean, or standard deviation of RT. In each group of simulations, valid RTs were generated from one of three distributions (i.e., ex-Gaussian, Erlangian, or lognormal), and the maximum likelihood parameter estimates were obtained for one of the three different models. For example, in one group of simulations the data were generated from an underlying ex-Gaussian distribution and maximum likelihood estimates were obtained for the best-fitting lognormal model for the randomly generated data. In fitting a model other than the one used to generate the data (i.e., fitting the "wrong" model), we intended to examine the robustness of each fitted model to violations of its assumptions about the distribution of the underlying data. We used three-parameter versions of the models both for generating the data and for fitting the simulated RTs, and parameter values used to generate the data were chosen to make the results as comparable as possible across simulations. For the Erlangian, the three parameters were m , λ , and a constant added to all the RTs, which were set to values of 2, 0.013, and 450, respectively, for generating RTs. For the lognormal, the three parameters were μ , σ , and a constant added

to all the RTs, which were set to values of 4.81, 0.64, and 450, respectively, for generating RTs. For the ex-Gaussian, the three parameters were μ , σ , and τ , which were set to values of 500, 20, and 100, respectively, for generating RTs. With these parameter values, the mean of the untruncated distribution was 600 in all three cases, and the standard deviations ranged from 102 to 106.

Within each of the nine groups of simulations, 16 individual simulations were conducted, structured according to a $4 \times 2 \times 2$ factorial design with the following factors:

1. The sample size N of the simulated experiment was 10, 20, 40, or 100.

2. The truncation cutoff was either the 90th or the 68th percentile point of the distribution used to generate the data. Thus, in different simulations either the top 10% or the top 32% of the observations were removed by truncation. Although 32% is clearly an unrealistically large percentage of observations to exclude by truncation in practice, we included this value in the simulations to see if the procedure failed catastrophically at larger truncation percentages. No truncation at the lower end of the distribution was used, because the results of the section Bias Effects of Truncation on Mean, Median, Standard Deviation, and Skewness indicate that such truncation creates smaller biases in the first place (e.g., Figure 2).

3. To check on the robustness of the procedure, the probability of a spuriously slow observation was set to either 0% or 10%. After all, Equation 77 is based on the assumption that only valid RTs are thrown out (i.e., 0% true outliers), but in practice researchers may also exclude some spurious observations exceeding the truncation cutoff. This tends to increase N_U , thereby influencing the parameter estimates that maximize Equation 77, and spurious observations may ultimately bias the maximum likelihood estimates. The simulations with 10% true outliers were conducted to check for such bias, and we assume that the bias obtained in these simulations is a reasonable upper bound on the bias in most common experimental settings. For these simulations, it was assumed that all spuriously slow RTs were sufficiently extreme to exceed the truncation cutoff, so that they would not contaminate the analysis after truncation. The case in which spuriously slow RTs come from a less extreme distribution, and therefore do contaminate the analysis, is considered below.

For each of the 144 different individual simulations described above (3 valid data distributions \times 3 fitted distributions \times 4 sample sizes \times 2 truncation cutoffs \times 2 probabilities of spuriously slow RTs), we simulated 500 identical subjects. The first step in simulating each subject was to generate a sample of N RTs. In the simulations with 0% spuriously slow observations, each RT was randomly generated from the valid data distribution being used for all the subjects in that simulation. In the simulations with 10% spuriously slow observations, each RT was generated from the valid data distribution with probability .9, and a spuriously slow RT exceeding the cutoff was generated with the complementary probability of .1. After the sample had been generated, all N_U RTs exceeding the truncation cutoff were excluded. Then, using those N_A sample RTs remaining after truncation, maximum likelihood estimates were obtained by

maximizing Equation 77 for the model being fitted.¹⁴ Finally, the untruncated mean, median, and standard deviation for that subject were estimated by calculating the mean, median, and standard deviation of the fitted distribution having parameter values equal to the obtained maximum likelihood estimates.

Across the 500 subjects in each simulation, we computed the averages of the untruncated mean, untruncated median, and untruncated standard deviation. To the extent that the procedure is successful in recovering the true, untruncated parameters (i.e., in removing truncation bias), these averages should be close to the true means, medians, and standard deviations of the valid data distributions from which the RTs were generated. For comparability with our earlier analyses (see, e.g., Equation 22 and Figure 2), these computations were summarized by computing relative bias, defined as the percentage deviation of each average (i.e., mean, median, and standard deviation) from the corresponding true value. For example, the relative bias, δ , of the estimated mean $\hat{\mu}$ was defined as

$$\delta \equiv \frac{\hat{\mu} - \mu}{\mu} \cdot 100\%. \quad (78)$$

The results of these simulations are summarized in Tables 14 and 15 for 0% and 10% spuriously slow RTs, respectively. For recovering means of the data distributions, the procedure worked rather well, especially when parameters were estimated by fitting the special Erlangian.¹⁵ With this fitted distribution, relative bias of the estimated means was no more than $\pm 1.5\%$ in any of the different simulation conditions. In contrast, the truncated means themselves had much greater relative biases (averaging -4.3% and -9.2% with 10% and 32% truncation, respectively). Thus, the maximum likelihood procedure yielded means that were much less biased than the means of the truncated samples, suggesting that this procedure might be of considerable practical value.

In addition to having small biases, the procedure appears to be robust for the estimation of means, given that the special Erlangian is used as the fitted distribution. This conclusion is suggested by the fact that the procedure produced good estimates of mean regardless of whether the true distribution of valid RTs was Erlangian, lognormal, or ex-Gaussian distribution and regardless of whether 0% or 10% spuriously slow RTs were present. This robustness suggests that the special Erlangian may safely be used to compensate for truncation effects on the mean in real experiments, when the true underlying distribution and proportion of outliers are unknown.

Finally, the means estimated using the special Erlangian also have acceptably low variance. This was measured by r , the ratio of the standard deviation of the 500 subject means estimated by the maximum likelihood procedure to the standard deviation of the actual subject means calculated from the 500 untruncated samples. With this definition, r values less than 1 indicate that the means obtained from the maximum likelihood procedure were less variable than the means of the samples themselves, and ratios greater than 1 indicate that the maximum likelihood procedure

yielded more variable means. As the r values are close to 1, it is clear using the maximum likelihood procedure to reduce bias does not add substantial random noise to the estimation process.

It is interesting to note that in a few cases the estimated means obtained with the maximum likelihood procedure actually had lower variance than the means of the full sample (i.e., cases with $r < 1$). Clearly, maximum likelihood estimators from a truncated sample use fewer observations than estimators from the untruncated distribution. As the accuracy of the estimation process increases with sample size, one might expect that estimates from a truncated sample would have greater variance than estimates from a full sample. However, this is not necessarily true (Rao, 1958, cited in Kendall & Stuart, 1967). As found here, an increase in estimation efficiency can result under moderate amounts of truncation (Kendall & Stuart, 1967, p. 524).

As with the mean, recovery of the median with the maximum likelihood procedure was also quite good, with relative biases less than $\pm 2.3\%$ when the special Erlangian was used as the fitted distribution. Unfortunately, recovery of the standard deviation was not nearly as successful as recovery of the mean and median. To recover the standard deviation, it seems best to fit the ex-Gaussian distribution, but this procedure still led to relative biases exceeding 10% in some cases. Still, the recovered standard deviations were much less biased than those obtained by using truncated samples (e.g., Figure 2), and so the procedure should still be seriously considered when examining standard deviations of truncated samples.

Compensation With Contaminated Distributions

To examine the situation in which spuriously fast or slow RTs contaminate the observed RT distributions even after truncation, we ran 16 additional sets of simulations in which the observed RTs were mixtures of valid RTs, spuriously fast RTs, and spuriously slow RTs. The spuriously fast and slow RTs came from ex-Gaussian distributions with parameters of ($\mu_f = 300$, $\sigma_f = 30$, $\tau_f = 10$) and ($\mu_s = 1,200$, $\sigma_s = 110$, $\tau_s = 100$), respectively. The 16 sets of simulations differed with respect to the probabilities of spuriously fast and slow RTs, which varied factorially over the values of 0, .025, 0.05, and .1.

Each of the 16 sets of simulations had the structure shown in Table 14 (i.e., 3 distributions used to generate valid RTs \times 3 distributions used to fit maximum likeli-

¹⁴ The first factor on the right side of the equation was always one, as no truncation from below was used.

¹⁵ When fitting the lognormal distribution, the estimation procedure converged to unreasonable parameter values for a small percentage of samples (approximately 0.4% overall, but as many as 4% in some conditions with sample sizes of 10). These parameter values yielded estimates of the untruncated mean greater than 2,000 ms, or else yielded estimates of the untruncated standard deviation that were greater than the mean. These estimates were excluded when computing the average relative biases reported in Tables 14 and 15.

Table 14
Average Relative Biases of Means and Standard Deviations Recovered From Truncated Samples Using Maximum Likelihood Estimation, and a Measure of Stability of Means From Simulations With 0% Spuriously Slow Reaction Times

Fitted distribution	Data distribution	Measure	Sample size							
			Truncate 32%				Truncate 10%			
			10	20	40	100	10	20	40	100
EG	EG	<i>M</i>	-0.1	0.0	0.5	0.2	0.9	0.3	0.1	-0.1
		<i>SD</i>	-12.2	-3.7	2.0	1.3	-6.1	-2.4	0.3	-0.8
		<i>r</i>	1.21	1.32	1.41	1.28	1.18	1.14	1.07	1.10
EG	ER	<i>M</i>	0.9	1.2	0.8	1.0	0.7	0.0	-0.3	-0.1
		<i>SD</i>	3.4	13.9	14.2	19.3	0.7	4.0	5.7	10.4
		<i>r</i>	1.51	1.70	1.59	1.50	1.37	1.15	1.10	1.15
EG	LN	<i>M</i>	0.1	-0.1	0.1	0.2	0.6	-0.1	0.1	-0.2
		<i>SD</i>	-12.5	-8.2	-2.1	-1.9	-12.5	-8.2	-4.2	-3.8
		<i>r</i>	1.10	1.23	1.32	1.22	1.06	0.98	1.05	1.01
ER	EG	<i>M</i>	-1.5	-1.3	-1.2	-1.4	0.5	-0.3	-0.3	-0.5
		<i>SD</i>	-28.3	-22.6	-20.8	-20.3	-18.5	-16.1	-14.5	-13.1
		<i>r</i>	0.96	1.04	0.96	0.96	1.00	0.98	0.99	0.97
ER	ER	<i>M</i>	-1.2	-0.9	-0.7	-0.7	-0.5	-0.4	-0.7	-0.6
		<i>SD</i>	-17.4	-11.8	-8.2	-6.1	-14.6	-9.9	-8.4	-5.3
		<i>r</i>	1.08	1.19	1.15	1.17	1.05	1.05	1.05	1.04
ER	LN	<i>M</i>	-1.0	-1.4	-1.2	-1.3	-0.4	-0.9	-0.6	-0.5
		<i>SD</i>	-27.9	-25.8	-22.7	-21.9	-24.1	-22.9	-19.0	-16.9
		<i>r</i>	1.02	0.94	0.94	0.97	0.94	0.89	0.91	0.90
LN	EG	<i>M</i>	-0.8	-0.9	-1.1	-1.7	0.1	-0.2	-0.4	-0.7
		<i>SD</i>	-18.8	-18.1	-19.4	-20.9	-12.3	-9.6	-11.4	-12.8
		<i>r</i>	1.15	1.24	1.07	1.01	1.06	1.04	0.99	1.00
LN	ER	<i>M</i>	2.4	2.7	3.1	2.1	1.8	1.2	0.7	0.8
		<i>SD</i>	52.3	52.9	57.5	42.8	37.8	30.9	28.7	29.4
		<i>r</i>	2.12	2.11	2.52	2.16	1.53	1.35	1.34	1.34
LN	LN	<i>M</i>	1.3	1.1	0.0	0.0	0.5	0.0	0.1	-0.3
		<i>SD</i>	12.8	7.9	-1.5	-0.4	-2.9	-3.0	-0.8	-2.8
		<i>r</i>	1.72	2.11	1.40	1.87	1.25	1.03	1.05	1.05

Note. EG = ex-Gaussian; ER = Erlangian; LN = lognormal; *r* = measure of stability of means.

hood estimates \times 4 sample sizes \times 2 levels of truncation from the upper tail of the valid RT distribution). Aside from the spuriously fast and slow RTs generated from the ex-Gaussian distributions, no additional outliers were generated, so the simulations were analogous to those in Table 14 rather than Table 15.

The results are summarized in Tables 16 and 17, which present the average relative biases of the recovered means and standard deviations, as well as the stability estimate *r*, as a function of the fitted distribution, the underlying data distribution used for the valid RTs, the percentage of observations truncated, and the probability of slow spurious observations. The tables show results averaged over sample size, which had negligible effects, separately for the simulations with 0% and 10% spuriously fast observations, which produced the most extreme results.

It is evident that the recovered means are still nearly unbiased, with relative bias values not too far from zero, and are reasonably stable, especially when the special Erlangian distribution is fitted. Recovery of the standard deviation was not, however, particularly successful.

Overall, the results of these simulations indicate that the maximum likelihood procedure for recovering the true mean from a truncated data set is rather robust. Even when spuriously fast and slow observations contaminate the RT distribution, the estimation procedure recovers nearly unbi-

ased estimates of the means of the distributions of valid RTs. Moreover, when the data are generated from a mixture distribution, estimated means are in many cases much more stable using truncation and maximum likelihood estimation than using the actual observed mean of all observations, particularly when the mean is estimated using the special Erlangian distribution. Of course, the high variability of the untruncated sample mean obtained from a mixture distribution is one of the main arguments in favor of truncation, but it appears that this maximum likelihood procedure can reduce this variability without introducing the substantial bias that is characteristic of truncation by itself. In sum, the simulations with untruncated spurious observations reinforce the conclusion that a combination of truncation with the maximum likelihood procedure may be a very useful approach in the presence of observations that are spuriously fast, observations that are spuriously slow, or both.

Other Maximum Likelihood Approaches

As noted above, we ran similar sets of simulations to evaluate two other modified maximum likelihood estimation procedures, but neither worked nearly as well as the one reported above. One was to maximize the likelihood function for the observations within the truncation interval, ignoring both the

Table 15
Average Relative Biases of Means and Standard Deviations Recovered From Truncated Samples Using Maximum Likelihood Estimation, and a Measure of Stability of Means From Simulations With 10% Spuriously Slow Reaction Times

Fitted distribution	Data distribution	Measure	Sample size							
			Truncate 32%				Truncate 10%			
			10	20	40	100	10	20	40	100
EG	EG	<i>M</i>	0.7	1.2	1.5	1.2	1.3	1.5	0.8	0.6
		<i>SD</i>	-5.1	3.4	8.5	7.7	-7.1	5.1	4.2	3.5
		<i>r</i>	1.28	1.51	1.56	1.44	1.29	1.38	1.17	1.16
EG	ER	<i>M</i>	1.2	2.2	2.7	2.3	1.6	0.9	0.4	0.3
		<i>SD</i>	5.1	15.9	26.1	27.7	3.9	6.5	10.3	13.4
		<i>r</i>	1.47	1.81	1.76	1.72	1.46	1.41	1.20	1.23
EG	LN	<i>M</i>	0.0	0.9	1.1	1.4	1.2	0.7	0.4	0.4
		<i>SD</i>	-14.1	-4.5	1.8	5.6	-10.0	-6.2	-2.1	-0.3
		<i>r</i>	1.30	1.42	1.49	1.46	1.27	1.10	1.15	1.14
ER	EG	<i>M</i>	-0.8	-0.4	-0.5	-0.5	-0.1	0.2	0.1	0.0
		<i>SD</i>	-23.3	-17.8	-17.0	-16.2	-21.2	-15.6	-12.8	-11.0
		<i>r</i>	1.04	1.27	1.14	1.06	1.09	1.08	1.04	1.06
ER	ER	<i>M</i>	0.0	0.5	0.5	0.2	0.4	-0.1	-0.2	-0.1
		<i>SD</i>	-12.5	-5.7	-3.1	-1.6	-13.0	-8.7	-5.8	-2.8
		<i>r</i>	1.14	1.32	1.34	1.28	1.17	1.12	1.13	1.13
ER	LN	<i>M</i>	-0.7	-0.2	-0.2	-0.6	-0.2	-0.1	-0.2	-0.3
		<i>SD</i>	-26.2	-21.0	-19.7	-18.9	-24.9	-20.2	-17.4	-15.9
		<i>r</i>	0.96	1.19	1.17	1.03	1.01	1.00	1.02	0.95
LN	EG	<i>M</i>	-0.4	-0.1	-0.7	-0.6	1.1	0.3	0.2	-0.2
		<i>SD</i>	-16.1	-13.8	-16.0	-15.3	-3.8	-7.2	-7.8	-9.4
		<i>r</i>	1.08	1.14	1.05	1.14	1.25	1.11	1.13	1.14
LN	ER	<i>M</i>	2.7	5.0	5.4	4.5	2.5	2.4	1.4	1.3
		<i>SD</i>	51.0	78.6	82.4	67.9	43.5	43.4	33.9	34.3
		<i>r</i>	1.83	2.48	2.84	2.51	1.71	1.58	1.45	1.45
LN	LN	<i>M</i>	0.6	1.1	1.3	1.0	1.0	1.1	0.4	0.3
		<i>SD</i>	4.3	5.3	7.0	5.1	0.5	6.7	1.1	2.0
		<i>r</i>	1.49	1.59	1.57	1.36	1.29	1.25	1.15	1.22

Note. *r* = measure of stability of means; EG = ex-Gaussian; ER = Erlangian; LN = lognormal.

observations outside this interval and the possibility of spurious observations. In this case the likelihood function is as follows (cf. Kendall & Stuart, 1967, p. 523):

$$L(t_1, \dots, t_{N_A} | \Theta) = \frac{\prod_{i=1}^{N_A} f(t_i | \Theta)}{[F(U | \Theta) - F(L | \Theta)]^{N_A}} \quad (79)$$

Unfortunately, with many samples (especially small ones) the parameter values that maximize this likelihood function do so by driving the denominator toward zero, thereby leading to unreasonable estimates. Ratcliff (1993) also examined the possibility of using this likelihood function to correct for truncation effects, and he found that it worked reasonably well with samples of 1,000 RTs uncontaminated by spurious observations.

The third modified maximum likelihood approach we considered attempted to allow for spurious observations outside the truncation interval. In this case, the likelihood function is

$$L(t_1, \dots, t_{N_A}, N_L, N_U | \Theta) = \{g_f + (1 - g_s - g_f) \cdot F(L | \Theta)\}^{N_L} \times (1 - g_f - g_s)^{N_A} \times \left\{ \prod_{i=1}^{N_A} f(t_i | \Theta) \right\} \times \{g_s + (1 - g_f - g_s) \cdot [1 - F(U | \Theta)]\}^{N_U} \quad (80)$$

where *g_f* and *g_s* are the probabilities of fast and slow spurious

observations falling outside the truncation interval. Note that this likelihood function reduces to Equation 77 when *g_f* = *g_s* = 0, so Equation 77 can be viewed as an approximation of it when *g_f* and *g_s* are close to zero. This likelihood function was more difficult to maximize than Equation 77 because there are two additional parameters (*g_f* and *g_s*). More important, it was less successful in reducing truncation bias.

Recommendations

Two kinds of recommendations emerge from the systematic analyses undertaken here. The first type, general recommendations about handling outliers in general and about using truncation in particular, are presented in this section. The second set, a set of specific computational procedures for use with truncated data sets, are presented in Appendix A. The latter lays out reasonable heuristics for counteracting truncation biases in a variety of common experimental settings.

Our investigations lead us to recommend that researchers pay much more attention to the consequences of their outlier rejection methods. In particular, either RT distributions should not be truncated at all or they should be truncated with such extreme cutoffs that at most a tiny fraction (<.5%) of valid RTs are excluded. There are two main reasons for this recommendation.

Table 16

Average Relative Biases of Means and Standard Deviations Recovered From Truncated Samples Using Maximum Likelihood Estimation, and Measure of Stability of Means for Mixture Distributions With 0% Spuriously Fast Observations

Fitted distribution	Data distribution	Measure	Percentage of slow spurious RTs							
			Truncate 32%				Truncate 10%			
			0	2.5	5	10	0	2.5	5	10
EG	EG	<i>M</i>	0.17	0.93	1.50	3.10	0.35	1.30	2.75	5.40
		<i>SD</i>	-2.60	1.80	5.55	15.08	-2.75	4.22	13.78	30.40
		<i>r</i>	1.31	0.97	0.83	0.73	1.12	0.80	0.77	0.71
EG	ER	<i>M</i>	-1.33	-0.68	-0.05	1.20	-0.47	0.68	1.70	3.83
		<i>SD</i>	-22.50	-19.73	-16.03	-10.65	-16.58	-10.68	-5.05	5.28
		<i>r</i>	1.00	0.68	0.61	0.53	0.97	0.70	0.64	0.59
EG	LN	<i>M</i>	-1.13	-0.80	0.03	1.38	-0.45	0.83	1.83	4.65
		<i>SD</i>	-19.08	-16.77	-12.60	-5.45	-11.88	-4.03	4.07	23.63
		<i>r</i>	1.08	0.77	0.66	0.60	1.02	0.78	0.73	0.72
ER	EG	<i>M</i>	0.90	1.95	2.80	4.75	0.00	1.20	2.53	5.60
		<i>SD</i>	12.75	19.05	22.78	35.13	4.45	12.98	21.02	39.53
		<i>r</i>	1.57	1.18	0.99	0.87	1.15	0.89	0.81	0.79
ER	ER	<i>M</i>	-0.53	-0.33	0.68	2.20	-0.60	0.65	1.58	3.93
		<i>SD</i>	-9.10	-8.43	-3.63	3.10	-9.63	-3.35	1.13	11.93
		<i>r</i>	1.19	0.85	0.70	0.64	1.04	0.79	0.69	0.63
ER	LN	<i>M</i>	3.35	4.72	5.63	8.10	0.73	2.58	4.35	8.15
		<i>SD</i>	60.33	73.50	80.68	107.47	27.42	46.28	61.38	100.58
		<i>r</i>	2.15	1.59	1.35	1.27	1.37	1.12	1.01	1.03
LN	EG	<i>M</i>	0.00	0.95	1.60	3.25	-0.05	1.25	2.33	5.28
		<i>SD</i>	-6.93	-1.50	3.00	11.65	-8.08	0.88	7.33	24.68
		<i>r</i>	1.24	0.95	0.85	0.71	1.03	0.82	0.74	0.69
LN	ER	<i>M</i>	-1.35	-0.70	-0.20	1.20	-0.53	0.57	1.53	3.63
		<i>SD</i>	-25.45	-22.38	-19.88	-13.80	-20.77	-15.30	-10.35	-0.65
		<i>r</i>	0.95	0.73	0.60	0.52	0.90	0.70	0.62	0.59
LN	LN	<i>M</i>	0.38	0.95	1.95	3.73	0.08	1.28	2.98	5.85
		<i>SD</i>	0.58	5.45	12.03	24.80	-1.18	8.48	21.43	46.35
		<i>r</i>	1.29	0.99	0.87	0.81	1.10	0.91	0.81	0.83

Note. *r* = measure of stability of means; EG = ex-Gaussian; ER = Erlangian, LN = lognormal; RTs = reaction times.

First, even truncating a small percentage of the valid RTs can have a major impact on summary measures of RT distributions, so it is clearly not safe to truncate extreme observations and then ignore the fact of truncation. For example, truncating RTs changes linear mean RT functions to nonlinear ones, and vice versa. The very extremity that causes observations to be excluded also causes them to have a disproportionately large influence on the results, and so model predictions change when these observations are excluded. Obviously, one cannot take much comfort from the fact that only a small percentage of observations are excluded, because a small fraction of extreme observations can have a substantial impact on summary statistics.

Second, we have seen that the effects of spurious observations may not be so bad after all. For example, when the goal of an experiment is to check the linearity of the relationship between mean RT and some independent variable or to check the additivity of the effects on mean RT of two experimental factors, such RTs have no influence on the qualitative predictions of various models in the first place (i.e., they destroy neither linearity nor factor additivity). As truncation can change qualitative predictions, substantially so for linearity, it seems clear that the cure is worse than the disease, at least for these designs. In most experiments, the worst effect of spurious observations is the reduction in power that they produce, and even this effect can generally

be overcome by increasing the number of subjects or observations per subject.

Of course, there may be research settings in which our recommendation is impractical, because of the sheer number of clearly spurious RTs (e.g., in research with clinical populations). In these cases, truncation or some other form of outlier exclusion may be needed to improve the signal to noise ratio of the data being analyzed, especially if it is difficult to get large samples. Even when truncation is unavoidable, though, the results of the present investigation have some important implications.

For example, given the potential influence of outlier exclusion on RT results, one obvious recommendation is that researchers should describe in detail the procedure they use to exclude outliers from their samples. This should be regarded as an aspect of statistical procedure that is just as relevant to the results of the study as the visual angle and intensity of stimuli, the characteristics of the response manipulanda, and so forth.

Another obvious recommendation is that the percentage of excluded trials should clearly be reported, separately for each of the major conditions of the study. Future workers will be in a much better position to evaluate a given set of experimental results if the possible effects of outlier exclusion can be evaluated, and this is most easily accomplished if the percentage of excluded trials is known. Moreover, if these

Table 17

Average Relative Biases of Means and Standard Deviations Recovered From Truncated Samples Using Maximum Likelihood Estimation, and Measure of Stability of Means for Mixture Distributions With 10% Spuriously Fast Observations

Fitted distribution	Data distribution	Measure	Percentage of slow spurious RTs							
			Truncate 32%				Truncate 10%			
			0	2.5	5	10	0	2.5	5	10
EG	EG	<i>M</i>	-5.15	-4.65	-4.08	-2.83	-4.08	-3.95	-1.83	1.20
		<i>SD</i>	0.40	12.13	9.45	14.85	10.93	24.75	30.00	51.48
		<i>r</i>	0.91	0.75	0.66	0.57	1.00	0.84	0.77	0.72
EG	ER	<i>M</i>	-3.88	-3.30	-2.80	-1.23	-4.13	-3.05	-1.90	0.30
		<i>SD</i>	12.95	24.83	18.03	27.50	13.90	21.08	24.90	39.80
		<i>r</i>	0.92	0.77	0.63	0.56	0.88	0.73	0.66	0.62
EG	LN	<i>M</i>	-2.58	-1.70	-1.25	0.48	-3.55	-2.35	-0.68	2.30
		<i>SD</i>	27.88	42.08	36.48	48.78	21.23	35.83	42.25	68.25
		<i>r</i>	1.23	0.99	0.84	0.81	1.00	0.84	0.81	0.83
ER	EG	<i>M</i>	-3.28	-2.45	-1.73	0.18	-3.40	-2.28	-0.82	2.03
		<i>SD</i>	8.58	14.40	19.63	30.68	6.60	16.63	26.28	45.58
		<i>r</i>	1.20	0.98	0.89	0.81	0.95	0.82	0.77	0.75
ER	ER	<i>M</i>	-3.80	-3.08	-2.33	-0.78	-3.60	-2.50	-1.35	1.50
		<i>SD</i>	-0.25	3.97	8.60	16.25	1.58	9.23	15.53	32.33
		<i>r</i>	0.97	0.79	0.71	0.64	0.89	0.74	0.67	0.66
ER	LN	<i>M</i>	-0.85	0.13	0.90	3.45	-2.38	-0.93	0.68	4.05
		<i>SD</i>	43.13	52.45	57.58	80.40	25.38	39.85	52.88	83.75
		<i>r</i>	1.55	1.20	1.18	1.13	1.12	0.95	0.90	0.93
LN	EG	<i>M</i>	-4.35	-3.65	-3.13	-1.43	-3.35	-2.28	-1.10	2.05
		<i>SD</i>	-10.68	-6.80	-1.53	6.85	-1.68	6.00	14.63	35.98
		<i>r</i>	0.95	0.79	0.69	0.65	0.89	0.76	0.66	0.69
LN	ER	<i>M</i>	-4.00	-3.50	-2.93	-1.35	-3.60	-2.63	-1.43	0.93
		<i>SD</i>	-10.35	-8.32	-5.22	2.53	-6.65	-1.07	7.25	21.40
		<i>r</i>	0.80	0.64	0.56	0.51	0.80	0.67	0.62	0.58
LN	LN	<i>M</i>	-2.40	-1.75	-0.70	1.00	-2.68	-1.33	-0.08	2.93
		<i>SD</i>	10.28	15.73	23.08	36.50	7.88	20.68	29.52	58.70
		<i>r</i>	1.19	0.94	0.92	0.84	0.94	0.82	0.78	0.82

Note. *r* = measure of stability of means; EG = ex-Gaussian; ER = Erlangian; LN = lognormal; RTs = reaction times.

percentages are specified, the present computations (e.g., Figure 2) can be used to estimate the maximum truncation biasing effects that might have been present in the experiment, as demonstrated in Appendix A.

Beyond these general recommendations, the results of these investigations provide a lot of information that RT researchers can use to avoid being misled by biasing effects of truncation when analyzing truncated data sets. One option is to use the maximum likelihood procedure described in the section *Compensating for Truncation Effects* to estimate the desired parameters of the untruncated distribution. Specifically, the special Erlangian can be fit to each set of observed RTs, and the estimated means and medians can be computed from the best-fitting special Erlangian.

Another option, which is computationally much less difficult, is to make some conservative adjustments of the truncated means or medians to allow for the biasing effects of truncation. Appendix A describes in detail how to do this for a variety of common experimental designs.

The bottom line is that the head-in-the-sand approach to truncation presents too many dangers to remain standard in the field. If extreme observations have been excluded, researchers simply must consider the possible effects of such exclusion on their hypothesis tests. Furthermore, theorists should be especially motivated to develop predictions and

hypothesis tests for summary measures that are relatively uninfluenced by truncation (e.g., median), to minimize the possibility that biasing effects of truncation will change the outcome of model tests.

Although this article has considered in detail only one outlier exclusion method, it seems safe to conclude that the general issues raised here will tend to apply to any possible RT-based method of outlier exclusion. The basic problem, noted in connection with truncation, is that when the RT is itself the basis for exclusion, there is no way to avoid the exclusion of valid but extreme observations. Exclusion of these observations will necessarily distort estimates of population parameters (e.g., $E[T]$ and $Var[T]$), and the nature of the distortion may vary slightly across conditions, thereby influencing hypothesis tests.

For example, one common alternative exclusion method is based on sample standard deviations. Trials on which the RT is longer than a fixed number of standard deviations from the sample mean are excluded. Unfortunately, this procedure does not take into account the skewness of the RT distribution, so more RTs will be excluded from the upper than the lower tail of RT distributions, thereby introducing bias into the estimated mean and skewness, as well as clearly reducing the standard deviation. This presents an especially serious problem in certain types of experiments, because the result-

ing estimation bias depends on sample size (Miller, 1991).

Another method for accommodating extreme RTs uses the so-called trimmed mean, which is computed from a random sample $T_1 < T_2 \dots < T_N$ of size N as follows: Choose a suitable positive integer k , $k < N/2$, such that the k smallest observations T_1, \dots, T_k and the k largest observations $T_N, T_{N-1}, \dots, T_{N-k+1}$ are deleted from the sample. The average of the $N - 2k$ intermediate observations $T_{k+1}, T_{k+2}, \dots, T_{N-k}$ is called the k th level trimmed mean

$$M_k = \frac{\sum_{i=k+1}^{N-k} T_i}{N - 2k}. \quad (81)$$

The trimmed mean excludes extreme observations from both tails of the RT distribution. Again, however, this method will also induce bias in the estimation of the moments of the untrimmed distribution, because RT distributions are skewed.

It is beyond the scope of this article to review all existing outlier identification and exclusion methods (cf. Barnett & Lewis, 1984; Hawkins, 1980; Lovie, 1986), but it is likely that each of them will have some effects on the analyses. We feel it is much more reasonable to understand and attempt to correct for exclusion biases than to search for a magic exclusion procedure that will be immune to such distortions, and in this spirit we conducted the present examination of truncation effects.

References

- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W. (1972). *Robust estimates of location: Survey and advances*. Princeton, NJ: Princeton University Press.
- Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, 2, 123-147.
- Ashby, F. G. (1982). Testing the assumptions of exponential, additive reaction time models. *Memory & Cognition*, 10, 125-134.
- Baddeley, A. D. (1990). *Human memory: Theory and practice*. Hillsdale, NJ: Erlbaum.
- Bain, L. J. (1978). *Statistical analysis of reliability and life-testing models*. New York: Marcel Dekker.
- Barlow, R. E., & Proschan, F. (1965). *Mathematical theory of reliability*. New York: Wiley.
- Barnett, V., & Lewis, T. (1978). *Outliers in statistical data*. New York: Wiley.
- Barnett, V., & Lewis, T. (1984). *Outliers in statistical data* (2nd ed.). New York: Wiley.
- Beckman, R. J., & Cook, R. D. (1983). Outlier.....s. *Technometrics*, 25, 119-149.
- Bree, D. S. (1975). The distribution of problem-solving times: An examination of the stages model. *British Journal of Mathematical and Statistical Psychology*, 28, 177-200.
- Briggs, G. E. (1974). On the predictor variable for choice reaction time. *Memory & Cognition*, 2, 575-580.
- Bundesen, C., Larsen, A., & Farrell, J. E. (1981). Mental transformations of size and orientation. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance*, Vol. 9 (pp. 279-294). Hillsdale, NJ: Erlbaum.
- Burbeck, S. L., & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics*, 32, 117-133.
- Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Memory & Cognition*, 41, 393-401.
- Crow, E. L., & Shimizu, K. (1988). *Lognormal distributions: Theory and applications*. New York: Marcel Dekker.
- Dixon, P., & Just, M. A. (1986). A chronometric analysis of strategy preparation in choice reactions. *Memory & Cognition*, 14, 488-500.
- Egeth, H. E., & Dagenbach, D. (1991). Parallel versus serial processing in visual search: Further evidence from subadditive effects of visual quality. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 551-560.
- Elandt-Johnson, R. C., & Johnson, N. L. (1980). *Survival distributions: Reliability applications in the biomedical sciences*. New York: Wiley.
- Everitt, B. S., & Hand, B. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.
- Hawkins, D. M. (1980). *Identification of outliers*. London: Chapman & Hall.
- Heathcote, A., & Mewhort, D. (1993). Representation and selection of relative position. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 488-516.
- Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response-time distributions: An example using the Stroop task. *Psychological Bulletin*, 109, 340-347.
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1985). *Exploring data tables, trends, and shapes*. New York: Wiley.
- Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 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.
- Huber, P. J. (1981). *Robust statistics*. New York: Wiley.
- Jensen, A. R. (1987). Individual differences in the Hick paradigm. In P. A. Vernon (Ed.), *Speed of information processing and intelligence* (pp. 101-175). Norwood, NJ: Ablex.
- Jones, W. P., & Anderson, J. R. (1982). Semantic categorization and high-speed scanning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 237-242.
- Kendall, M. G., & Stuart, A. (1967). *The advanced theory of statistics. Volume II: Inference and relationship* (2nd ed.). London: Griffin.
- Kotz, S. (1973). Normality vs. lognormality with applications. *Communications in Statistics*, 1, 113-132.
- Lambert, A. J., & Hockey, R. J. (1986). Selective attention and performance with a multidimensional visual display. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 484-495.
- Lovie, P. (1986). Identifying outliers. In A. D. Lovie (Ed.), *New developments in statistics for psychology and the social sciences* (pp. 44-69). New York: Methuen.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford, England: Oxford University Press.
- Massaro, D. W., & Cowan, N. (1993). Information processing models: Microscopes of the mind. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (pp. 383-425). Stanford, CA: Annual Reviews.
- McClelland, J. L. (1979). On the time relations of mental processes: A framework for analyzing processes in cascade. *Psychological Review*, 86, 287-330.
- McCormack, P., & Wright, N. (1964). The positive skew observed in reaction time. *Canadian Journal of Psychology*, 18, 43-51.

- McGill, W. (1963). Stochastic latency mechanisms. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. 1* (pp. 309–360). New York: Wiley.
- McGill, W. J., & Gibbon, J. (1965). The general-gamma distribution and reaction times. *Journal of Mathematical Psychology, 2*, 1–18.
- Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed–accuracy decomposition. *Psychological Review, 95*, 183–237.
- Meyer, D. E., Yantis, S., Osman, A. M., & Smith, J. E. K. (1985). Temporal properties of human information processing: Tests of discrete versus continuous models. *Cognitive Psychology, 17*, 445–518.
- Miller, J. O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*, 247–279.
- Miller, J. O. (1983). Can response preparation begin before stimulus recognition finishes? *Journal of Experimental Psychology: Human Perception and Performance, 9*, 161–182.
- Miller, J. O. (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 539–543.
- Miller, J. O. (1991). Reaction time analysis with outlier exclusion: Bias varies with sample size. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology, 43*, 907–912.
- Miller, J. O. (1993). A queue-series model for reaction time, with discrete-stage and continuous-flow models as special cases. *Psychological Review, 100*, 702–715.
- Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.
- Murphy, T. D., & Eriksen, C. W. (1987). Temporal changes in the distribution of attention in the visual field in response to precues. *Perception & Psychophysics, 42*, 576–586.
- Parzen, E. (1962). *Stochastic processes*. San Francisco: Holden-Day.
- Pashler, H. E. (1984a). Evidence against late selection: Stimulus quality effects in previewed displays. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 429–448.
- Pashler, H. E. (1984b). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 358–377.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). *Numerical recipes: The art of scientific computing*. Cambridge, England: Cambridge University Press.
- 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. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.
- Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*, 190–214.
- Roberts, S., & Sternberg, S. (1992). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV. Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611–653). Cambridge, MA: MIT Press.
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York: Wiley.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.
- Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology, 18*, 105–139.
- Schweickert, R., & Townsend, J. T. (1989). A trichotomy: Interactions of factors prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *Journal of Mathematical Psychology, 33*, 328–347.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.
- Shimizu, K., & Crow, E. L. (1988). History, genesis, and properties. In E. L. Crow & K. Shimizu (Eds.), *Lognormal distributions: Theory and applications* (pp. 1–25). New York: Marcel Dekker.
- Sternberg, S. (1967). Two operations in character-recognition: Some evidence from reaction time measurements. *Perception & Psychophysics, 2*, 45–53.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 276–315.
- Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics. Volume 1: Distributional theory*. London: Charles Griffin.
- Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology, 21*, 136–152.
- Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology, 28*, 363–400.
- Townsend, J. T. (1992). On the proper scales for reaction time. In H. G. Geissler, S. W. Link, & J. T. Townsend (Eds.), *Cognition, information processing, and psychophysics: Basic issues* (pp. 105–120). Hillsdale, NJ: Erlbaum.
- Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge, England: Cambridge University Press.
- Townsend, J. T., & Schweickert, R. (1985). Interactive effects of factors affecting processes in stochastic PERT networks. In G. d'Ydewalle (Ed.), *Cognition, information processing, and motivation* (pp. 255–276). Amsterdam: North-Holland.
- Townsend, J. T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology, 33*, 309–327.
- Ulrich, R., & Miller, J. O. (in press). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*.
- Ulrich, R., & Wing, A. M. (1991). A recruitment theory of force-time relations in the production of brief force pulses: The parallel force unit model. *Psychological Review, 98*, 268–294.
- Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology*. New York: Henry Holt.
- Yantis, S., Meyer, D. E., & Smith, J. E. K. (1991). Analyses of multinomial mixture distributions: New tests for stochastic models of cognition and action. *Psychological Bulletin, 110*, 350–374.

(Appendixes follow on next page)

Appendix A

How to Estimate and Control for Truncation Biases: A Cookbook Approach

This appendix is offered as a practical guide for RT researchers who wish to allow for potential truncation effects in the analysis of a given set of truncated data and thereby avoid reaching conclusions that are really artifacts of the truncation procedure. In particular, it explains (a) how to estimate potential biases introduced by truncation and (b) how to control potential truncation effects in inferential tests using reaction time (RT).

The procedures described here are applicable to RT experiments including multiple conditions with the same truncation cutoffs. Three cases are distinguished, depending on the outcome of the experiment. If mean RT varies considerably over conditions but the percentages of RTs discarded from the lower and upper tails of the RT distributions remain constant, then it is very likely that the truncation procedure eliminated only spurious RTs (i.e., the tails of the true RT distributions did not extend into the rejection region). In this case, it does not seem necessary to worry about truncation artifacts.

On the other hand, if the percentages of truncated RTs vary with mean RT across conditions, then it is very likely that the true RT distributions extend past the truncation cutoffs, and the possibility of truncation artifacts must be considered. For example, if the percentages of RTs truncated from the lower tails of the distributions decrease as mean RT increases, this would indicate that extremely fast yet valid RTs were truncated from the true RT distributions. Analogously, if the percentages of RTs truncated from the upper tails increase with mean RT, then this would indicate that valid but long RTs were discarded from the true distributions. Therefore, there are three cases in which truncation artifacts are a realistic possibility, which can be empirically distinguished as follows:

Case 1. As mean RT increases, the percentage P of RTs truncated from the lower tail decreases, and the percentage Q truncated from the upper tail increases.

Case 2. The percentage P of RTs discarded from the lower tail decreases as mean RT increases, but the percentage Q of RTs discarded from the upper tail remains constant.

Case 3. The percentage P of RTs truncated from the lower tail remains constant as mean RT increases, but the percentage Q of RTs truncated from the upper tail increases.

Assessment of Truncation Effects in a Single Condition

In this section we describe a procedure to assess truncation effects on the mean, standard deviation, or median of RT in a single condition. This procedure is then adapted in the following sections for use in the analysis of multicondition experiments. The computational steps of the procedure establish an interval estimate for the desired parameter (i.e., mean, standard deviation, or median). The procedure is easy to implement on a computer, and the width of the obtained interval estimate can be narrowed with additional distributional information and computations.

To some extent, this procedure can be regarded as a simpler alternative to the maximum likelihood approach described in the text section *Compensating for Truncation Effects*. However, the maximum likelihood approach requires a complete data set, whereas this procedure can also be applied to aggregated data sets. This feature might be especially useful in examining published results for potential truncation artifacts. As the procedure requires less computational effort than the maximum likelihood approach, it

might also be routinely applied in RT research. If the procedure suggests that truncation artifacts may have been large enough to explain the experimental effects, a researcher should use the maximum likelihood approach to compensate for truncation artifacts.

Rationale for the Procedure

The idea behind the procedure is fairly simple. If percentages p and q of valid RTs are truncated from the fast and slow tails, respectively, then the problem is to estimate the relative bias $\delta(p, q)$ for the probability density function (PDF) f . In most practical situations, however, the researcher will not know the underlying RT distribution from which the sample was taken; not only its parameters but also its shape (e.g., Erlangian, lognormal, or ex-Gaussian) are therefore unknown. In the presence of this uncertainty, a conservative strategy is to allow for the worst case by determining the most extreme bias values in Table 2. Accordingly, two extreme bias values would be obtained, namely $\delta_{\min} = \min[\hat{\delta}_i(p, q)]$ and $\delta_{\max} = \max[\hat{\delta}_i(p, q)]$, where i indexes the alternative PDFs considered in Table 2. These two extreme values define the endpoints of a range of plausible bias values, encompassing several plausible RT distributions. Given the range of plausible bias values $[\delta_{\min}, \delta_{\max}]$, it is a simple matter to obtain an interval estimate of the untruncated mean with Equation 24.

The Procedure

We first describe and illustrate how the procedure can be used to estimate the mean RT, and then we indicate how the steps are modified to estimate the median or standard deviation. Let M be the truncated mean and r_0 be the irreducible minimum of RT. A reasonably conservative estimate of r_0 is 100 ms (see Luce, 1986, p. 62). Then an interval estimate of the untruncated mean is obtained with the following steps:

1. If Case 1 applies, then set $p = P$ and $q = Q$
If Case 2 applies, then set $p = P$ and $q = 0$
If Case 3 applies, then set $p = 0$ and $q = Q$
2. Compute $M_w = M - r_0$.
3. Determine with Table 2 the smallest bias, that is,

$$\delta_{\min} = \min[\hat{\delta}_i(p, q)], \quad (\text{A1})$$

where i runs over all approximated bias functions that are represented in Table 2. Compute $M_u = M_w/(1 + \delta_{\min}/100) + r_0$.

4. Analogously, determine

$$\delta_{\max} = \max[\hat{\delta}_i(p, q)] \quad (\text{A2})$$

and calculate $M_l = M_w/(1 + \delta_{\max}/100) + r_0$.

5. The untruncated mean RT should be in the interval $[M_l, M_u]$.

The interval estimates obtained from this procedure are analogous to, but different from, confidence intervals. Like confidence intervals, they identify a range of values within which the desired parameter (i.e., mean, median, or standard deviation of the untruncated distribution) might lie. Unlike confidence intervals, however, these interval estimates do not allow for sampling error in the data, only for uncertainty about the effect of truncation, which stems from uncertainty about the true underlying distribution of valid RTs. If

a researcher wanted an interval that allowed for both sorts of uncertainty, it would be possible to extend the interval computed here by 2-3 standard errors in each direction, using the appropriate measure of standard error estimated from the data.

A Numerical Example

Table A1 provides a numerical example involving four hypothetical subjects. The percentages *P* and *Q* truncated from the lower and upper tails, respectively, of each subject's distribution varies between subjects. The resulting truncated mean, standard deviation, and median are listed for each subject.

The computational steps listed in the previous section are illustrated for the mean of 340 ms for Subject 1 under Case 1:

1. Set $p = 0.6$ and $q = 2.5$
2. Assume $r_0 = 100$ and compute $M_w = 340 - 100 = 240$.
3. Use $\hat{\delta}(p, q) = A \cdot p^B + C \cdot q^D$. The coefficients *A*, *B*, *C*, *D* are given in Table 2 for various distributions. For the Erlangian, we obtain the following bias values $\hat{\delta}(.6, 2.5)$ for $m = 2, \dots, 15$: -5.49, -4.16, -3.42, -2.94, -2.61, -2.36, -2.17, -2.01, -1.88, -1.77, -1.67, -1.59, -1.52, and -1.45. For the lognormal the bias values are -6.44, -4.88, -4.00, -3.43, -3.02, -2.72, -2.48, -2.29, -2.14, -2.00, -1.89, -1.79, -1.71, and -1.63. Finally, the bias values for the ex-Gaussian are -5.37, -4.09, -3.38, -2.93, -2.61, -2.36, -2.17, -2.02, -1.89, -1.78, -1.69, -1.60, -1.53, and -1.47. Taking all values into account, we find $\delta_{\min} = -6.44$ and $\delta_{\max} = -1.45$, and hence

$$M_u = 240 / (1 + (-6.44) / 100) + 100 = 357$$

and

$$M_l = 240 / (1 + (-1.45) / 100) + 100 = 344.$$

4. Thus, the obtained interval estimate of the untruncated mean is [344,357].

Analogous computational steps apply for Cases 2 and 3. Similar steps are needed to obtain estimates of the median and the standard deviation of RT. However, r_0 should always be set to zero in Step 2 when obtaining an interval for the standard deviation, because for standard deviations the tables are independent of r_0 .

Cases 1, 2, and 3 were applied to each hypothetical subject and each summary measure for comparison. Table A2 summarizes obtained intervals for the mean, standard deviation, and median.

To facilitate this procedure, we computed the most extreme bias values, δ_{\min} and δ_{\max} , obtained with the mean, median, and standard deviation for various values of *p* and *q*, and these are given in Tables A3 and A4. Two-dimensional interpolation (Press, Flannery, Teukolsky, & Vetterling, 1986, pp. 95-101) can be used to obtain quite accurate estimates of δ_{\min} and δ_{\max} for values of *p* and *q* between those in the tables. These tables greatly simplify the third step in the

Table A1
Data for Numerical Example Using Four Hypothetical Subjects

Subject	<i>M</i>	<i>Mdn.</i>	<i>SD</i>	<i>P</i> (%)	<i>Q</i> (%)
1	340	320	65	0.6	2.5
2	295	270	95	1.0	1.5
3	370	345	80	0.5	3.0
4	415	385	100	0.3	5.0
Average	355	330	85	0.6	3.0

Note. Shown for each subject are the truncated mean, median, and standard deviation of reaction time (RT), and the percentages of RTs discarded from the lower (*P*) and upper (*Q*) tails of each subject's RT distribution.

Table A2
Interval Estimates Obtained for the Hypothetical Subjects in Table A1

Subject	<i>M</i>	<i>Mdn.</i>	<i>SD</i>	<i>p</i>	<i>q</i>
Case 1: $p = P$ and $q = Q$					
1	[344,357]	[321,324]	[73,84]	0.6	2.5
2	[296,303]	[270,271]	[104,116]	1.0	1.5
3	[375,392]	[347,351]	[90,107]	0.5	3.0
4	[425,454]	[389,398]	[118,147]	0.3	5.0
Average	[360,377]	[332,336]	[96,114]		
*	[360,375]	[332,335]	[96,114]	0.6	3.0
Case 2: $p = P$ and $q = 0$					
1	[338,339]	[319,320]	[65,66]	0.6	0.0
2	[292,294]	[268,269]	[95,97]	1.0	0.0
3	[368,369]	[344,345]	[80,81]	0.5	0.0
4	[413,414]	[384,385]	[100,101]	0.3	0.0
Average	[353,354]	[329,330]	[85,86]		
*	[353,354]	[329,330]	[85,86]	0.6	0.0
Case 3: $p = 0$ and $q = Q$					
1	[343,358]	[322,325]	[72,84]	0.0	2.5
2	[297,305]	[271,272]	[102,116]	0.0	1.5
3	[376,393]	[347,352]	[89,107]	0.0	3.0
4	[425,455]	[389,399]	[117,147]	0.0	5.0
Average	[362,378]	[332,337]	[95,114]		
*	[361,377]	[332,337]	[95,113]	0.0	0.0

Note. The last row in each section, indicated with an asterisk, shows the intervals obtained when the procedure is applied to the group average data of Table A1. Intervals were computed assuming $r_0 = 100$ ms and $2 \leq m \leq 15$. *p* = proportion truncated from lower tail; *q* = proportion truncated from upper tail.

above procedure, as the values of δ_{\min} and δ_{\max} may be obtained from these tables, obviating the need to compute $\hat{\delta}(p, q)$ for each of the 42 distributions under consideration.

Extension to Other Summary Measures

The procedure just described can easily be extended to the estimation of standard deviation and median, using Tables 3 and 4 instead of Table 2. The only cautionary note concerns the irreducible minimum of RT: Relative biases of medians were also computed assuming $r_0 = 0$, so Equation 24 should also be used in that case, replacing means with medians. Relative biases of standard deviations were computed without this assumption, however, so an equation analogous to Equation 23 must be used with that summary measure (i.e., set $r_0 = 0$).

Bias Correction in Group Data

The procedure described here might also be useful in evaluating potential truncation artifacts in previously published group data, where data of single subjects are not available, as illustrated in Tables A1 and A2. The last row in Table A1 shows the averages of mean, standard deviation, median, *P*, and *Q* over the four hypothetical subjects. Cases 1-3 were applied to these group means, and the resulting intervals for the mean, standard deviation, and the median of the group data are shown in Table A2 (see rows denoted with an asterisk). These results agree very well with the averages of the individual subject intervals given in Table A2, suggesting that approximately the same results are obtained whether one computes

Table A3

Most Negative Relative Bias (δ_{\min}) of Mean, Standard Deviation, and Median, as a Function of $p\%$ and $q\%$

p (%)	Parameter	q (%)							
		0.00	0.25	0.50	1.00	2.00	4.00	6.00	10.00
0.00	M	0.0	-1.4	-2.3	-3.7	-6.0	-9.7	-12.8	-18.2
	SD	0.0	-7.5	-10.5	-14.7	-20.6	-28.8	-35.1	-45.0
	Mn	0.0	-0.2	-0.5	-1.0	-1.9	-3.8	-5.7	-9.4
0.25	M	0.2	-1.2	-2.1	-3.5	-5.8	-9.4	-12.6	-18.0
	SD	-0.9	-7.5	-10.5	-14.7	-20.6	-28.9	-35.1	-45.0
	Mn	0.1	0.0	-0.2	-0.7	-1.7	-3.5	-5.4	-9.2
0.50	M	0.3	-0.9	-1.8	-3.2	-5.5	-9.2	-12.3	-17.7
	SD	-1.4	-7.6	-10.6	-14.8	-20.6	-28.9	-35.2	-45.1
	Mn	0.2	0.1	0.0	-0.5	-1.4	-3.3	-5.2	-8.9
1.00	M	0.5	-0.5	-1.4	-2.8	-5.1	-8.8	-11.9	-17.3
	SD	-2.4	-7.6	-10.6	-14.8	-20.7	-29.0	-35.3	-45.1
	Mn	0.3	0.2	0.2	0.0	-1.0	-2.8	-4.7	-8.5
2.0	M	1.0	0.3	-0.6	-2.0	-4.3	-8.0	-11.1	-16.5
	SD	-4.0	-7.8	-10.8	-15.0	-20.9	-29.2	-35.4	-45.3
	Mn	0.6	0.5	0.5	0.3	0.0	-1.9	-3.8	-7.5
4.0	M	1.8	1.5	1.0	-0.4	-2.7	-6.4	-9.5	-14.9
	SD	-6.6	-9.2	-11.2	-15.5	-21.3	-29.6	-35.9	-45.8
	Mn	1.2	1.2	1.1	0.9	0.6	0.0	-1.8	-5.6
6.0	M	2.6	2.3	2.1	1.0	-1.2	-4.9	-8.0	-13.5
	SD	-8.9	-11.5	-12.8	-15.9	-21.8	-30.0	-36.3	-46.2
	Mn	1.9	1.8	1.7	1.6	1.2	0.6	0.0	-3.6
10.0	M	4.1	3.8	3.6	3.2	1.6	-2.1	-5.2	-10.6
	SD	-13.0	-15.5	-16.8	-18.8	-22.8	-31.0	-37.3	-47.2
	Mn	3.1	3.0	2.9	2.8	2.5	1.9	1.3	0.0

Note. δ_{\min} was computed across 42 distributions defined by the Erlangian, lognormal, or ex-Gaussian with parameters corresponding to $2 \leq m \leq 15$. p = proportion truncated from lower tail; q = proportion truncated from upper tail.

an interval from the averages or computes averages from the intervals.^{A1}

Constraining the Search Through the Bias Tables

The above procedure assumes that the researcher has no information about the true underlying distribution of valid RTs and must therefore allow for the worst case across any of the distributions contained in the bias tables. If some information about the true underlying distribution is available, however, the range of distributions can be constrained, and it is very likely that this will produce a narrower range of $[\delta_{\min}, \delta_{\max}]$, producing smaller interval estimates for the untruncated values. As an extreme example, if the underlying distribution is known to be a special Erlangian with $m = 2$, then the untruncated value can be estimated precisely (i.e., the interval $[M_b, M_u]$ contains a single point.)

At issue, then, is the question of which theoretical RT distributions provide acceptable models for an observed distribution of RTs. A useful technique for addressing this question is the quantile-quantile plot, or Q-Q plot (see Hoaglin, Mosteller, & Tukey, 1985, Chapter 10, for technical details of this technique). This is a plot of selected quantiles of the observed RT distribution against the same quantiles calculated from a given theoretical distribution (e.g., the special Erlangian with $m = 2$). The resulting plot will be a straight line when the two distributions have the same shape, so the goodness of fit may be measured with a product-moment correlation coefficient. In fact, linearity will be obtained even if the distributions differ in scale and location, so the technique does not require the estimation of such parameters. If the shapes differ, of course, the plot will curve in some way. To reduce the conservatism of the present procedure, then, the researcher could use this technique to assess many different underlying distributions and reject those yielding significant curvature, thereby reducing the number of dis-

tributions across which minimum and maximum bias are obtained. In addition, if the researcher is willing to assume that all subjects have the same underlying distributional shape, the Q-Q plots can be computed using vincentized RT distributions (Thomas & Ross, 1980).

^{A1} The mathematical relation between these two procedures shows why they yield almost identical intervals. Let the subjects in a sample be numbered $1, \dots, N$ and let M_1, \dots, M_N denote the truncated means of the subjects, respectively. Furthermore, let δ_j be the relative bias of mean M_j , and define $B_j \equiv 1/(1 + \delta_j/100)$. Then the bias-corrected mean M_{cj} of subject j is given by

$$M_{cj} = M_j \cdot B_j.$$

Using the definition of covariance $s_{x,y}$ of two variables x and y

$$s_{x,y} = \frac{\sum_{j=1}^N x_j y_j}{N} - \bar{x} \bar{y},$$

and the bias corrected sample mean \bar{M}_c is computed as

$$\begin{aligned} \bar{M}_c &= \frac{1}{N} \sum_{j=1}^N M_j \cdot B_j \\ &= \bar{M} \cdot \bar{B} + s_{M,B} \\ &\approx \bar{M} \cdot \bar{B}. \end{aligned}$$

This approximation holds, if $s_{M,B}$ is close to zero or if it is negligibly small with respect to the product $\bar{M} \cdot \bar{B}$. We found this to be a reasonable approximation in a variety of realistic simulations.

Furthermore, let $\bar{p} = (1/N) \sum_{j=1}^N p_j$ and $\bar{q} = (1/N) \sum_{j=1}^N q_j$. Taylor series expansion of \bar{B} (cf. Mood, Graybill, & Boes, 1974, p. 181)

Table A4

Most Positive Relative Bias (δ_{max}) of Mean, Standard Deviation, and Median, as a Function of $p\%$ and $q\%$

p (%)	Parameter	q (%)							
		0.00	0.25	0.50	1.00	2.00	4.00	6.00	10.00
0.00	M	0.0	-0.3	-0.5	-0.9	-1.5	-2.7	-3.7	-5.6
	SD	0.0	-2.1	-3.3	-5.1	-8.0	-12.6	-16.4	-22.8
	Mn	0.0	-0.1	-0.2	-0.3	-0.6	-1.2	-1.9	-3.1
0.25	M	0.4	-0.1	-0.3	-0.7	-1.4	-2.5	-3.5	-5.4
	SD	-0.0	-2.7	-3.9	-5.7	-8.6	-13.2	-17.0	-23.4
	Mn	0.2	-0.0	-0.1	-0.2	-0.5	-1.2	-1.8	-3.0
0.50	M	0.8	0.0	-0.2	-0.6	-1.2	-2.4	-3.4	-5.2
	SD	-0.1	-3.1	-4.3	-6.2	-9.1	-13.6	-17.4	-23.9
	Mn	0.5	0.2	-0.0	-0.2	-0.5	-1.1	-1.7	-2.9
1.00	M	1.4	0.4	0.1	-0.3	-0.9	-2.1	-3.1	-4.9
	SD	-0.1	-3.8	-5.1	-6.9	-9.8	-14.4	-18.2	-24.6
	Mn	0.9	0.7	0.5	0.0	-0.3	-0.9	-1.5	-2.8
2.0	M	2.5	1.4	0.8	0.2	-0.4	-1.6	-2.6	-4.4
	SD	-0.3	-4.6	-6.2	-8.3	-11.2	-15.7	-19.5	-26.0
	Mn	1.9	1.7	1.4	0.9	0.0	-0.6	-1.2	-2.5
4.0	M	4.5	3.4	2.6	1.6	0.5	-0.6	-1.6	-3.5
	SD	-0.8	-5.4	-7.4	-10.1	-13.5	-18.0	-21.8	-28.3
	Mn	3.8	3.6	3.4	2.9	1.9	0.0	-0.6	-1.8
6.0	M	6.4	5.3	4.5	3.2	1.7	0.2	-0.8	-2.6
	SD	-1.2	-6.2	-8.2	-11.2	-15.2	-20.1	-23.9	-30.3
	Mn	5.8	5.5	5.3	4.8	3.9	2.0	0.1	-1.2
10.0	M	10.0	8.8	8.1	6.8	4.7	2.2	0.8	-1.0
	SD	-2.2	-7.8	-9.8	-12.8	-17.3	-23.4	-27.6	-34.0
	Mn	9.7	9.4	9.2	8.7	7.8	5.9	4.0	0.3

Note. δ_{max} was computed across 42 distributions defined by the Erlangian, lognormal, or ex-Gaussian with parameters corresponding to $2 \leq m \leq 15$. p = proportion truncated from lower tail; q = proportion truncated from upper tail.

Some Conservative Correction Procedures for Hypothesis Testing Situations

In many experiments the results will be so clear-cut that they cannot be explained as truncation artifacts, and it is useful to have some simple methods capable of establishing that this is the case without too much work. Such methods are described in this section, for a variety of different experimental situations and comparisons. These methods are all conservative, in the sense that they allow at least enough for the effects of truncation. Thus, if these methods show that truncation artifacts cannot explain the results, the researcher can proceed to interpret the results confidently, without any worry that truncation artifacts are responsible. On the other hand, if these methods indicate that truncation artifacts may have been large enough to explain the results, then the researcher should use

shows that

$$\bar{B} \approx \left[1 + \frac{A\bar{p}^B + C\bar{q}^D}{100} \right]^{-1}$$

Simulations revealed that this approximation is also excellent. Finally, if we substitute the previous expression into $\bar{M}_c \approx \bar{M} \cdot \bar{B}$, then we arrive at the useful approximation

$$\bar{M}_c \approx \frac{\bar{M}}{1 + \frac{A\bar{p}^B + C\bar{q}^D}{100}}$$

for group data, which was used to correct the group mean in the numerical example of Table 19. An analogous argument holds for computation of intervals for the median and standard deviation from aggregated data sets.

the more complicated but less conservative maximum likelihood method to evaluate the *truncation artifact hypothesis* further.

The basic strategy is the same in all these conservative correction procedures. Using the observed proportions of RTs truncated from below and above, a researcher decides which of the three possible cases discussed above applies to the data and computes estimates of the true parameter values for untruncated distributions. These estimates may be made under various distributional assumptions. From each set of estimates (i.e., each type of distributional assumption), a new value of the hypothesis-testing statistic is computed. If all values are significant (i.e., regardless of distributional assumptions), then the researcher can confidently reject the null hypothesis, H_0 , even allowing for truncation effects. If none are significant, then the researcher can confidently conclude that truncation biases were not responsible for the failure to reject H_0 . If some values are significant and others are not, no firm conclusions can be drawn from these procedures and the less conservative but more difficult maximum likelihood method should be used.

Experiments Comparing Two Conditions, H_0 Rejected

Suppose an experimenter finds a statistically reliable difference between the means of two conditions. In the present context, the question is whether the difference might actually have been an artifact of the truncation procedure, and the answer clearly depends on the proportions of observations truncated from above and below in each condition.

A relatively conservative way to address this question is to ask how small the larger untruncated mean might reasonably be, and to ask how large the smaller untruncated mean might reasonably be. We then repeat the statistical test with the two extreme means to find out whether the result remains statistically reliable. If so, we can safely reject the possibility that the difference in means is due to a

truncation artifact. Thus, the overall strategy is to find the minimal plausible difference between two untruncated means.

Table A5 gives a concrete numerical example. This hypothetical example displays mean RTs for a sample of 15 subjects, with each subject tested in two conditions. The averaged truncated means in Conditions 1 and 2 were 267 and 293 ms, respectively. A t test for correlated samples on the truncated mean RT revealed a significant effect $t_{df=14} = 5.7, p < .001$.

For each subject, hypothetical percentages P and Q of truncated RTs are also given. The values of P do not vary with condition, but Q s are larger in Condition 2 than in 1, indicating that the right tail of the distribution of valid RTs extends into the upper rejection region. Thus, Case 3 applies. For each subject and condition the above procedure was used to compute an interval estimate for the untruncated mean. Because the smaller truncated mean was observed in Condition 1, we selected the largest estimate of the untruncated mean, M_u , for each subject in this condition. Similarly, as the larger truncated mean was obtained under Condition 2, we selected the smallest estimate of the untruncated mean in this condition. Having allowed for truncation bias in this rather conservative fashion, the resulting group means are 274 and 296 ms under Conditions 1 and 2, respectively, and the t test still shows a significant effect $t_{df=14} = 4.8, p < .001$. Thus, we may safely conclude that the RT difference between Condition 1 and 2 is not an artifact of the truncation procedure. In experiments with two independent samples, similar reasoning would apply.

Experiments Comparing Two Conditions, H_0 Not Rejected

If H_0 is not rejected, of course, interest turns to the issue of whether or not truncation effects might have concealed a true difference between means. This is an inherently more difficult problem, of course, because of the well-known fact that failure to reject a null hypothesis does not imply the truth of that hypothesis. Nonetheless, when considering truncation effects, it makes sense to ask whether a statistically significant difference might have been concealed by truncation.

Table A5
Numerical Example of Testing for Significant Difference
in Mean Reaction Time Between Two Conditions

Subject	Condition 1				Condition 2			
	M	P	Q	M_u	M	P	Q	M_l
1	250	1.0	1.2	257	300	0.5	1.3	302
2	300	1.0	0.4	304	310	0.8	0.7	311
3	220	0.5	0.6	223	250	0.6	1.1	251
4	240	0.5	0.3	242	240	0.9	0.6	241
5	310	0.8	1.2	319	330	1.0	1.5	333
6	270	0.5	1.0	277	300	0.6	1.2	302
7	280	0.8	0.7	285	310	0.7	1.6	313
8	210	0.6	0.3	212	240	0.8	0.5	241
9	320	1.3	1.4	331	350	1.1	1.4	353
10	340	0.8	2.0	355	330	0.9	2.3	334
11	350	1.3	2.5	369	400	1.1	2.5	406
12	230	0.4	2.4	240	280	0.6	2.2	283
13	220	0.5	1.3	226	250	0.7	1.8	252
14	210	0.4	0.3	212	220	0.5	0.8	221
15	250	0.9	0.9	255	290	1.1	1.5	292
Average	267	0.8	1.1	274	293	0.8	1.4	296

Note. M_u is the maximum value in the interval estimate of the untruncated mean in Condition 1; M_l is the minimum value in the interval estimate of the Condition 2 mean. Both sets of intervals were computed assuming that $r_0 = 100$ ms and $2 \leq m \leq 15$. Data shown are for 15 hypothetical subjects.

A conservative approach to this question is to find the maximum difference in means consistent with the observed truncation proportions. If this maximum difference would still have been statistically insignificant, then we can be reasonably sure that truncation effects were not responsible for the lack of a significant difference.

In this case, the experimenter must examine two complementary comparisons: (a) Compare the maximum estimates (i.e., M_u) in Condition 1 against the minimum estimates (i.e., M_l) in Condition 2 and (b) compare the minimum estimates (i.e., M_l) in Condition 1 against the maximum estimates (i.e., M_u) in Condition 2. If H_0 cannot be rejected in either comparison, then there is no reason to suspect that truncation may have concealed a significant effect. If H_0 can be rejected in at least one comparison, however, then it is possible that the null hypothesis would have been rejected were it not for the truncation procedure.

Experiments Comparing More Than Two Conditions

When more than two conditions are compared in an experiment, an analysis of variance provides an omnibus F statistic used to test the null hypothesis that all conditions have the same true mean. If this F is significant, we must ask whether the significance might be due to truncation effects; if it is not significant, we must ask whether the failure to get a significant F might be due to truncation.

A straightforward, although tedious, extension of the procedure used in the previous two sections may be applied in this case. Taking into account whether Case 1, 2, or 3 applies, we can compute the extreme estimates M_u and M_l for each condition. For each possible set of these extreme estimates, we can compute an F_* . Across the different sets, we can find the maximum and minimum values of this F_* . If the minimum F_* is statistically significant, or if the maximum F_* is statistically nonsignificant, then the conclusion will be obvious, given the conservatism of the procedure. Unfortunately, if the minimum F_* is statistically nonsignificant and the maximum F_* is statistically significant, then it will be necessary to use the maximum likelihood method to provide a more precise test of the null hypothesis.

In experiments comparing more than two conditions, it is frequently useful to make post hoc comparisons involving pairs of conditions or involving contrasts between the averages of some conditions and the averages of others. Fortunately, the conservative approach described above for two-sample comparisons extends easily to these cases. When comparing a pair of conditions, one would simply find the minimum or maximum difference between untruncated means (depending on whether the hypothesis was rejected or not in the truncated comparison), using the methods described earlier. When comparing averages, one would similarly find the minimum or maximum difference in average of untruncated means, replacing each condition mean with its minimum or maximum untruncated estimate as appropriate.

Experiments Examining Linear RT Functions

The linear component across an observed set of conditions is simply a special type of contrast across those conditions, so the methods of the previous section apply.

Factorial Experiments

The deviation from additivity across four conditions in a 2×2 factorial design is also a special type of contrast, so again the earlier contrast methods apply. With more levels on one or both of the factors, the contrast will have to be made separately for each factorial combination of four levels. Alternatively, one could compute F_s s under various assumptions as was done before, and examine the maximum and minimum F_s s.

Comparisons of Parameters Other Than Means

The same basic strategy can be used for tests involving other parameters (e.g., variance). Using the observed proportions of truncated observations, a new value of the test statistic can be computed

for each set of assumptions about the form of the underlying distribution. If the minimum value of the test statistic is significant, or the maximum is nonsignificant, then firm conclusions can be reached. Otherwise, an alternative procedure (e.g., maximum likelihood) must be used to reach a decision.

Appendix B

Proofs of Equation 6 and Propositions 1 and 2

Proof of Equation 6

Let $0 < p < 1$. A quantile of order p with respect to the truncated distribution is a value t_p ($L < t_p < U$) such that the conditional probability $P[\mathbf{T} \leq t_p | L \leq \mathbf{T} \leq U] = p$ holds. In particular, the median is given by the value $t_{.5}$.

The quantile t_p of the truncated cumulative probability density function (CDF) is easily expressed in terms of the untruncated CDF by noting the following relations:

$$p = P[\mathbf{T} \leq t_p | L \leq \mathbf{T} \leq U] \tag{B1}$$

$$= \frac{\int_L^{t_p} f(t) dt}{F(U) - F(L)} \tag{B2}$$

$$= \frac{F(t_p) - F(L)}{F(U) - F(L)} \tag{B3}$$

Rearranging yields,

$$F(t_p) = p \cdot [F(U) - F(L)] + F(L) \tag{B4}$$

$$= p \cdot F(U) + (1 - p) \cdot F(L). \tag{B5}$$

Hence, in general, the quantile t_p is given by

$$t_p = F^{-1}[p \cdot F(U) + (1 - p) \cdot F(L)], \tag{B6}$$

which yields Equation 6 for $p = .5$.

Proof of Proposition 1

$$E[\mathbf{T}^i | L \leq \mathbf{T} \leq U] \tag{B7}$$

$$= \frac{\int_L^U t^i f(t) dt}{F(U) - F(L)}$$

$$= \frac{\int_L^U t^i \frac{\lambda e^{-\lambda t} (\lambda t)^{m-1}}{(m-1)!} dt}{F(U) - F(L)} \tag{B8}$$

$$= \frac{\int_L^U \frac{\lambda e^{-\lambda t} \lambda^{m+i-1} t^{m+i-1}}{(m+i-1)!} dt}{F(U) - F(L)} \cdot \frac{(m+i-1)!}{(m-1)!} \lambda^{-i} \tag{B9}$$

Note that the integrand of the above integral is simply the PDF of a special Erlangian with shape parameter $(m + i)$ and scale parameter λ , so the integral is $G(U) - G(L)$, with G defined in the proposition. The proof is complete.

Proof of Proposition 2

$$E[\mathbf{T}^i | L \leq \mathbf{T} \leq U] = \frac{\int_L^U \frac{t^i}{\sqrt{2\pi\sigma t}} \exp\left[-\frac{(\ln t - \mu)^2}{2\sigma^2}\right] dt}{\Phi\left[\frac{\ln U - \mu}{\sigma}\right] - \Phi\left[\frac{\ln L - \mu}{\sigma}\right]} \tag{B10}$$

Because $\mathbf{T}^i = e^{i \ln \mathbf{T}}$, we may rewrite the above equation as

$$E[e^{i \ln \mathbf{T}} | L \leq \mathbf{T} \leq U] = \frac{\int_L^U \frac{e^{i \ln t}}{\sqrt{2\pi\sigma t}} \exp\left[-\frac{(\ln t - \mu)^2}{2\sigma^2}\right] dt}{\Phi\left[\frac{\ln U - \mu}{\sigma}\right] - \Phi\left[\frac{\ln L - \mu}{\sigma}\right]} \tag{B11}$$

Define the numerator of Equation B11 as function $g(i)$ and substitute y for $\ln t$. This yields

$$g(i) = \int_{\ln L}^{\ln U} \frac{e^{iy}}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(y - \mu)^2}{2\sigma^2}\right] dy. \tag{B12}$$

Substituting $\sigma z + i\sigma^2 + \mu$ for y gives

$$g(i) = \int_a^b \frac{e^{i(\sigma z + i\sigma^2 + \mu)}}{\sqrt{2\pi}} \exp\left[-\frac{(z\sigma + i\sigma^2)^2}{2\sigma^2}\right] dz, \tag{B13}$$

where the lower and upper limits of integration are $a = (\ln L - i\sigma^2 - \mu)/\sigma$ and $b = (\ln U - i\sigma^2 - \mu)/\sigma$, respectively. Simplifying the above expression yields

$$g(i) = \exp\left[i\mu + \frac{i^2\sigma^2}{2}\right] \times \int_a^b \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \tag{B14}$$

$$= \exp\left[i\mu + \frac{i^2\sigma^2}{2}\right] \times [\Phi(b) - \Phi(a)]. \tag{B15}$$

Replacing the numerator of Equation B11 with the right-hand side of Equation B15 completes the proof.

Appendix C

Models Predicting Asymptotic Hazard Functions

A asymptotic value of the hazard function is predicted by models in which RT is a sum of components, at least one of which is exponential. For example, Luce (1986, pp. 103–105) noted that the hazard function of a convolution approaches an asymptotic value of λ whenever one of the convolved variables is bounded and the other is exponential (λ) in its tail. Ashby (1982, Theorem 3) showed that an asymptotic hazard function also results from the convolution of an exponential with a sum of other components having nondecreasing hazard functions. These suggestions can be extended, because the hazard function of T is asymptotic under even more general conditions. To see this, let $T = E + R$ with E having an exponential distribution with rate λ and R being a nonexponential component. Because $f_T(t) = \int_0^t f_E(t-x)f_R(x)dx$ and $f_E(t) = \lambda e^{-\lambda t}$, it is easily seen that

$$f_T(t) = \lambda e^{-\lambda t} \int_0^t e^{\lambda x} f_R(x) dx, \quad (C1)$$

and

$$F_T(t) = F_R(t) - e^{-\lambda t} \int_0^t e^{\lambda x} f_R(x) dx, \quad (C2)$$

and hence the hazard function of T is

$$h(t) = \frac{\lambda e^{-\lambda t} \int_0^t e^{\lambda x} f_R(x) dx}{1 - F_R(t) + e^{-\lambda t} \int_0^t e^{\lambda x} f_R(x) dx}. \quad (C3)$$

Now let $G(\lambda, t) \equiv \int_0^t e^{\lambda x} f_R(x) dx$; then Equation 77 simplifies to

$$h(t) = \frac{\lambda}{\frac{1 - F_R(t)}{e^{-\lambda t} G(\lambda, t)} + 1}. \quad (C4)$$

$G(\lambda, \infty)$ is simply the moment generating function of R , which should be finite under plausible distributional assumptions. From Equation C4, it can be seen that $h(t)$ approaches λ whenever the tail probability $1 - F_R(t)$ decreases to zero faster than the tail probability $1 - F_E(t) = e^{-\lambda t}$. As the upper tail of an exponential distribution is relatively thick, this will hold for many assumed R distributions, including unbounded ones.

Received November 18, 1992

Revision received July 7, 1993

Accepted September 28, 1993 ■

Low Publication Prices for APA Members and Affiliates

Keeping You Up-to-Date: All APA members (Fellows; Members; Associates, and Student Affiliates) receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*.

High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they can subscribe to the *American Psychologist* at a significantly reduced rate.

In addition, all members and affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential Resources: APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the APA*, the *Master Lectures*, and *Journals in Psychology: A Resource Listing for Authors*.

Other Benefits of Membership: Membership in APA also provides eligibility for low-cost insurance plans covering life, income protection, office overhead, accident protection, health care, hospital indemnity, professional liability, research/academic professional liability, student/school liability, and student health.

For more information, write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242, USA