An evaluation of the Vincentizing method of forming group-level response time distributions

JEFFREY N. ROUDER and PAUL L. SPECKMAN University of Missouri, Columbia, Missouri

Vincentizing (quantile averaging) is a popular means of pooling response time distributions across individuals to produce a group average. The benefit of Vincentizing is that the resulting histogram "looks like" an average of the individuals. In this article, we competitively test Vincentizing against the more mundane approach of averaging parameter estimates from fits to individuals. We simulate data from three distributions: the ex-Gaussian, the Weibull, and the shifted-Wald. For the ex-Gaussian and the shifted-Wald, parameter averaging outperforms Vincentizing. There is only an advantage of Vincentizing for the Weibull and only when there are few observations per participant. Overall, we recommend that researchers use Vincentizing only in select circumstances and with the knowledge that Vincentized estimates are often inconsistent estimators of averaged parameters.

Response time (RT), the time taken to complete a task, is a common dependent variable in experimental psychology and has been used to draw inferences about the nature of mental processing (Luce, 1986). Most researchers tend to analyze the mean RT, but a growing number of researchers are examining whole RT distributions as a means of providing more extensive and insightful tests of cognitive and perceptual theories (e.g., Ashby, Tien, & Balakrishnan, 1993; Hockley, 1984; Logan, 1992; Ratcliff, 1978; Ratcliff & Rouder, 1998, 2000; Rouder, 2000; Spieler, Balota, & Faust, 1996). For many paradigms and research questions, it is advantageous to use several individuals and attempt estimating a group response time distribution that reflects the distribution of an average or typical individual. To this end, it is imperative to have a set of tools for forming group RT distributions.

The current state of the art in forming group RT distributions is a nonparametric technique called Vincentizing¹ (Vincent, 1912). Figure 1 shows an example of Vincentizing. Quantiles² of individual RT distributions are computed and then averaged to produce a group distribution. Panel A shows three individual distributions of four observations each. For each individual, the smallest observation is the .2 quantile, the next is the .4 quantile, and so on. The .2 quantile of the Vincentized distribution is the average of the .2 quantile of each participant, and so on. To produce a group RT distribution, this averaging procedure is applied to several quantiles. Ratcliff (1979) provided a landmark study in which he claimed that the

This research was supported by National Science Foundation Grant SES-0095919 to J.N.R., D. Sun, and P.L.S., University of Missouri Research Board Grant 00-77 to J. N.R., and National Science Foundation Grant DMS-9972598 to D. Sun and P.L.S. We are grateful to Michael Stadler, who allowed us use of his data. Correspondence concerning this article may be addressed to J. N. Rouder, Department of Psycho-

logical Sciences, 210 McAlester Hall, University of Missouri, Colum-

bia, MO 65211 (e-mail: jeff@banta.psyc.missouri.edu).

mean of the Vincentized distribution was about the average of the individuals' means, the variance was about the average of individuals' variances, and the shape was about the average of individuals' shapes. Panel B demonstrates these properties. The first three histograms show data from three hypothetical individuals (1,000 observations each). The fourth histogram is from the Vincentized distribution, and it appears to have average mean, variability, and shape.

Because of these properties, Vincentizing has become a useful tool in cognitive and perceptual psychology. Some authors have used it to provide a pictorial view of group RT distributions (e.g., Heathcote, Popiel, & Mewhort, 1991; Madden et al., 1999; Penner-Wilger, Leth-Steensen, & LeFevre, 2002). Others have fit parametric forms to the Vincentized distribution and have used the resulting parameter values to draw conclusions about latent psychological processes. For example, Logan (1992) tested the instance theory of practice with Vincentizing. He took blocks of six RTs per item and averaged quantiles across participants. Then he estimated Weibull-distribution parameters from the Vincentized distributions. The question of interest was whether the Weibull parameters varied systematically with practice. Another example comes from Andrews and Heathcote (2001). They used Vincentizing to generate RT distributions across a number of word identification tasks, including word naming, lexical decision, and the task of deciding whether a target word was a member of a previously presented semantic category. They fit an ex-Gaussian (Hohle, 1965) to Vincentized distributions and assessed whether word frequency effects were manifested in the same ex-Gaussian parameter across tasks. A third example comes from Spieler et al. (1996), who looked at which ex-Gaussian components of RT were affected by Stroop interference across young adults, healthy older adults, and older adults with Alzheimer's-type dementia. A fourth example comes from Ratcliff and Rouder (2000), who Vincentized RT

(A)	Pa	articipa	nts	Quantile	Vincentized
	1	2	3	Probability	Distribution
	315	459	688	.2	487
	372	525	755	.4	551
	584	534	825	.6	648
	684	574	915	.8	724

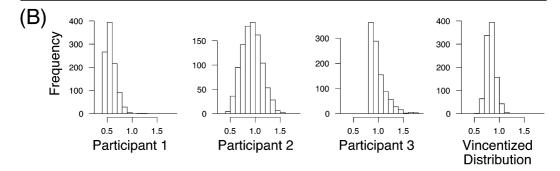


Figure 1. Examples of Vincentizing, Panel A shows how quantiles are averaged. Panel B shows that Vincentized distribution is an "average" of the individual distributions.

distributions in a letter identification task. These Vincentized distributions were then fit with a diffusion model (Ratcliff, 1979; Ratcliff & Rouder, 1998), and the resulting parameters were used to draw conclusions about the dynamics of letter recognition.

To evaluate Vincentizing as a means of obtaining group RT estimates, we start with a definition of a true or population group-level distribution. There is a fairly natural definition of true group distribution within parametric frameworks. Suppose that each individual's true distribution follows a certain parametric form with unique parameters. The true group RT distribution is that certain parametric form with parameters given by an average of the individual parameters. For example, the Weibull distribution has three parameters: one corresponding to location, another to scale, and a third to shape. Let us suppose that each individual's true distribution is a Weibull but that they all have unique location, scale, and shape parameters. By this definition, the true group distribution is a Weibull. The true group location parameter is the average of each individual's true location parameter, and so on. Here, we study how well Vincentizing allows for the recovery of these true group parameters.

Parametric analysis affords an alternative to Vincentizing in constructing group RT distributions. Individual-level parameters may be estimated from individual RT distributions and then averaged to form group-level estimates. For the Weibull example, estimated group location parameter is the mean of all the individual location estimates, and so on. This approach is termed *parameter averaging*. In this article, we competitively test Vincentizing against two parameter-averaging methods to see which best recovers group-level parameters.

There is reason at the outset to suspect that Vincentizing may be inferior to parameter averaging. Estimators in psychology are usually consistent. A consistent estimator will converge to its true value with increasing sample size; an inconsistent estimator will not. Consistency is desirable because estimation can be made arbitrarily accurate by choosing a sufficiently large sample. Almost all common estimators are consistent, including sample mean, sample variance, and sample correlation. Typical approaches to estimating individual RT distributions, such as maximizing likelihood and minimizing chi-square, are also consistent, at least under mild technical conditions (Hogg & Craig, 1978). Averaging individual parameter estimates is certainly a consistent estimator of the underlying population average if the individual parameter estimates are consistent. But estimates of true group parameters from Vincentized distributions are generally not consistent. Thomas and Ross (1980) showed that for Vincentizing to yield consistent parameter estimates, the individual distributions must be of a specific type: They must be from a common *location-scale family*. Examples of such distributions include the normal and exponential but do not include distributions in which the shape changes. We discuss the sampling properties of Vincentizing for location-scale families elsewhere (Jiang, Rouder, & Speckman, 2004). Here, we evaluate Vincentizing vis-à-vis parameter averaging for three non-location-scale distributions: the ex-Gaussian, the Weibull, and the shifted-Wald distributions. For all three of these distributions, group parameter estimates from Vincentized distributions are necessarily inconsistent.

Inconsistency, in and of itself, is not fatal. An estimator could still well approximate the parameter it is intended to estimate. But it must be realized that with a sufficiently

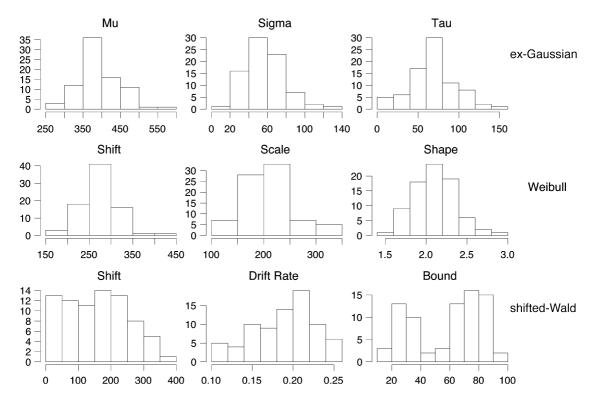


Figure 2. Histograms of *truth values* used in the simulations for the three distributions. All three parameters of ex-Gaussian, the shift and scale parameter of the Weibull, and the shift parameter of the shifted-Wald are measured in milliseconds. The shape of the Weibull and the bound of the shifted-Wald are dimensionless quantities. Drift rate is measured in units of 1/msec.

large sample, a consistent estimator will always outperform an inconsistent estimator. Therefore, we already have a partial answer to our question. We know that, for those cases in which Vincentizing produces inconsistent estimators, there necessarily exists a sample size above which parameter averaging is the superior method.

SIMULATION STUDY

To assess the relative performance of Vincentizing and parameter averaging, we performed a set of simulation studies. The relative performance of the methods depends on how much variability there is in RT across individuals. To ensure a realistic degree of across-participants variability³ and, hence, increase the generalizability of the results, we based our simulations on a large data set consisting of 80 observations for each of 80 individuals. We fit each individual's RT distribution with three parametric forms. The resulting parameter values served as *truth values* for the subsequent simulations. These truth values were used to generate artificial data for each individual. Group-level parameters were then estimated from the artificial data by either Vincentizing or parameter averaging.

Parametric Distributions

We used three distributions to test Vincentizing and parameter averaging. Our choice was not meant to be exhaustive but was representative of the range of options in the modeling of RT.

Ex-Gaussian. The ex-Gaussian is a popular form that describes the addition of an exponentially distributed and a normally distributed stage. There are three parameters, one that describes the scale of the exponential (denoted τ) and two that describe the mean and the scale of the normal (denoted μ and σ , respectively). The density of the ex-Gaussian is given by

$$f(t) = \frac{\exp\left(-\frac{(t-\mu)}{\tau} + \frac{\sigma^2}{2\tau^2}\right)}{\tau} \Phi\left(\frac{t-\mu}{\sigma} - \frac{\sigma}{\tau}\right), \quad \sigma, \tau > 0,$$

where Φ is the cumulative distribution function of the standard normal distribution. Typically, the exponential distribution represents the duration of decision processes that are thought to be central and consciously controlled (Balota & Spieler, 1999). The normal distribution represents the duration of sensory processes that are thought to be peripheral and automatic. Histograms of the indi-

vidual truth values for the ex-Gaussian used in the simulations are shown in Figure 2. These values were obtained by maximum likelihood estimation.

Weibull. The Weibull is a three-parameter distribution with parameters of shift, scale, and shape. The density of a Weibull is given by

$$f(x) = \left(\frac{\beta}{\theta}\right) \left(\frac{t - \psi}{\theta}\right)^{\beta - 1} \exp\left[-\left(\frac{t - \psi}{\theta}\right)^{\beta}\right],$$

where (ψ, θ, β) are the shift, scale, and shape parameters, respectively, $\tau > \psi$, and $\theta, \beta > 0$. Logan (1988, 1992) noted that the Weibull describes race processes and used it to model performance gains from practice. Alternatively, Rouder, Sun, Speckman, Lu, and Zhou (2003) use the preceding distinction between peripheral and central processes in their interpretation of Weibull parameters. Changes in shape across two conditions or populations are assumed to reflect structural changes in central processing, such as insertion of stages (e.g., Ashby & Townsend, 1980; Balota & Chumbley, 1985), Changes in scale reflect change in the rate or speed of central processing but no changes in structure or strategy. Finally, changes in location reflect changes in sensory processes. Histograms of the truth values for the Weibull used in the simulations are shown in Figure 2. These were obtained by Rouder et al. through a hierarchical Bayesian approach. The values are similar to maximum likelihood estimates.

Shifted-Wald. The shifted-Wald is a three-parameter form that describes a diffusion process for RT tasks with a single response—for example, simple detection or the go/no-go task. The parameters⁵ are the drift rate of the process, the boundary, and a residual shift parameter. This residual shift parameter denotes the time taken for nondecisional processes, such as encoding the stimulus and motor processes responsible for making the response (this parameter is termed $T_{\rm ER}$ in Ratcliff's diffusion model). Recently, Schwarz (2001) has introduced an ex-Wald distribution—a model of the addition of a Wald distributed stage with an exponentially distributed stage. Here, we use the shifted-Wald instead of the ex-Wald, since the shifted-Wald is more similar to two-choice diffusion models. The density of the shifted-Wald is given by

$$f(t) = \frac{\alpha}{\sigma \sqrt{2\pi(t-\psi)^3}} \exp\left\{-\frac{\left[\alpha - \mu(t-\psi)\right]^2}{2(t-\psi)}\right\},\,$$

where (ψ, α, μ) are the shift, bound, and drift rate parameters, respectively, $\tau > \psi$, and $\alpha, \mu > 0$. Histograms of the truth values for the shifted-Wald used in the simulations are shown in Figure 2. These were obtained by the nonlinear least-squares method described below. For 3 of the 80 participants, the value of the shift was below zero. Negative shift values are psychologically implausible, and the simulation study was based on the remaining 77 participants.

Three Different Estimation Methods

Group-level parameters were estimated by three different methods.

Vincentizing + least squares (V+LS). First, sample quantiles are computed and averaged—that is, Vincentized. Then, the estimated parameters are those that minimized the sum square difference between these averaged quantiles and the predicted quantiles. This least-squares method is asymptotically optimal for averaged quantiles for location-scale random variables, such as the normal and the exponential (Jiang et al., 2004), and serves as a suitable method here.

Maximum likelihood + parameter averaging (ML+PA). Maximum likelihood is an accepted and recommended method in both the statistics literature (e.g., Lehmann, 1991) and the psychological literature (e.g., Dolan, van der Maas, & Molenaar, 2002; Van Zandt, 2000). Individual-level maximum likelihood estimates are averaged (across individuals) to produce group-level estimates.

Least squares + parameter averaging (LS+PA). This is a quantile-based approach similar to that in Heath-cote, Brown, and Mewhort (2002). Individual parameter estimates are those that minimize the sum square difference between individual's quantiles and predicted quantiles. Individual estimates are averaged to produce group-level estimates. This approach is comparable to V+LS, except that the order of averaging and estimating is reversed.

Method

Figure 3 provides a schematic for the simulations with the Weibull distribution. The simulations for the other two distributions were run in an identical manner. There is a separate individualized true value for the shift, scale, and shape parameters (denoted with "T-Shift" in Box 1 in the figure). Using the truth values, we defined the true group parameters as the arithmetic average across the true individual parameters—for example, the true group shift parameter is the average of individual true shift parameters (see Box 2 in the figure). Data were simulated from individual truth values, as is indicated in Box 3. Parameters were estimated from the data set with either parameter averaging (ML+PA, LS+PA; see Boxes 4 and 5) or Vincentizing (V+LS; see Boxes 6 and 7). The figure shows the case with 80 observations per participant. We also ran a second set of simulations with 20 observations per participant to explore the small sample behavior of the estimation methods.

Results

Ex-Gaussian. The estimation errors for the three parameters are displayed as boxplots in Figure 4. The top row is for the first data set with 80 observations per individual. The bottom row is for the second data set with 20 observations per individual. The results show that Vincentizing (V+LS) is biased and the least effective method of estimation. The parameter-averaging techniques are fairly accurate. The two parameter-averaging methods

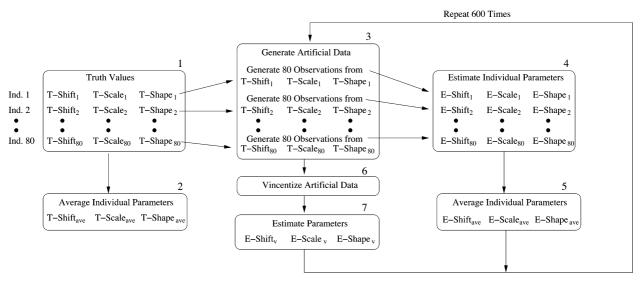


Figure 3. Schematic of the simulation method.

are fairly equivalent for 80 observations per individual, whereas LS+PA outperforms ML+PA with 20 observations per individual. Overall, researchers should use parameter averaging (perhaps LS+PA) when fitting the exGaussian.

Why does LS+PA outperform the Vincentizing method? As was previously mentioned, the Vincentizing method necessarily produces inconsistent estimators. For all sample sizes, the Vincentized estimators will be biased. For this distribution, LS+PA approaches the unbiased asymptotic limit quickly, even with sample sizes as small as 20 observations per individual. We speculate

that the increased performance of LS+PA over ML+PA at 20 observations per individual reflects the nonasymptotic properties of the estimators for this distribution and does not necessarily generalize to other distributions. Although we are surprised at the result, it is consistent with the results of Heathcote et al. (2002; see also Speckman & Rouder, 2004, and Heathcote & Brown, 2004, in this issue). They reported an advantage of a related quantile-based estimation method that they termed *quantile maximum likelihood estimation*.

Weibull. Figure 5 shows the estimation errors for the Weibull distribution as boxplots. With 80 observations

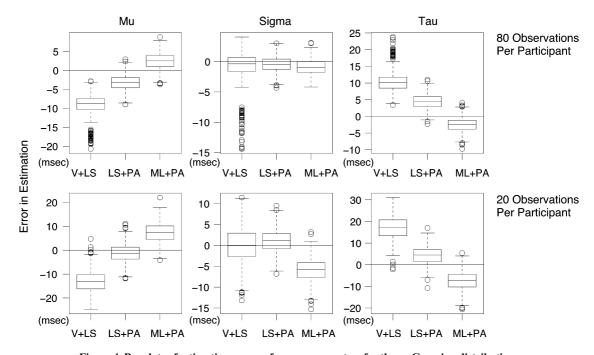


Figure 4. Boxplots of estimation errors of group parameters for the ex-Gaussian distribution.

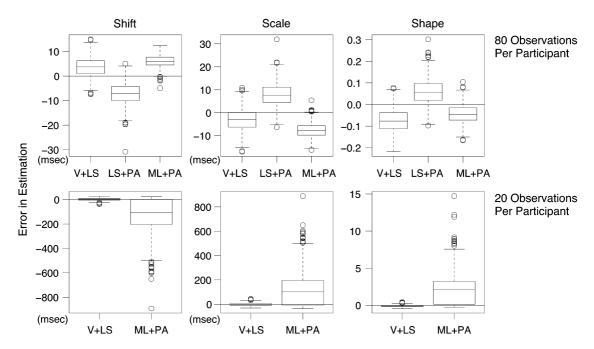


Figure 5. Boxplots of estimation errors of group parameters for the Weibull distribution.

per individual, all three techniques provided estimates with modest degrees of bias and fairly comparable standard errors. The situation changed drastically when there were only 20 observations per individual. LS+PA estimates were wildly biased and variable; for example, the shift parameter had a mean error on the order of -1,000 sec and a standard error on the order of 10,000 sec. It is impossible to fit the boxplots of the group parameter estimates on the same graph with the other two methods. Overall, the two parameter-averaging methods fail spectacularly, whereas the Vincentizing method does fairly well. Importantly, these failures do not reflect failures of the numerical optimization routine to find maxima or

minima of the objective functions. In almost all cases, outputs of the numerical routines converged to their preset criteria. Instead, the failures reflect an underlying inability of the parameter-averaging methods to provide robust estimates with small sample sizes.

The dramatic failure of parameter-averaging methods comes about because the Weibull is an "ill-behaved" distribution. The main problem is that for high values of shape parameter, corresponding increases in the parameter do not substantially affect the distribution. Figure 6 shows this phenomenon. Each density has the same mean and variance, but the shape parameter is varied from 7 to 7,000. The consequence for estimation is that small

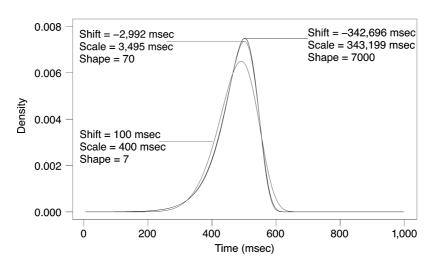


Figure 6. The Weibull distribution is "ill-behaved" in that distributions with vastly different parameters mimic each other.

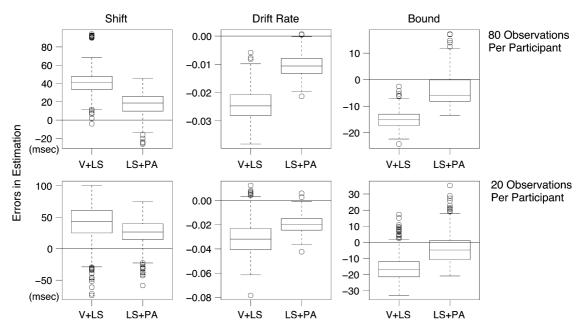
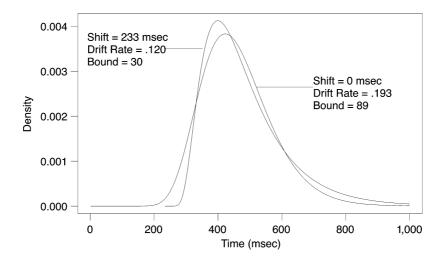


Figure 7. Boxplots of estimation errors of group parameters for the shifted-Wald distribution.

changes in the data result in huge changes in parameters that span several orders of magnitude. Fortunately, this parameter tradeoff is present only for Weibull distributions with high shape values (those that skew left). If the Weibull has shape parameter values in the ranges typical for RT distributions (1.3–3.4), the tradeoff is not present.

The instability demonstrated in Figure 6 raises problems when Weibull parameters are estimated from small samples. Due to random variability, the sample distribution for small samples may be symmetric or even skewed left. In this case, the Weibull estimates may vary by several orders of magnitude. This is exactly what happened in the simulation with 20 observations per individual. Vincentizing is not susceptible to the same instabilities as parameter averaging. The Vincentized distribution maintains a moderate degree of right skew and is not affected disproportionately by extreme samples. Weibull parameters most concordant with the Vincentized samples are not in the range in which the instability occurs. The upshot is a massive advantage in Vincentizing for small sample sizes.

With the results from the two different sample sizes, it is easy to make recommendations. For sample sizes smaller than 80 samples per individual, Vincentizing is superior to parameter averaging. For sample sizes around 80, the methods are fairly equivalent. But because parameter-averaging methods are consistent and Vincentizing is not, the parameter-averaging methods will be more ac-



Figure~8.~The~shifted-Wald~distribution~is~``ill-behaved"~in~that~different~parameter~settings~can~yield~similar~distributions.

curate with larger sample sizes. Our results indicate that the crossover point is around 80 observations per individual. Of course, these results hold for this data set; but on the grounds that this is a representative set for quick accurate decisions, it can be generalized to data from similar tasks.

Shifted-Wald. Figure 7 shows the estimation errors for the shifted-Wald as boxplots for the LS+PA and V+LS methods. ML+PA estimation errors are not graphed, since they were even larger and more variable than those displayed. For example, the ML+PA estimate for shift has an average error of -118 msec and a standard deviation of 130 msec. Although LS+PA outperforms Vincentizing in terms of variability and bias, both methods yield surprisingly biased and variable estimates. Figure 8 shows two shifted-Wald densities that are fairly similar. Even though the densities are similar, the corresponding parameters are quite different. Small changes in the distribution from sampling variability result in large changes in the parameters. Although this ill-behavior is not as extreme as that with the Weibull, it is somewhat more endemic and occurs for right-skewed distributions. Because of this ill-behavior, parameter estimation within the shifted-Wald is difficult, and researchers need hundreds of observations per participant to sufficiently localize parameters. Nevertheless, the simulations reveal that averaging individual least-squares parameters is the best method for estimating group parameters.

CONCLUSION

The preceding analyses demonstrate that Vincentizing should not be the de facto means of constructing group RT distributions. Group estimates from Vincentized quantiles are often less accurate than those obtained from averaging individual parameter estimates. For non-location-scale families, such as the ex-Gaussian, Weibull, and shifted-Wald distributions, Vincentized estimates of group parameters are not consistent—they do not converge to true values. This inconsistency means that for some sample size, parameter averaging will outperform estimates from Vincentized distributions. For the ex-Gaussian and shifted-Wald, we found through simulation that estimates from parameter averaging are less error prone than those from Vincentizing (at least for sample sizes larger than 20 observations per participant). We advise researchers using the ex-Gaussian or shifted-Wald to consider parameter averaging. However, there are some cases in which Vincentizing is more accurate. In the Weibull distribution, instability of individual parameter estimates with small sample sizes leads to horrible estimation properties in both individual and group parameter estimates. These instabilities do not affect Vincentized estimates. Of course, these results are obtained from simulated data, and it is unknown how robust the estimation methods are to misspecification.

We have two concluding recommendations. First, we recommend that researchers precisely define what they mean by a true group RT distribution. We explore the para-

metric approach to modeling RT, in which it is fairly natural to define true group RT distributions as those with parameters given by averages of true individual parameters. Only within the context of a precisely defined true group RT distribution can one estimate group parameters or perform inference. Second, we recommend that researchers explore, through simulation, both parameter averaging and Vincentizing to find out which is best for their particular application. Perhaps the strongest result from this set of studies is that although Vincentizing may work well when distributions are ill-behaved, it is less than optimal for other applications.

REFERENCES

- Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27, 514-544.
- ASHBY, F. G., TIEN, J.-Y., & BALAKRISHNAN, J. D. (1993). Response time distributions in memory scanning. *Journal of Mathematical Psychology*, **37**, 526-555.
- ASHBY, F. G., & TOWNSEND, J. T. (1980). Decomposing the reaction time distribution: Pure insertion and selective influences revisited. *Journal of Mathematical Psychology*, **21**, 93-123.
- BALOTA, D. A., & CHUMBLEY, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? Journal of Memory & Language, 12, 590-606.
- BALOTA, D. A., & SPIELER, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, 128, 32-55.
- Dolan, C. V., van der Maas, H. L. J., & Molenaar, P. C. M. (2002).
 A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavior Research Methods, Instruments, & Computers*, 34, 304-323.
- Feller, W. (1968). Introduction to probability theory and its applications (Vol. 1, 3rd ed.). New York: Wiley.
- HEATHCOTE, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments*, & *Computers*, **28**, 427-445.
- HEATHCOTE, A., & Brown, S. (2004). Reply to Speckman and Rouder: A theoretical basis for QML. *Psychonomic Bulletin & Review*, 11, 577-578.
- HEATHCOTE, A., BROWN, S., & MEWHORT, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, **9**, 394-401.
- 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.
- HOCKLEY, W. E. (1984). Analysis of reaction time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10, 598-615.
- Hogg, R. V., & Craig, A. T. (1978). Introduction to mathematical statistics. New York: Macmillan.
- HOHLE, R. H. (1965). Inferred components of reaction time as a function of foreperiod duration. <u>Journal of Experimental Psychology</u>, 69, 382-386.
- HYNDMAN, R. J., & FAN, Y. N. (1996). Sample quantiles in statistical packages. *American Statistician*, 50, 361-365.
- JIANG, Y., ROUDER, J. N., & SPECKMAN, P. L. (2004). A note on the sampling properties of the Vincentizing (quantile averaging) procedure. Journal of Mathematical Psychology, 48, 186-195.
- LEHMANN, E. L. (1991). *Theory of point estimation*. Pacific Grove, CA: Wadsworth & Brooks.
- LOGAN, G. D. (1988). Toward an instance theory of automization. <u>Psy-</u> chological Review, 95, 492-527.

- Logan, G. D. (1992). Shapes of reaction time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 883-914.
- Luce, R. D. (1986). Response times. New York: Oxford University Press.
- MADDEN, D. J., GOTTLOB, L. R., DENNY, L. L., TURKINGTON, T. G., PROVENZALE, J. M., HAWK, T. C., & COLEMAN, R. E. (1999). Aging and recognition memory: Changes in regional cerebral blood flow associated with components of reaction time distributions. *Journal of Cognitive Neuroscience*, 11, 511-520.
- Penner-Wilger, M., Leth-Steensen, C., & LeFevre, J.-A. (2002). Decomposing the problem-size effect: A comparison of response time distributions across cultures. *Memory & Cognition*, **30**, 1160-1167.
- RATCLIFF, R. (1978). A theory of memory retrieval. <u>Psychological Review</u>, 85, 59-108.
- RATCLIFF, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446-461.
- RATCLIFF, R., & ROUDER, J. N. (1998). Modeling response times for decisions between two choices. *Psychological Science*, 9, 347-356.
- RATCLIFF, R., & ROUDER, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception & Performance*, 26, 127-140.
- ROUDER, J. N. (2000). Assessing the roles of change discrimination and luminance integration: Evidence for a hybrid race model of perceptual decision making in luminance discrimination. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 359-378.
- ROUDER, J. N., SUN, D., SPECKMAN, P. L., LU, J., & ZHOU, D. (2003). A parametric hierarchical framework for inference with response time distributions. *Psychometrika*, **68**, 587-604.
- SCHWARZ, W. (2001). The ex-Wald distribution as a descriptive model of response times. <u>Behavior Research Methods, Instruments</u>, & <u>Computers</u>, 33, 457-469.
- SPECKMAN, P. L., & ROUDER, J. N. (2004). A comment on Heathcote, Brown, and Mewhort's QMLE method for response time distributions. *Psychonomic Bulletin & Review*, 11, 574-576.
- SPIELER, D. H., BALOTA, D. A., & FAUST, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception & Performance*, 22, 461-479.

- THOMAS, E. A. C., & Ross, B. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology*, **21**, 136-152.
- VAN ZANDT, T. (2000). How to fit a response time distribution. <u>Psy-chonomic Bulletin & Review</u>, 7, 424-465.
- VINCENT, S. B. (1912). The function of vibrissae in the behavior of the white rat. *Behavioral Monographs*, 1 (Whole No. 5).

NOTES

- 1. There are a few variants of the Vincentizing procedure, as has been discussed by Heathcote (1996). In this article, we average sample quantiles across individuals. See Heathcote or Van Zandt (2000) for a discussion of alternative averaging methods.
- 2. The term *quantile* is similar to the term *percentile*; the difference is that whereas percentiles range from 0 to 100, quantiles range from 0 to 1. The *p*th quantile of a distribution is the value such that a probability of observing an observation below the value is p. We refer to p as the probability of the quantile—for example, the .1 quantile is the same as the 10th percentile. The *i*th ranked observation served as the quantile estimator for p = i/(M+1), where M is the number of observations. This formula for estimating a probability of a quantile is commonly used in statistical software packages (e.g., SAS's Proc-Univariate). There are other alternatives (see, e.g., Hyndman & Fan, 1996).
- 3. Variability in shape across participants is a key determinant of the properties of Vincentizing. If there is no variability in shape across individuals, RT can be modeled as a location-scale distribution, and Vincentizing yields consistent estimation (Jiang et al., 2004).
- 4. We thank Michael Stadler for use of his data. Participants had to identify in which of four locations an asterisk was presented by depressing one of four computer keyboard keys. Each participant responded to 120 such trials. The last 80 correct observations between 200 and 1,200 msec served as data.
- 5. There is an additional parameter, termed *the coefficient of drift* (Feller, 1968), which serves to scale the units of the other parameters. Without any loss of generality, we set this parameter to 1 msec⁻¹.

(Manuscript received October 11, 2002; revision accepted for publication June 30, 2003.)