A HIERARCHICAL BAYESIAN STATISTICAL FRAMEWORK
FOR RESPONSE TIME DISTRIBUTIONS

JEFFREY N. ROUDER, DONGCHU SUN, PAUL L. SPECKMAN, JUN LU, AND DUO ZHOU

UNIVERSITY OF MISSOURI-COLUMBIA

This paper provides a statistical framework for estimating higher-order characteristics of the response time distribution, such as the scale (variability) and shape. Consideration of these higher order characteristics often provides for more rigorous theory development in cognitive and perceptual psychology (e.g., Luce, 1986). RT distribution for a single participant depends on certain participant characteristics, which in turn can be thought of as arising from a distribution of latent variables. The present work focuses on the three-parameter Weibull distribution, with parameters for shape, scale, and shift (initial value). Bayesian estimation in a hierarchical framework is conceptually straightforward. Parameter estimates, both for participant quantities and population parameters, are obtained through Markov Chain Monte Carlo methods. The methods are illustrated with an application to response time data in an absolute identification task. The behavior of the Bayes estimates are compared to maximum likelihood (ML) estimates through Monte Carlo simulations. For small sample size, there is an occasional tendency for the ML estimates to be unreasonably extreme. In contrast, by borrowing strength across participants, Bayes estimation "shrinks" extreme estimates. The results are that the Bayes estimators are more accurate than the corresponding ML estimators.

Key words: Bayesian analysis, hierarchical models, response time, MCMC, Weibull distribution.

The time taken to complete a task in a psychology experiment is called the response time (RT). RT is a popular dependent measure; for example, it is the dominant measure in paradigms for assessing human memory (Ratcliff, 1978; Sternberg, 1966; Wixted & Rohrer, 1993), attention, (Heathcote, Popiel, & Mewhort, 1991; Posner, 1978; Spieler, Balota, and Faust, 2000; Treisman & Gelade, 1980), word identification (Andrews & Heathcote, 2001; Madden et al., 1999; Plourde & Besner, 1997), learning (Logan, 1988, 1992), perceptual judgments (Petrusic, Baranski, & Kennedy, 1999, Rouder, 2000), and object recognition (Cave & Squire, 1992; Mitchell & Brown, 1988). Most researchers tend to analyze 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. (An incomplete and selective list includes the work of Andrews & Heathcote, 2000; Ashby, Tien, & Balakrishnan, 1993; Hockley, 1984; Logan, 1992; Ratcliff, 1978; Ratcliff & Rouder, 1998, 2000; Rouder, 2000; Spieler, Balota, & Faust, 1996; Smith, 1995; and Van Zandt, Colonius, & Proctor, 2000). In this paper, we propose a Bayesian framework for estimating the shift, scale (variability), and shape of RT distributions.

Although estimating higher order properties is always more difficult than estimating the mean, there is an additional complication for RT distributions. In typical applications, researchers collect a limited number of observations from each of several participants. Researchers seeking to increase power can often increase the number of participants, but not the number of observations per participant. In these applications, RT distributions vary with the participants. Some participants respond relatively quickly while others respond relatively slowly. To gain accurate

589

estimates of distributional properties such as shift, scale, and shape, it is critically important to take account of this variability across participants. If this variability is ignored and all observations are treated as from a common source, then the estimates will be grossly distorted.

If researchers were solely interested in the mean RT across several conditions or populations, then general linear model-based techniques such as ANOVA or regression would be appropriate. There has been steady and continued development in advanced techniques to account for the hierarchical structure of data within the Generalized Linear Model (GLM) framework (e.g., Dey, Ghosh, & Mallick, 2000; Kreft & de Leeuw, 1998). With these techniques, researchers can increase efficiency in estimation and power in inference. The drawback is that these techniques do not provide direct information about shift, scale (variability) or shape of distributions. To fill this void, we propose a statistical framework for estimating these distributional properties while still accounting for participant variability in RT distributions.

Our framework, hierarchical parametric Bayesian framework (HPBF), has three key features. First, the framework is parametric; each observation is assumed to be from a three-parameter Weibull distribution (Johnson, Kotz, & Balakrishnan, 1994). Second, it is hierarchical. At the first level, the Weibull parameters vary from participant to participant. But these sets of Weibull parameters are assumed to come from a common trivariate distribution that depends on the condition or context. For example, suppose there are ten participants in an experiment with five college-aged and five elderly. Then each of the ten participants will have their own sets of Weibull parameters. The five sets of Weibull parameters for the five college-aged participants are assumed to come from one distribution while the five sets of Weibull parameters for the five elderly are assumed to come from a possibly different distribution. A schematic of the framework is shown in Figure 1. In the schematic, the shift parameter for each participant is a sample from one of two distributions. If the participant is college-aged, then the shift parameter is from the distribution on the left. If the participant is elderly, then the shift parameter is from the distribution on the right. Although each participant has his or her own parameters (indicated by X's for college-aged participants and by O's for elderly participants), the distribution of shift parameters for the college-aged participants is smaller than that for the elderly participants. Figure 1 only depicts the hierarchical structure for the shift parameter. But, in general, all three parameters
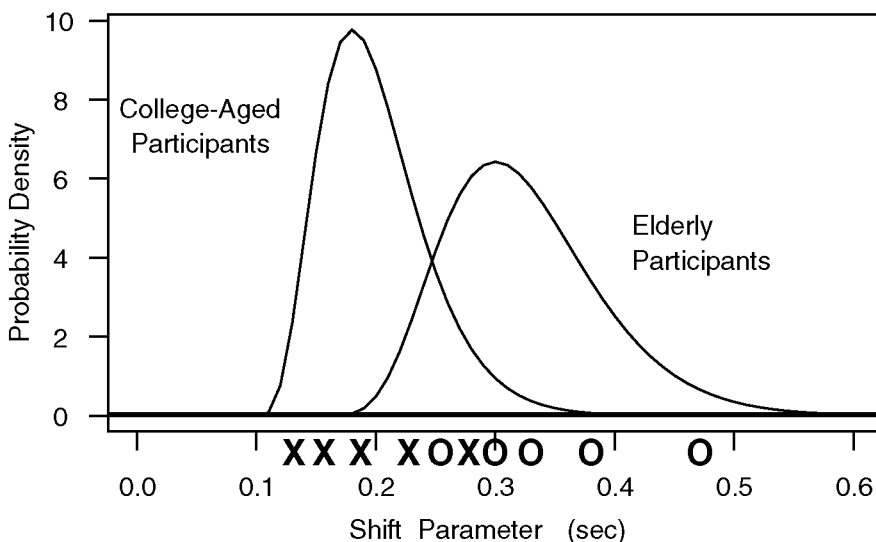

FIGURE 1.
The Hierarchical Framework. Each curve depicts the density from which the shift parameters are sampled. College-aged participants' shift parameters are denoted with X's while elderly participants' shift parameters are denoted with O's.

may have a hierarchical structure. The third feature of the framework is that it is Bayesian. Prior distributions are assumed on parameters and the calculation of the posterior distributions is done through Monte Carlo Markov Chain integration.

The main reason to postulate a hierarchical structure is to provide for better estimation. The gains from "borrowing strength" in a hierarchical, Bayesian framework have been well established in a number of contexts (Gelman, Carlin, Stern, & Rubin, 1995; Tanner, 1993). One recent example of the success of Bayesian analysis in Psychology is that in item response theory (Fox & Glas, 2001; Wang, Bradlow, & Wainer, 2002). In HPBF, parameter estimation and inference are done with Bayesian Monte Carlo Markov Chain methods. As will be shown, the hierarchical framework offers significant advantages over a nonhierarchical framework.

Motivation

In HPBF, we use a parametric distribution. This is a common approach because of the well-known instability of higher order sample moments: sample skew and sample kurtosis require thousands of observations and are very sensitive to outliers (see Heathcoate et al., 1991; and Ratcliff, 1979). Researchers have used standard parametric forms including the Weibull (Logan, 1992) and the log-normal (Ulrich & Miller, 1993), as well as a less-known form termed the ex-Gaussian (Hohle, 1965). The shape of the ex-Gaussian, a convolution of a normal and an exponential density, is sometimes fairly similar in form to the Weibull distribution. The reason we chose the Weibull is four-fold: (a) RT distributions and the Weibull are bounded from below, (b) RT distributions and the Weibull have an elongated right tail, (c) the Weibull is computationally tractable for this application, and (d) the Weibull has broad applicability in statistics—our research is informed by and informs others working with the Weibull.

A major problem with the current state-of-the-art is a lack of powerful tools for estimation and inference about distributional characteristics with small sample sizes. As discussed previously, in many applications, psychologists collect a few observations from each of several participants per experimental condition. In order to arrive at stable population parameter estimates, psychologists pool data across participants with a technique from Vincent (1912; see also Ratcliff, 1979). In this technique, referred to as Vincentizing, quantiles are averaged across participants to produce one composite distribution. For example, the tenth quantiles of the composite distribution is the average of the tenth quantiles across all of the participants. If individual participants' distributions are members of the same location-scale family, then (a) the composite distribution is a member of that family and (b) the parameters of the composite distribution are the arithmetic means of the component distributions (Jiang, Rouder, & Speckman, in press; Thomas & Ross, 1980).

In practice, response times are modeled from three-parameter families that are not location-scale families. For these cases, the large-sample, asymptotic properties of the composite distribution are unknown. It is not at all clear what the Vincentized distribution represents or estimates. In spite of this critical drawback, Vincentizing is still commonly used with nonlocation-scale distributions because the method provides a composite distribution that "looks like" an average of the components (Ratcliff, 1979). The disadvantage of using vincentiles is severe. Since one is not sure what the composite distribution is estimating, inference and the ensuing interpretation of differences across populations or conditions are highly problematic. What is sorely lacking is a principled and powerful statistical platform for combining observations across participants. The HPBF fills this void by jointly modeling variability within and across participants simultaneously.

The Weibull and Its Psychological Interpretation

Each participant provides a series of observations. Let Y_{ij} denote the response time of participant i on trial j ($1 \leq i \leq I$; $1 \leq j \leq J_i$). Each observation is assumed to be independent and identically distributed from a three-parameter Weibull distribution with density
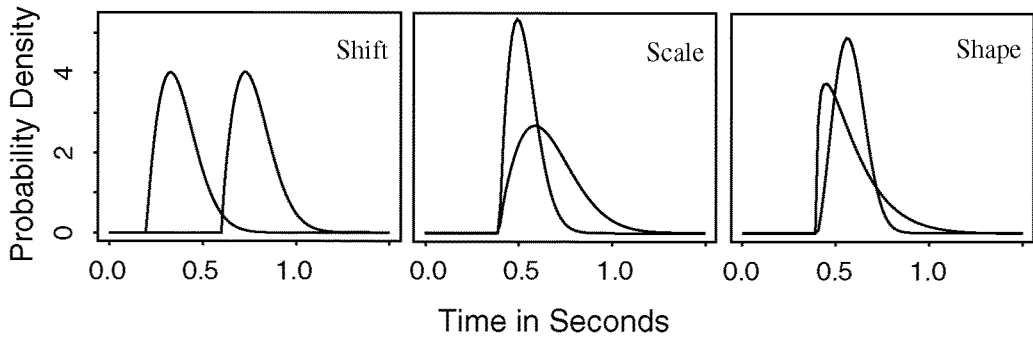
FIGURE 2.
The Weibull parameters of shift, scale and shape. Each plot shows the effect of changing one parameter while holding the other two constant.

$$f(y_{ij} \mid \psi_i, \theta_i, \beta_i) = \frac{\beta_i (y_{ij} - \psi_i)^{\beta_i - 1}}{\theta_i^{\beta_i}} \exp\left\{ -\frac{(y_{ij} - \psi_i)^{\beta_i}}{\theta_i^{\beta_i}} \right\}, \quad y_{ij} > \psi_i. \tag{1}$$

Weibull parameters (ψ, θ, β) are interpretable; they correspond to the heuristics of location (a shift parameter characterized as the lower bound), scale, and shape of the distribution, respectively. The role of these three parameters is shown in Figure 2. The Weibull is quite flexible encompassing any scale and shift. The shape can be varied from highly right skewed to nearly symmetric.

The parameters of the Weibull have a psychological interpretation. In experimental psychology, there is a broad, long-standing distinction between two types of processes: central and peripheral[1] (Balota & Spieler, 1999; Dzhafarov, 1992; Luce, 1986). Peripheral processes are quick sensory processes which occur automatically, whereas central processes are processes that require conscious control and attention (e.g., Hasher & Zacks, 1979, Jacoby, 1991, Luce, 1986, Schnieder & Shiffrin, 1977). For example, moving your eyes to a location of a bright flash relies on peripheral processes while maintaining a ten digit phone number in memory relies mainly on central processes. It is common to assume that the latency of peripheral processes has little variability while the latency of central processes is variable and skewed.

Differences in the structure of central processes across groups or conditions would be manifested as a difference in the shape parameter. Difference in the structure of central processes would include the insertion of stages (e.g., Ashby & Townsend, 1980; Balota & Chumbley, 1984) or changes in search strategy (such as those frequently encountered in visual search tasks, e.g., Treisman & Gelade, 1980). However if the central processes follow the same structure across different groups or conditions, but the speed of execution is different, then there would be differences in the scale parameter but not the shape parameter. Finally, differences in the speed of peripheral processes are manifested largely in changes in the shift parameter (e.g., Balota & Spieler, 1999; Hockley, 1984; Ratcliff, 1979).

The shift parameter, ψ_i, connotes the quickest possible time for a stimulus-initiated response. This is the minimum time it takes for the stimulus to be converted to nervous system energy, propagate to the brain, and then to the part of the body making the response. Most substantive theories incorporate the concept of a minimum RT in one way or another. For example, in stochastic process models of decision making (e.g., Busemeyer & Townsend, 1993; Link, 1975; Ratcliff, 1979; Rouder, 1996, 2001; Townsend & Ashby, 1983), the predicted RT can be ex-

---

[1] The terms central and peripheral do not necessarily refer to locations in the body or brain. It is known that central processes are located in the cortex, but peripheral process may be cortical or subcortical. Peripheral processes include processes that may occur outside of the brain, for example, motor processes.

pressed as $RT = a + D$ where a is a constant and $D \geq 0$ is a random variable denoting the time for central process completion. This constant denotes the minimum (it is also called the *residual*, Luce, 1986, the *time-for-encoding-and-response*, Ratcliff, 1979; Ratcliff & Rouder, 1998, or the *irreducible minimum*, Dhzafarov, 1992; Hsu, 1999). Estimation of an irreducible minimum provides an important constraint on substantive theories. Substantive theories often make the prediction that the irreducible minimum does not change across manipulations hypothesized to affect only central processing.

Single Population Model

The Model

In this section, we implement the HPBF for a single population as a first step in model development. In the Weibull model, (1), the parameters ψ_i, θ_i and β_i represent the i-th individual's shift, scale and shape parameters, respectively. The parameters $(\psi_i, \theta_i, \beta_i)$ are assumed to come from a prior distribution.[2] In this model, the prior on each individual's shift parameter ψ_i is flat, for example, uniform over a finite range. The other two parameters, shape and scale, are assumed to be samples from a two-stage hierarchical prior distribution, whose first stage prior is given by

$$(\beta_i \mid \eta_1, \eta_2) \overset{\text{iid}}{\sim} \text{Gamma}(\eta_1, \eta_2) \text{ restricted to } \beta_i > 0.01, \tag{2}$$

$$(\theta_i^{-\beta_i} \mid \xi_1, \xi_2) \overset{\text{iid}}{\sim} \text{Gamma}(\xi_1, \xi_2), \tag{3}$$

where Gamma (η_1, η_2) denotes the gamma distribution with density

$$f(t \mid \eta_1, \eta_2) = \eta_2^{\eta_1} t^{\eta_1 - 1} \exp(-\eta_2 t) / \Gamma(\eta_1)$$

for $t > 0$.

For the first stage, the prior for the shape parameter β_i is a gamma distribution with parameters η_1 and η_2 restricted to the range $\beta_i > 0.01$. This somewhat unusual restriction is a technical one needed to insure that posterior moments exist for the θ_i. The additional restriction is easily implemented in the MCMC simulation by sampling β_i from the posterior density in (c) below restricted to $[0.01, \infty)$. Although there is a theoretical problem with existence of the posterior mean of θ_i with the more common gamma prior, the change in the sampling procedure for the β_i had no effect at all on the data analyses reported below. For given shape parameter β_i, the conditional distribution of the scale parameter θ_i is sampled so that $\theta_i^{-\beta_i}$ is from a Gamma(ξ_1, ξ_2) distribution. Equivalently, $\theta_i^{\beta_i}$ is sampled from an inverse gamma distribution with parameters ξ_1 and ξ_2, that is, $t = \theta_i^{\beta_i}$ has density $g(t \mid \xi_1, \xi_2) = \xi_2^{\xi_1} e^{-\xi_2/t} / [t^{\xi_1+1} \Gamma(\xi_1)]$ for $t > 0$. This distribution will be called Inverse Gamma(ξ_1, ξ_2). This choice of priors is quite general in the sense that it is flexible and can model a reasonably broad class of prior distributions. The setup is also convenient for computational reasons, as will be seen in the computational section below, and generalizes the prior specification advocated by Berger and Sun (1993).

Hierarchical priors are used for β and θ^β, but not for ψ. The uniform prior on ψ is relatively uninformative and the model is computationally simple. This simplicity allowed us to explore more easily the properties of the Bayesian procedure through simulation. In the current model, the hierarchical priors in β and θ^β yield shrinkage in all three parameters. It is not clear that a hierarchical prior on ψ would yield any additional gain.

The parameters $(\xi_1, \xi_2, \eta_1, \eta_2)$ of the first stage prior serve as *hyper-parameters*. These hyper-parameters describe how the shape and scale vary across individuals within the population. The second stage prior is given by mutually independent distributions:

---

[2]We refer to distributions of parameters and hyper-parameters as *prior distributions* with no distinction between distributions of variables we regard as latent (such as shift, shape and scale) and distributions reflecting a priori beliefs about a parameter.

$$\xi_k \sim \text{Gamma}\,(a_k, b_k), \qquad k = 1, 2, \tag{4}$$

$$\eta_k \sim \text{Gamma}\,(c_k, d_k), \qquad k = 1, 2. \tag{5}$$

The main goal of practitioners is to do inference about differences in shift, scale and shape as a function of condition or population. In terms of HPBF, this means that practitioners will make statements about second-stage parameters, $(\eta_1, \eta_2, \xi_1, \xi_2)$. One critical benefit of the model is that it does indeed provide for inference on these second-stage parameters. That is, one is able to gather estimates and prediction intervals about the distribution of scale and shape across a population.

Parameter Estimation

In Bayesian frameworks, the goal is to obtain posterior quantities such as the posterior mean, variance, and the marginal posterior distribution for a parameter. We implemented parameter estimation with Gibbs Sampling (see Gelfand & Smith, 1990) as follows.

Full-conditional distributions. To implement Gibbs sampling, we need the full conditional posterior distributions, given here and derived in the Appendix. In the following, we use $[t|\cdot]$ to denote the conditional density of t given its argument, and $(t|\cdot)$ is the corresponding conditional distribution.

Fact 1. The full conditional posterior distributions of ($\boldsymbol{\psi}$, $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\boldsymbol{\eta}$) given $\mathbf{y}$ are as follows.

(a) For given ($\boldsymbol{\theta}$, $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\boldsymbol{\eta}$; $\mathbf{y}$), the ψ_i are independent, and the conditional posterior density of ψ_i is

$$[\psi_i | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\eta}; \mathbf{y}] \propto \left\{ \prod_{j=1}^{J_i} (y_{ij} - \psi_i) \right\}^{\beta_i - 1} \exp\left\{ -\frac{\sum_{j=1}^{J_i} (y_{ij} - \psi_i)^{\beta_i}}{\theta_i^{\beta_i}} \right\},$$

where $0 < \psi_i < \min_{1 \le j \le J_i} y_{ij}$.

(b) For given ($\boldsymbol{\psi}$, $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\boldsymbol{\eta}$; $\mathbf{y}$), the θ_i are independent and the conditional posterior distribution of $\theta_i^{\beta_i}$ is

$$(\theta_i^{\beta_i} | \boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\eta}; \mathbf{y}) \sim \text{Inverse Gamma}\left(\xi_1 + J_i, \xi_2 + \sum_{j=1}^{J_i} (y_{ij} - \psi_i)^{\beta_i} \right).$$

(c) For given ($\boldsymbol{\psi}$, $\boldsymbol{\theta}$, $\boldsymbol{\xi}$, $\boldsymbol{\eta}$; $\mathbf{y}$), the β_i are independent and the conditional posterior density of β_i is

$$[\beta_i | \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\eta}; \mathbf{y}] \propto \frac{\beta_i^{J_i + \eta_1}}{\theta_i^{(J_i + \xi_1)\beta_i}} \exp\left\{ -\frac{\xi_2 + \sum_{j=1}^{J_i} (y_{ij} - \psi_i)^{\beta_i}}{\theta_i^{\beta_i}} - \eta_2 \beta_i \right\} \prod_{j=1}^{J_i} (y_{ij} - \psi_i)^{\beta_i}.$$

(d) For given ($\boldsymbol{\psi}$, $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, ξ_2, $\boldsymbol{\eta}$; $\mathbf{y}$), the conditional posterior density of ξ_1 is

$$[\xi_1 | \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \xi_2, \boldsymbol{\eta}; \mathbf{y}] \propto \left(\frac{\xi_2^I}{\theta_i^{\beta_i}} \right)^{\xi_1} \frac{\xi_1^{a_1 - 1}}{\{\Gamma(\xi_1)\}^I} e^{-b_1 \xi_1}.$$

(e) For given ($\boldsymbol{\psi}$, $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, ξ_1, $\boldsymbol{\eta}$; $\mathbf{y}$), the conditional posterior distribution of ξ_2 is

$$(\xi_2 | \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \xi_1, \boldsymbol{\eta}; \mathbf{y}) \sim \text{Gamma}\left(a_2 + I\xi_1, b_2 + \sum_{i=1}^{I} \frac{1}{\theta_i^{\beta_i}} \right).$$

(f) For given $(\boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}, \eta_2; \mathbf{y})$, the conditional posterior density of η_1 is

$$[\eta_1 \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}, \eta_2; \mathbf{y}] \propto \frac{\left(\eta_2^I \prod_{i=1}^I \beta_i\right)^{\eta_1}}{\{\Gamma(\eta_1)\}^I} \eta_1^{c_1 - 1} \exp\{-d_1 \eta_1\}.$$

(g) For given $(\boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}, \eta_1; \mathbf{y})$, the conditional posterior distribution of η_2 is

$$(\eta_2 \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}, \eta_1; \mathbf{y}) \sim \text{Gamma}\left(c_2 + I\eta_1, d_2 + \sum_{i=1}^I \beta_i\right).$$

Sampling full-conditional distributions. Sampling from the full conditionals given in (b), (e), and (g) in Fact 1 is relatively straightforward as they are functions of Gamma and Inverse Gamma densities. The method we used is the modified rejection technique of Ahrens and Dieter (1974, 1982). The sampling of the remaining full-conditionals is somewhat more complex. One advantage of the current choices of priors is that they yield full-conditional densities that are log concave. (A density fit is log concave if $\frac{d^2}{dt^2}(\log f(t)) \le 0$ for all t.)

Fact 2.

(i)   The conditional density in (a) is log-concave if $\beta_i \ge 1$.
(ii) The conditional posterior density in (c) is log-concave.
(iii) If $I + 2a_1 - 2 \ge 0$, the conditional posterior density in (d) is log-concave.
(iv) If $I + 2c_1 - 2 \ge 0$, the conditional posterior density in (f) is log-concave.

The proof of the above fact is also given in the Appendix. With log-concave full-conditionals, we can use the adaptive-rejection method of sampling due to Gilks and Wild (1992; Wild & Gilks, 1993; see also Berger & Sun, 1993). If the full conditional densities are not log-concave, the adaptive rejection technique can be replaced with a Metropolis-Hastings step. The advantage of the adaptive rejection technique (without a Metropolis-Hastings step) for log-concave distributions is that of computational speed.

Software. The Gibbs Sampling technique was implemented as a C-language program for UNIX/Linux platforms. Random number generation was done through the ran2 routine (Press, Teukolsky, Vetterling & Flannery, 1992). Source code, including a version of the Gilks and Wild adaptive-rejection sampling, is freely available by request.

Benchmarking the Single Population Model

The above single-population model was benchmarked vis-a-vis a nonhierarchical maximum likelihood (ML) approach. A three-parameter Weibull distribution was fit to each individual participant.[3] The nonhierarchical ML approach is the best yardstick to benchmark the model because it typifies the state-of-the art in experimental psychology.

Data Set

For a test, we used an unpublished data set in implicit learning provided by Michael Stadler. The data we analyze were from a practice block in Stadler's experiment. Participants were given one of four simple stimuli (an asterisk at one of four locations on a computer screen) and had

---

[3] The negative log-likelihood function was minimized with repeated applications of the simplex algorithm (Nelder & Mead, 1965).

to make one of four responses by pressing a key on a computer keyboard. Each stimulus was mapped to a unique response. Eighty participants responded to 120 such trials.

Before analysis, we trimmed outlying RTs of less than 0.2 seconds and greater than 1.2 seconds. The justification is that responses less than 0.2s and greater than 1.2s are too fast or slow to be related to the stimulus onset. In Stadler's data, less than .25% of the observations were trimmed. For each participant, we used the last 80 trials with correct responses and RTs in the 0.2s to 1.2s range as our benchmark data set. The choice of using the last 80 trials that met the above criterion was motivated by the desire to minimize the impact of practice. The first 40 trials will be impacted to a larger degree by practice effects than will be the latter 80. The model was not adapted to include trimming. Given the extremely small number of trimmed observations, the lack of accounting for trimming is negligible. Future models will be more complex.

A summary analysis of the data revealed great variability of response times both within participants and across participants. The mean of the sample standard deviations serves as an index of the amount of variability within participants while the standard deviation of the sample means serves as an index of the amount of variability between participants. The within-participants index (.093s) is only about twice as big as the between-participants index (.055s). This means that between-participants variability is moderately large and such variability may affect estimation. To further illustrate the types of variability in the data, histograms from six participants are shown in Figure 3. The left column shows two participants who differ greatly in shift, especially with regard to the onset of the distribution. The following columns show differences among participants in scale and shape. The data for these participants are fairly extreme and define the range of variability on these distribution characteristics.
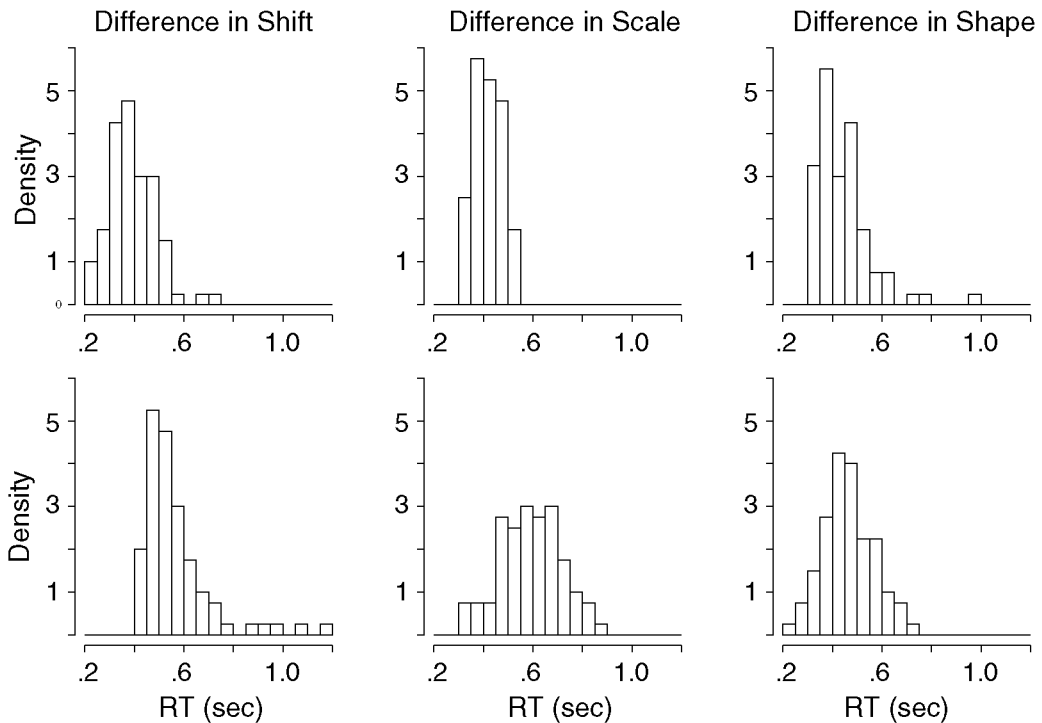

FIGURE 3.
Histograms of RTs for selected pairs of participants. The left column emphasizes differences in shift across participants. The middle and right columns emphasize differences in scale and shape, respectively.

Bayesian Estimation

Choice of priors. To compute estimates with HPBF, a prior density must be placed on primary parameters ψ_i as well as hyper-parameters η_k and ξ_k by choosing values for (a_k, b_k, c_k, d_k), $k = 1, 2$. The prior for ψ_i is a uniform of finite range. All that is necessary is that the the top bound be greater than the value of the minimum observations. We used a uniform from 0 sec to 1 sec. Indeed, the minimum observation for each participant was less than 1 sec. The choices for the hyper-parameters take some ingenuity as they specify priors (hyper-priors) for the parameters of the first level priors on the actual parameters (θ_i, β_i) of interest. The priors used were chosen to be "vaguely informative", that is, they are not so diffuse as to be completely noninformative, but they were designed with prior knowledge of the kinds of response times that are obtained in practice.

The following computations will be used in choosing priors. To find the hyper-parameters (a_k, b_k) and (c_k, d_k), $k = 1, 2$, we need the marginal priors of β_i and θ_i. Letting $f(\cdot|\cdot)$ denote the gamma density as above, the marginal prior of β_i is

$$m_1(\beta) = \int_0^\infty \int_0^\infty f(\beta|\eta_1, \eta_2) f(\eta_1 \mid c_1, d_1) f(\eta_2 \mid c_2, d_2) \, d\eta_1 \, d\eta_2 \tag{6}$$

$$= \frac{d_1^{c_1} d_2^{c_2}}{\Gamma(c_1)\Gamma(c_2)} \int_0^\infty \int_0^\infty \frac{\eta_2^{\eta_1} \beta^{\eta_1-1} \eta_1^{c_1-1} \eta_2^{c_2-1}}{\Gamma(\eta_1)} \exp\left\{-d_1\eta_1 - (d_2 + \beta)\eta_2\right\} d\eta_1 \, d\eta_2$$

for $\beta > 0$. Using the Inverse Gamma density $g(\cdot|\cdot)$ given above and changing variables, the marginal prior for θ_i is

$$m_2(\theta) = \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \beta\theta^{\beta-1} g(\theta^\beta \mid \xi_1, \xi_2) f(\beta \mid \eta_1, \eta_2) \tag{7}$$

$$\times \prod_{k=1}^2 f(\xi_k \mid a_k, b_k) f(\eta_k \mid c_k, d_k) \, d\beta \, d\eta_1 \, d\eta_2 \, d\xi_1 \, d\xi_2$$

$$= \left\{\prod_{k=1}^2 \frac{b_k^{a_k} d_k^{c_k}}{\Gamma(a_k)\Gamma(c_k)}\right\} \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty G \, d\beta \, d\xi_1 \, d\xi_2 \, d\eta_1 \, d\eta_2,$$

where

$$G = \frac{\beta^{\eta_1} \xi_1^{a_1-1} \xi_2^{\xi_1+a_2-1} \eta_1^{c_1-1} \eta_2^{\eta_1+c_2-1}}{\theta^{\xi_1\beta+1}\Gamma(\xi_1)\Gamma(\eta_1)} \exp\left\{-b_1\xi_1 - b_2\xi_2 - d_1\eta_1 - (\beta + d_2)\eta_2 - \frac{\xi_2}{\theta^\beta}\right\}.$$

To obtain these priors and judge the effects of the choice of hyper-parameters, we used the marginal priors given in (6) and (7). The marginal prior $m_1(\beta)$ can be evaluated by a one dimensional numerical integration. By studying graphs of various combinations of parameters (c_k, d_k), $k = 1, 2$, the parameters in Table 1 labeled "vague prior" were chosen. Priors were picked based on prior knowledge of response times to simple stimuli (without examining the data). RT typically range from about 200ms to perhaps 800ms. A uniform prior on [0, 1 sec] was taken for ψ_i because we wanted to supply very little prior information. Note that we allow ψ_i to be smaller than 0.2s even though we assume that response times must be larger than 0.2s. From inspection of previous research, the shape parameter for response times is generally between 1.4 and 3.3 with a central tendency around 2.0 (see Logan, 1992). The marginal distribution on shape induced from the vague prior of Table 1 is graphed in the bottom panel of Figure 4. Because the marginal prior for θ, $m_2(\theta)$ given in (7), cannot be analytically reduced to less than three integrals, we used histograms of Monte Carlo simulations from potential marginal distributions to see the effects of the hyper-parameters on the marginal prior distribution. To simulate, we first

TABLE 1.
Prior parameters for HPBF single-population model and four alternatives.

| | $\theta^{-\beta}$ | | | | β | | | |
| | ξ_1 | | ξ_2 | | η_1 | | η_2 | |
	a_1	b_1	a_2	b_2	c_1	d_1	c_2	d_2
Vague Prior	2.0	0.1	2.0	2.85	2.0	0.02	2.0	0.04
Alternative Prior 1	2.5	1.0	2.5	1.0	0.3	0.01	0.28	0.02
Alternative Prior 2	5.0	0.5	5.0	0.5	0.3	0.01	0.28	0.02
Alternative Prior 3	2.0	0.1	2.0	2.85	1.0	1.0	1.0	1.0
Alternative Prior 4	5.0	0.5	5.0	0.5	1.0	1.0	1.0	1.0

drew parameters $(\eta_1, \eta_2, \xi_1, \xi_2)$ from the second-level prior distributions specified by trial values of the (a_k, b_k, c_k, d_k). Values of β and $t = \theta^\beta$ were then drawn from the respective Gamma and Inverse Gamma distributions, and finally $\theta = t^{1/\beta}$ was obtained. The values of (a_1, b_1) and (a_2, b_2) for the "vague prior" in Table 1 were chosen after considerable experimenting. The marginal prior as shown in Figure 4 is reasonably flat with a long right tail. Again, this is a rather noninformative choice as the prior experience suggested that mass should be between 50ms and 400ms.

To study the effect of our choice of prior hyper-parameters, we also analyzed the data using four other priors, listed in Table 1. The four alternative priors were chosen to be extreme alternatives to the vaguely informative prior of Table 1. The marginal prior on β induced by Priors 1 and
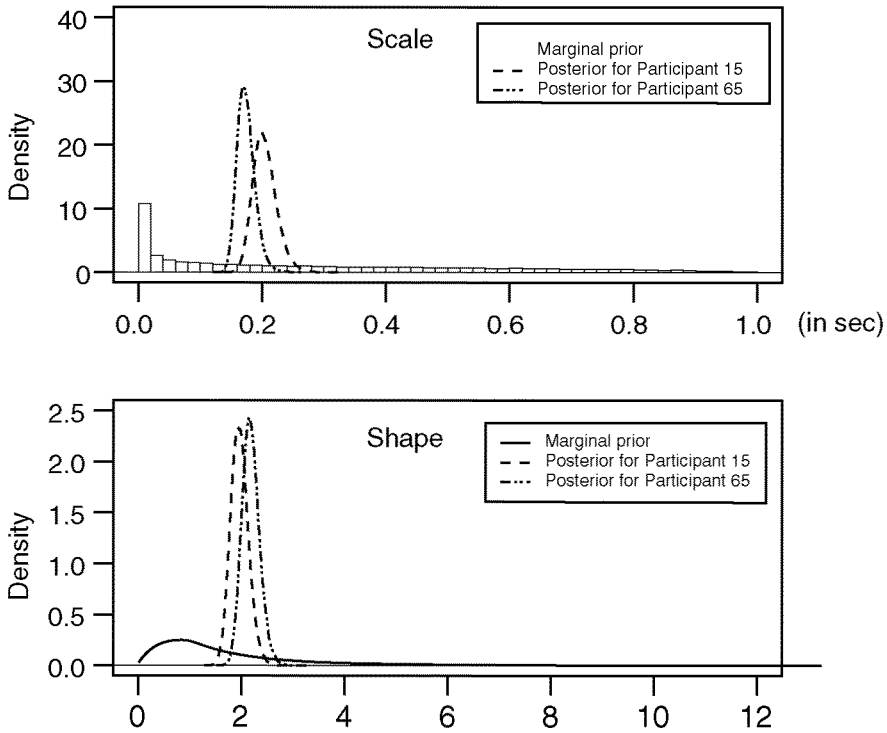

FIGURE 4.
Prior and posterior densities of the scale (θ) and shape (β) parameters for two representative participants.

2 has mass shifted toward zero; that induced by Priors 3 and 4 has mass shifted toward implausibly large values. The marginal prior on θ induced by all four alternative priors is bimodal with a second mode around 1 sec. Priors 1, 3, and 4 have significant mass above 1 sec. Even though these priors are extreme, the resulting posterior estimates are very close to those obtained with the vague prior. Hence, for this model and data set, parameter estimates are not overly sensitive to specification of the prior.

Figure 4 also depicts posterior density estimates for two representative participants. As can be seen, the posterior densities have far less spread than the prior densities, indicating that the knowledge gained about the parameters has dramatically increased.

Chain convergence. We conducted extensive experiments to determine MCMC convergence with respect to initial values. The primary parameters ψ, β and θ converge rapidly. Convergence is slower for the hyper-parameters in general; parameters η_2 and ξ_2 were the slowest to converge. To formally assess convergence, we followed Gelman and Rubin (1992) and estimated the scale reduction factor, $\hat{R}$, of parallel sequences. We experimented with initial values 0.0001, 5.0, 80.0 and 100.0 for η_1; 0.1, 1, 2 and 4 for η_2; 0.0001, 5.0, 30.0, and 40.0 for ξ_1; 0.1, 1, 2 and 4 for ξ_2; and 0.0001, 0.8, 2.5 and 3.0 for β_j. Because the posterior distribution of ψ_j must belong to the interval $(0, \psi_j^u)$, where $\psi_j^u = \min_k y_{jk}$, we tried initial values $\psi_j^u/5$, $2\psi_j^u/5$, $3\psi_j^u/5$, and $4\psi_j^u/5$. Since θ_j can be simulated at the beginning of each of the Gibbs cycles, no initial values are needed. In all of these cases, the estimated scale reduction factors $\hat{R}$ for the hyper-parameters were less then 1.05 when the Gibbs sample sizes are larger than 3,000. This value compares favorably with Gelman and Rubin's recommendation that $\hat{R}$ be less than 1.20. These experiments resulted in convergence of the MCMC simulation regardless of starting values.

For the comparisons described in the next section, we used a very conservative burn-in period of 10,000 cycles with an additional 50,000 for estimation.

Comparing ML and HPBF estimates

To obtain point estimates from HPBF, we estimated the means of the marginal posterior distributions by computing sample averages from the MCMC simulations. The point estimates from both ML and HPBF are displayed in Figure 5. The top row of the figure shows boxplots of the estimates across participants. The bottom row shows scatter plots in which the ML estimates are plotted on the x-axis and the Bayesian estimates are plotted on the y-axis. The Bayesian and ML estimates tend to have similar medians across participants. For participants with moderate estimates, the two estimators tend to be highly similar. Overall, there is less variability in the Bayesian estimates as evidenced in both the boxplots as well as by a slope less than unity in the scatter plots. This effect is most noticeable for the shape parameter. The most extreme ML estimators tend to be more extreme than their Bayesian counterparts, especially for small values of shift and large values of shape.

For several participants, ML and HPBF estimates are nearly identical. However, for some participants, there are differences. Figure 6 highlights these differences and also shows the effect of shrinkage in the HPBF estimates. In the figure, the Weibull density is fit to the empirical histograms. The solid lines denote the case in which the parameters are the MLEs; the dotted lines denote the case in which the parameters are the posterior-mean point estimates from HPBF. For the top two panels, Participants 3 and 5, the model fits and data show a fair amount of skew. But, the ML fit is more skewed than the HPBF fit. In the bottom two panels, Participants 38 and 63, the model fits and data show a fair amount of symmetry. But, the ML fit is more symmetrical than the HPBF fit. In these cases, the differences in shape are accompanied by substantial differences in shift. As shown by the arrows, the shift estimate for the ML method is several tens of milliseconds lower than that of the HPBF method.
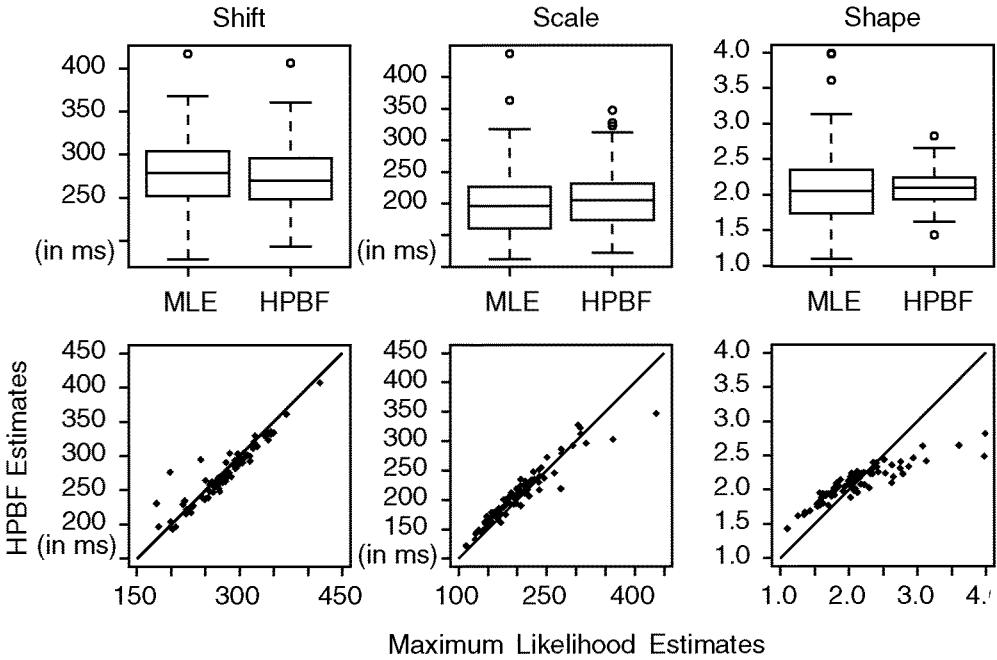
FIGURE 5.
MLE and Bayesian estimates for Stadler's data.

Simulation Studies

The question facing researchers is which method better estimates the true values of the parameters. To help answer this question, we performed a small Monte Carlo study in which the estimates from Stadler's data served as "true values". In Simulation 1, each participant's ML estimates served as the "true values" and were used to generate a set of 80 observations per participant. We then fit the simulated data set with both the ML and HPBF methods. A total of 1055 such tests were performed.

To speed convergence in the simulation experiments, we used starting values based on the ML estimates. For example, to determine a suitable set of starting values for η_1 and η_2, we used the method of moments on the ML estimates $\hat{\beta}_j$, equating the sample mean to η_1/η_2 and the sample variance to η_1/η_2^2. A similar procedure was used with the estimates $\hat{\lambda}_j = \hat{\theta}_j^{-\hat{\beta}_j}$. The sample mean and variance of $\hat{\lambda}_j$ were equated to ξ_1/ξ_2 and ξ_1/ξ_2^2 respectively to obtain initial values. With these carefully chosen starting values, a burn-in period of 3,000 cycles with an additional 3,000 samples for estimation were sufficient.

Based on these simulations, we calculated the root mean squared error (RMSE) and mean absolute error (MAE) across all participants. The reported numbers are the average of these RMSEs and MAEs across tests. Simulation 2 was identical to Simulation 1 except that each participant's HPBF estimates were used as "true values". There were 854 tests. Because the HPBF Bayes estimates are optimal when data are actually sampled from a HPBF, we anticipated that the Bayesian estimates would do especially well in Simulation 2.

Table 2 shows the results of the simulations. There is a substantial advantage for the HPBF estimators over their ML counterparts in terms of RMSE. This advantage is due in part to the fact that the actual sampling distributions of the MLEs have rather heavy tails, and occasional extreme estimates are produced. RMSE tends to be more heavily influenced by outliers than mean absolute error. By borrowing strength, the Bayes estimates tend to be much better in samples
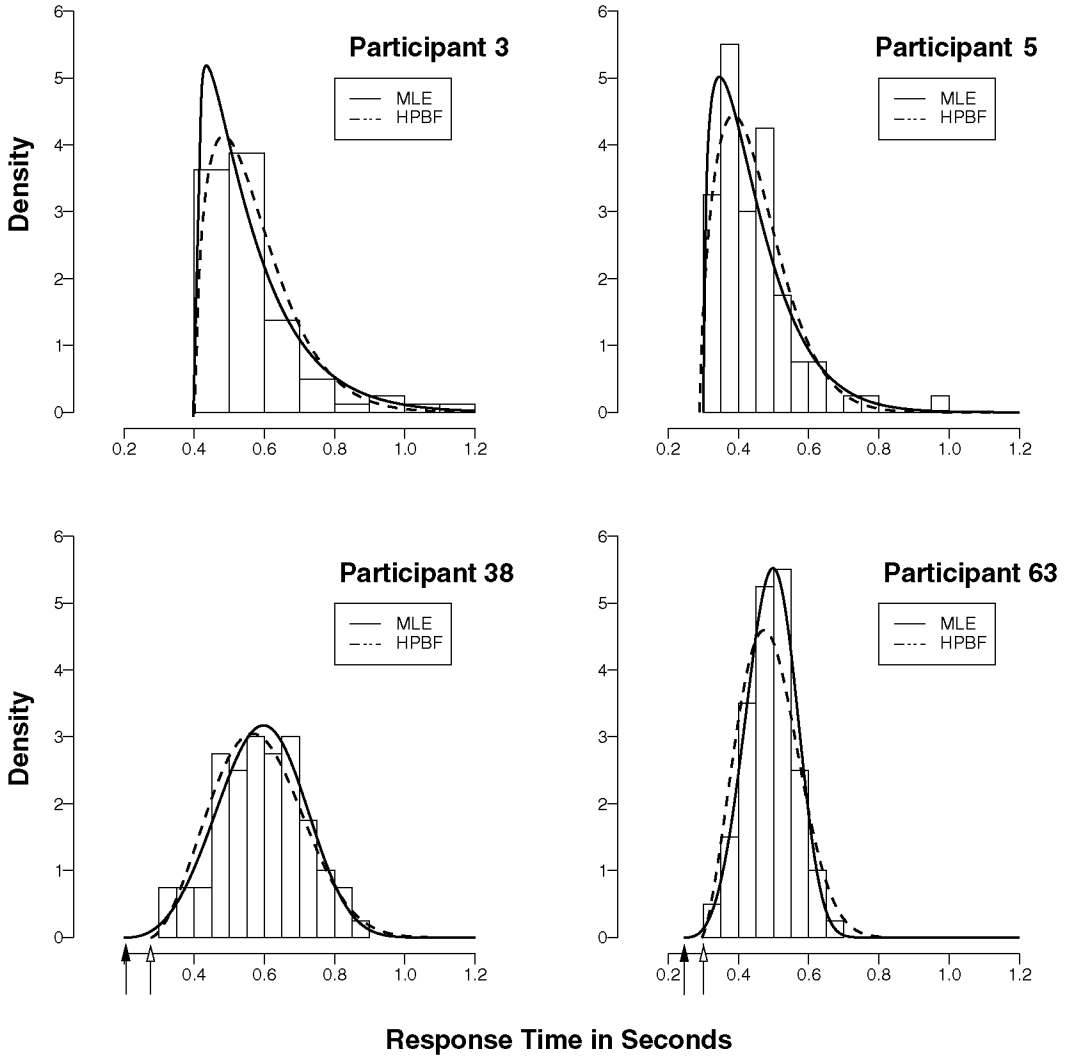
where ML performs poorly. This effect is so pronounced that the ratio of RMSE of the MLE to the Bayes estimates is greater for "MLE truth" than for "Bayes truth".

In the above simulation with 80 observations per participant, we showed an advantage of HPBF over ML. We further benchmarked the two approaches with small sample sizes. Instead of analyzing 80 observations per participant, we limited the data set to the last 20 observations per participant. Parameters were estimated as in the previous analysis. Figure 7 shows box plots and scatter plots for both Bayesian and ML estimates. One of the participants was removed because of extreme MLE estimates (shift of −12 sec, shape of 250). Even after this outlying participant is removed, there is more variability in the MLE estimators for shift and scale and dramatically more variability for shape.

We performed a small Monte Carlo simulation to help interpret these differences in estimation. The simulation was similar to those previously described. MLE estimates from all eighty observations for each of the eighty participants were used as "truth values". Then, data sets,

TABLE 2.
Benchmark of MLE vs. HPBF

	MLE as true values		HPBF as true values	
	RMSE	MAE	RMSE	MAE
Shift (ψ)				
ML Estimates	.0351	.0172	.0216	.0155
HPBF Estimates	.0221	.0148	.0152	.0116
Ratio	1.59	1.16	1.42	1.33
Scale (θ)				
ML Estimates	.0402	.0228	.0274	.0207
HPBF Estimates	.0273	.0195	.0201	.0156
Ratio	1.47	1.17	1.36	1.34
Shape (β)				
ML Estimates	.5457	.2998	.3488	.2652
HPBF Estimates	.3896	.2739	.1944	.1525
Ratio	1.40	1.09	1.79	1.74

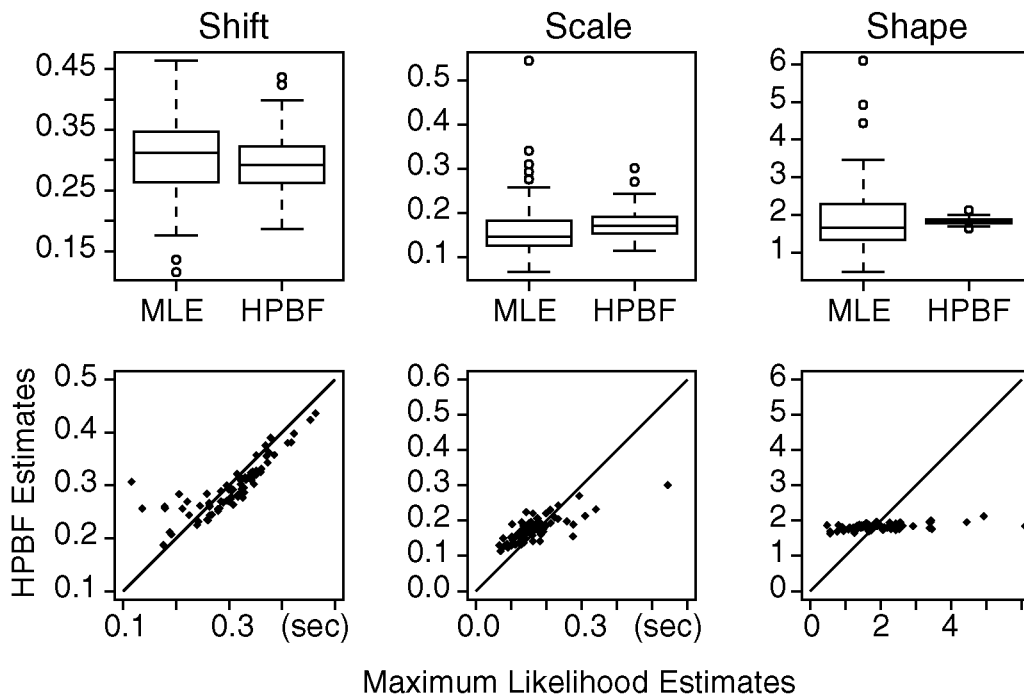RMSE: Root-mean-squared error.
MAE: Mean absolute error.

FIGURE 7.
MLE and Bayesian estimates for the last 20 trials in Stadler's data. Top row shows boxplots of estimates with one participant's estimates removed (MLE estimates were exceedingly extreme). Bottom row shows the scatter plot of Bayesian estimates as a function of MLE estimates with the outlying participant removed.

TABLE 3.
Benchmark of MLE vs. HPBF for small sample sizes

	Shift (ψ)		Scale (θ)		Shape (β)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
MLE	1.264	.226	1.269	.238	21.363	3.855
HPBF	.038	.027	.044	.032	.549	.397
Ratio	33.6	8.4	28.7	7.4	39.0	9.7

consisting of twenty observations for each of the eighty participants were generated. Based on these generated data sets, MLE and HPBF estimates were obtained. There were 866 such tests. The RMS-error between the true values and the estimates is shown in Table 3. As can be seen, HPBF estimates are dramatically more accurate than the ML estimates. The large RMS-error for the MLE reflects the presence of extreme estimates. The extreme estimates from the data are neither unexpected nor due to misspecification of the Weibull model; they are common even in estimates from data simulated from the Weibull distribution. Thus, the ML approach is flawed for estimating shift, scale, and shape Weibull parameters with small sample sizes.

The difficulty of ML estimation with the three-parameter Weibull is well known in the statistics literature (e.g., Smith & Naylor, 1987) and is a result of violations of the regularity condition (see Johnson, Kotz, & Balikrishnan, 1994). Psychologists have also been aware of this problem with MLE estimation for the three-parameter Weibull (Dolan, van der Maas, & Molenaar, in press; Heathcote, 1996; Van Zandt, 2000). Even with these well-known problems, ML estimation is considered the gold-standard for estimation of RT distributions. Van Zandt, in her recent review of several estimation techniques, finds no method that performs as well as MLE. The hierarchical approach presented here is new and was not available to Van Zandt for evaluation. Clearly it greatly outperforms ML estimation for small-sample sizes and provides an attractive alternative for psychologists who need to pool data across participants.

Concluding Points

In this paper we presented a model of RT distributions. The approach is hierarchical, accounting for variability within and between participants simultaneously. Estimation and inference are done with Bayesian Monte Carlo Markov Chain sampling methods. Overall, estimation is superior to the nonhierarchical, maximum likelihood approach. We believe this statistical platform for higher-order distribution properties will aid in developing and testing psychological theory.

The Appendix

Proof of Fact 1. For (a), since $[\psi] \propto 1$, we have

$$[\psi \mid \theta, \beta, \xi, \eta; y] \propto [y \mid \psi, \theta, \beta] \propto \prod_{i=1}^{I} \prod_{j=1}^{J_i} [y_{ij} \mid \psi_i, \theta_i, \beta_i].$$

The result follows. For (b), note that

$$[\theta \mid \psi, \beta, \xi, \eta; y] \propto [y \mid \psi, \theta, \beta] \, [\theta \mid \beta, \xi]$$

$$\propto \prod_{i=1}^{I} \frac{1}{\theta_i^{(\xi_i + J_i)\beta_i + 1}} \exp \left\{ -\frac{\xi_2 + \sum_{j=1}^{J_i} (y_{ij} - \psi_i)^{\beta_i}}{\theta_i^{\beta_i}} \right\}.$$

The result is immediate. For (c), we get

$$[\boldsymbol{\beta} \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\eta}; \mathbf{y}] \propto [\mathbf{y} \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}]\, [\boldsymbol{\theta} \mid \boldsymbol{\beta}, \boldsymbol{\xi}]\, [\boldsymbol{\beta} \mid \boldsymbol{\eta}].$$

Other conditional distributions can be proved similarly. □

Proof of Fact 2. For (i), if $\psi_i \in (0, \min_{1 \le j \le J_i} y_{ij})$, the first derivative of the logarithm of the conditional posterior density of ψ_i is

$$\frac{\partial \log[\psi_i \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\eta}; \mathbf{y}]}{\partial \psi_i} = -\sum_{j=1}^{J_i} \frac{\beta_i - 1}{y_{ij} - \psi_i} + \frac{\beta_i}{\theta_i^{\beta_i}} \sum_{j=1}^{J_i} (y_{ij} - \psi_i)^{\beta_i - 1}.$$

The second derivative of the logarithm of this conditional density is

$$\frac{\partial^2 \log[\psi_i \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\eta}; \mathbf{y}]}{\partial \psi_i^2} = -\sum_{j=1}^{J_i} \frac{\beta_i - 1}{(y_{ij} - \psi_i)^2} - \frac{\beta_i(\beta_i - 1)}{\theta_i^{\beta_i}} \sum_{j=1}^{J_i} (y_{ij} - \psi_i)^{(\beta_i - 1)(\beta_i - 2)},$$

which is negative if $\beta_i > 1$. Part (i) holds. For (ii), we have

$$\frac{\partial^2 \log[\beta_i \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\eta}; \mathbf{y}]}{\partial \beta_i^2} = -\frac{J_i + \eta_1}{\beta_i^2} - \sum_{j=1}^{J_i} \left(\frac{y_{ij} - \psi_i}{\theta_i} \right)^{\beta_i} \log^2\left(\frac{y_{ij} - \psi_i}{\theta_i} \right) < 0,$$

for any $\beta_i > 0$. For (iii), we get

$$\frac{\partial^2 \log[\xi_1 \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \xi_2, \boldsymbol{\eta}; \mathbf{y}]}{\partial \xi_1^2} = -\frac{a_1 - 1}{\xi_1^2} - I \frac{\partial^2}{\partial \xi_1^2} \log \Gamma(\xi_1).$$

It follows from Formula (1.46) of Bowman and Shenton (1988) that the second derivative of the log gamma function has the expression

$$\frac{\partial^2 \log \Gamma(\alpha)}{\partial \alpha^2} = \frac{1}{\alpha} + \frac{1}{2\alpha^2} + \frac{2\pi}{\alpha} \int_0^\infty \frac{\sqrt{t}\, e^{2\pi\sqrt{t}}}{(\alpha^2 + t)(e^{2\pi\sqrt{t}} - 1)^2} dt, \text{ for } \alpha > 0.$$

We have then,

$$\frac{\partial^2 \log[\xi_1 \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \xi_2, \boldsymbol{\eta}; \mathbf{y}]}{\partial \xi_1^2} = -\frac{I}{\xi_1} - \frac{I + 2a_1 - 2}{2\xi_1^2} - \frac{2\pi}{\xi_1} \int_0^\infty \frac{\sqrt{t}\, e^{2\pi\sqrt{t}}}{(\xi_1^2 + t)(e^{2\pi\sqrt{t}} - 1)^2} dt,$$

which is negative if $I + 2a_1 - 2 \ge 0$. Finally, for (iv),

$$\frac{\partial^2 \log[\eta_1 \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}, \eta_2; \mathbf{y}]}{\partial \eta_1^2} = -\frac{c_1 - 1}{\eta_1^2} - I \frac{\partial^2}{\partial \eta_1^2} \log \Gamma(\eta_1)$$

$$= -\frac{I}{\eta_1} - \frac{I + 2c_1 - 2}{2\eta_1^2} - \frac{2\pi}{\eta_1} \int_0^\infty \frac{\sqrt{t}\, e^{2\pi\sqrt{t}}}{(\eta_1^2 + t)(e^{2\pi\sqrt{t}} - 1)^2} dt,$$

which is negative if $I + 2c_1 - 2 \ge 0$. □

References

Ahrens, J., & Dieter, U. (1974). Computer methods for sampling from gamma, beta, poisson and binomial distributions. *Computing, 12,* 223–246.

Ahrens, J., & Dieter, U. (1982). Generating gamma variates by a modified rejection technique. *Communications of the Association for Computing Machinery (ACM), 25,* 47–54.

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., & Balikrishnan, 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. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 340–357.

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.

Berger, J.O., & Sun, D. (1993). Bayesian analysis for the poly-Weibull distribution. *Journal of the American Statistical Association, 88*, 1412–1418.

Bowman, K.O., & Shenton, L.R. (1988). *Properties of estimators for the gamma distribution.* New York, NY: Marcel Dekker.

Busemeyer, J.R., & Townsend, J.T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*, 432–459.

Cave, C., & Squire, L. (1992). Intact and long-lasting repetition priming in amnesia. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 509–520.

Dey, D., Ghosh, S., & Mallick, B. (2000). *Generalized linear models: A Bayesian perspective.* New York, NY: Marcel Dekker.

Dolan, C., van der Maas, H., & Mollenaar, C. M. (in press). A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavioral Research Methods, Instruments, and Computers.*

Dzhafarov, E.N. (1992). The structure of simple reaction time to step-function signals. *Journal of Mathematical Psychology, 36*, 235–268.

Fox, J.P., & Glas, C.A.W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika, 66*, 271–288.

Gelfand, A., & Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*, 398–409.

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (1995). *Bayesian data analysis.* London, U.K.: Chapman and Hall.

Gelman, A., & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science, 7*, 457–511.

Gilks, W. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics, 41*, 337–348.

Hasher, L., & Zacks, R.T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388.

Heathcote, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavioral Research Methods, Instruments, and Computers, 28*, 427–445.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (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.

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

Hsu, Y.F. (1999). *Two studies on simple reaction times: I. On the psychophysics of the generalized Pieron's law. II. On estimating minimum detection times using the time estimation paradigm.* Unpublished doctoral dissertation. Irvine, CA: University of California.

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic fromm intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Jiang, Y., Rouder, J.N., & Speckman, P.L. (in press). A note on the sampling properties of the Vincentizing (quantile averaging) procedure. *Journal of Mathematical Psychology.*

Johnson, N.L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions, Volume 2* (2nd ed.). New York, NY: John Wiley & Wiley.

Kreft, I.G.G., & de Leeuew, J. (1998). *Introducing multilevel modeling.* London, U.K.: Sage.

Link, S.W. (1975). The relative judgement theory of two choice response time. *Journal of Mathematical Psychology, 12*, 114–135.

Logan, G.D. (1988). Towards an instance theory of automization. *Psychological 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, NY: 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.

Mitchell, D. & Brown, A. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*, 213–222.

Nelder, J.A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7*, 308–313.

Petrusic, W.M., Baranski, J.V., & Kennedy, R. (1999). Similarity comparisons with remembered and perceived magni-
    tudes: Memory psychophysics and fundamental measurement. *Memory & Cognition, 26*, 1041–1055.
Plourde, C.E., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian
    Journal of Experimental Psychology, 51*, 181–194.
Posner, M.I. (1978). *Chronometric explorations of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, F.P. (1992). *Numerical recipes in C: The art of scientific
    computing* (2nd ed.). 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., & Rouder, J.N. (1998). Modeling response times for decisions between two choices. *Psychological Science,
    9*, 347–356.
Ratcliff, R., & Rouder, J. (2000). A diffusion model analysis of letter masking. *Journal of Experimental Psychology:
    Human Perception and Performance, 26*, 127–140.
Rouder, J.N. (1996). Premature sampling in random walks. *Journal of Mathematical Psychology, 40*, 287–296.
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 and Performance, 26*, 359–378.
Rouder, J.N. (2001). Testing evidence accrual models by manipulating stimulus onset. *Journal of Mathematical Psychol-
    ogy, 45*, 334–354.
Schnieder, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search,
    and attention. *Psychological Review, 84*, 1–66.
Smith, P.L. (1995). Multiple detector models of visual simple reaction time. *Psychological Review, 102*, 567–593.
Smith, R.L., & Naylor, J.C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-
    parameter Weibull distribution. *Applied Statistician, 36*, 358–369.
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 and
    Performance, 22*, 461–479.
Spieler, D.H., Balota, D.A., & Faust, M.E. (2000). Levels of selective attention revealed through analyses of response
    time distributions. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 506–526.
Sternberg, S. (1966). High-speed scanning in human memory. *Scuebce, 153*, 652–654.
Tanner, M.A. (1993). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood
    functions*. Berlin, Germany: Springer.
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.
Townsend, J.T., & Ashby, F.G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, U.K.:
    Cambridge University Press.
Treisman, A.M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.
Ulrich, R., & Miller, J.O. (1993). Information processing models generating lognormally distributed reaction times.
    Journal of Mathematical Psychology, 37, 513–525.
Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin and Review, 7*, 424–465.
Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual
    matching. *Psychonomic Bulletin and Review, 7*, 208–256.
Vincent, S.B. (1912). The function of vibrissae in the behavior of the white rat. *Behavioral Monographs, 1*, No. 5.
Wang, X.H., Bradlow, E.T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied
    Psychological Measurement, 26*, 109–128.
Wild, P., & Gilks, W.R. (1993). Adaptive rejection sampling from log-concave density functions. *Applied Statistics, 42*,
    701–708.
Wixted, J.T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental
    Psychology: Learning, Memory, & Cognition, 19*, 1024–1039.