PSYCHOMETRIKA—VOL. 80, NO. 2, 491–513
JUNE 2015
DOI: 10.1007/S11336-013-9396-3

THE LOGNORMAL RACE: A COGNITIVE-PROCESS MODEL OF CHOICE AND LATENCY WITH DESIRABLE PSYCHOMETRIC PROPERTIES

JEFFREY N. ROUDER AND JORDAN M. PROVINCE

UNIVERSITY OF MISSOURI

RICHARD D. MOREY

UNIVERSITY OF GRONINGEN

PABLO GOMEZ

DEPAUL UNIVERSITY

ANDREW HEATHCOTE

UNIVERSITY OF NEWCASTLE

We present a cognitive process model of response choice and response time performance data that has excellent psychometric properties and may be used in a wide variety of contexts. In the model there is an accumulator associated with each response option. These accumulators have bounds, and the first accumulator to reach its bound determines the response time and response choice. The times at which accumulator reaches its bound is assumed to be lognormally distributed, hence the model is race or minima process among lognormal variables. A key property of the model is that it is relatively straightforward to place a wide variety of models on the logarithm of these finishing times including linear models, structural equation models, autoregressive models, growth-curve models, etc. Consequently, the model has excellent statistical and psychometric properties and can be used in a wide range of contexts, from laboratory experiments to high-stakes testing, to assess performance. We provide a Bayesian hierarchical analysis of the model, and illustrate its flexibility with an application in testing and one in lexical decision making, a reading skill.

Key words: cognitive psychometrics, response-times models, race models.

Cognitive modelers and psychometricians analyze human performance data to measure individual abilities, to measure the effects of covariates, and to assess latent structure in data. Yet, there is relatively little overlap between psychometric models and cognitive-process models. This lack of overlap is in many ways understandable. Cognitive process modelers and psychometricians not only have different goals, they come from different traditions.

For cognitive modelers, the emphasis is on using structure in data to draw inferences about the processes and mental representations underlying perception, memory, attention, decision-making, and information-processing. These researchers specify complex, detailed nonlinear models that incorporate theoretical insights on the acquisition, storage, and processing of mental information. Consider, for example, McClelland and Rumelhart' *interactive activation model* (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), which describes how people recognize letters and words by making specific representation and processing assumptions. The goal of cognitive modelers is to uncover the true structures underlying mental life, and consequently, models are benchmarked by their ability to fit fine-grained features in the data. The

Requests for reprints should be sent to Jeffrey N. Rouder, University of Missouri, Columbia, MO, USA. E-mail: rouderj@missouri.edu

disadvantages of this approach are as follows: (1) It may be difficult to analyze complicated nonlinear cognitive process models, and in many cases parameter estimation proceeds with the help of heuristics or trial-and-error. (2) It is almost always difficult to assess a constellation of issues related to model fit including adequately characterizing the complexity of the models, characterizing the data patterns that could not be fit (Roberts & Pashler, 2000), and understanding the robustness to misspecification. (3) It is often not clear how process models may be extended to account for covariates or used in a psychometric setting to understand variation across people.

Psychometricians, on the other hand, place an emphasis on the usefulness of models. Good models may not fit the data in the finest detail, but they are statistically tractable, convenient to analyze, incorporate covariates of interest, and may be used to understand variation in latent traits and abilities across people. Psychometricians retain these desirable properties by using models that cognitive modelers may consider too simple. Consider, for example, the Rasch IRT model in which the key structural specification is that performance reflects the difference between participant ability and item difficulty (Rasch, 1960). Because this specification is simple, the model retains desirable statistical properties including ease of analysis and the ability to understand misspecification. Moreover, because the base IRT model is so simple, it is straightforward to integrate sophisticated models on item and participant effects including factor models and structural equation models (Skrondal & Rabe-Hesketh, 2004; Embretson, 1991).

In our view, cognitive modelers and psychometricians make different tradeoffs: Psychometricians gain the ability to more easily measure latent abilities by stipulating simple psychological processes and representations; cognitive modelers gain the ability to understand the details of these processes and representations by stipulating detailed and complex models, but at the expense of statistical tractability. There is, however, a fertile middle ground of models, sometimes called cognitive psychometrics (Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002), where the goal is to develop psychological process models with good psychometric and statistical properties.

In this paper, we develop a model for the joint distribution of response choice and response time (RT) that we believe occupies this middle ground. The model captures a processing commitment: evidence for a response grows gradually in time in separate, independent, racing accumulators. The response is made when one of these accumulators first wins the race, that is, it first attains sufficient evidence. We adopt perhaps the simplest version of this race-between-accumulators theme, and show how it retains desirable properties of useful psychometric models. Not only is the model statistically tractable, it is straightforward to place sophisticated model components, such as autoregressive components, on model parameters.

In the next section, we provide a brief and selective overview of the literature on accounting for response time and response choice to provide context for our approach. We then present a *base model* that accounts for response choice and response time for a single participant observing a number of replicate trials. Following that, the analysis of the base model with Bayesian methods is discussed. The base model is easily extended for real-world applications, and we provide two examples. The first is a Rasch-inspired testing application where for each item the participant may choose among n options. The second is an application in lexical decision where participants decide if a presented letter string is a word or a nonword. The model describes how latent accumulation for word and nonword information varies across experimental conditions and people.

1. Modeling Response Time and Response Choice Jointly

Making choices always produces two dependent measures, the choice selected and the time it takes to make the selection (response time or RT). Given that experimenting and testing is often carried out via computer, RT is often readily available. The most popular approach in experimental psychology and psychometrics is to model one dependent measure, either RT or choice, and to ignore the other. Although ignoring one of the measures may seem naive, it is often natural in real-world settings. Analyzing one measure seems appropriate when there is variation in only one, or where the joint variation is so thoroughly correlated that all the information is effectively contained in one measure. Modeling both RT and accuracy jointly entails additional complications over modeling one of these measure, and in some contexts the gain may be marginal. Whether jointly modeling RT and accuracy is worthwhile will vary from situation to situation, and may depend to some large extent on the researchers' goals and desired interpretations of parameters.

One conventional psychometric approach to modeling RT and choice jointly is to first place separate regression models on choice and RT (e.g., Thissen, 1983; van Breukelen, 2005; van der Linden, Scrams, & Schnipke, 1999; van der Linden, 2007, 2009). It is common to place a two-parameter lognormal model on RT although other specifications, including the Weibull and a more general proportional hazard model are precedented. Conventional 1PL, 2PL, or 3PL IRT models may be placed on choice, where choice is coded as either a correct response or an incorrect response. Choice and RT may then be linked by allowing RT to be a covariate in the IRT model on choice or by allowing choice to be a covariate in the lognormal model on RT, or both (e.g., van Breukelen, 2005). Similarly, choice and RT may be linked through shared latent parameters. For example, the person parameters in the IRT model and the lognormal model may be modeled as coming from a bivariate structure with a shared variance component. These approaches provides for tractability and flexibility in understanding how people and item covariates affect choice and RT. Moreover, because the modeling is within conventional IRT and mixed linear model frameworks, analysis follows a familiar path. Nonetheless, the models are statistical and are neither informed by nor inform theories of cognitive process.

Perhaps the most popular cognitive-process alternative is based on Ratcliff and colleagues' diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998). In Ratclff's diffusion model, latent evidence evolves gradually in time and is modeled as a diffusion process which terminates on one of two latent bounds. The model is applied to tasks where participants make dichotomous decisions in response to stimuli. Each of the two responses is associated with one of the two bounds. The decision occurs when the process is absorbed, and the response is the one associated with the absorbing bound. The key advantage of this approach is the separation of the latent bound from the rate of information accumulation. This separation provides for a natural account of speed-vs.-accuracy tradeoffs in decision making.

Recently, there has been a trend toward adapting the diffusion model so that it may be used in psychometric settings: Wagenmakers, van der Maas, and Grasman (2007) provide a simplified version that has somewhat increased statistical tractability over Ratlciff's (1978) original version. Tuerlickx and De Boek (2005) show that the 2PL IRT model on choice data may be interpreted as a diffusion model with a simple relationship between item parameters and diffusion model parameters. van der Maas, Molenaar, Maris, Kievit, and Borsboom (2001), Vandekerckhove, Verheyen, and Tuerlinckx (2010), and Vandekerckhove, Tuerlinckx, and Lee (2011) show how latent-variable submodels may be added to diffusion model parameters to account for variation among people, items, and conditions. This process-based approach is laudable and exciting because it is both substantively and psychometrically attractive.

Though the diffusion model approach has a number of benefits, it also has a few distinct disadvantages. First, it is not straightforward to extend the diffusion model to an arbitrary number of choices. The analysis of absorption of a multidimensional diffusion process is complex though some results are possible under certain conditions (see Diederich & Busemeyer, 2003). Second, the approach is not altogether convenient because the likelihood function is comprised of infinite sums of products of sinusoids and exponentials. The current approaches to analysis, such as those in Vanderckhove et al. (2010, 2011), are based on default Metropolis Hastings (MH) sampling in JAGS and Winbugs. Because the likelihood is not easily analyzed and because analysis relies on several MH steps, it is not clear whether more complex models, such as autoregressive models, may be placed on parameters.

In this paper, we propose a simpler and more tractable cognitive process: a race between accumulators (Audley & Pike, 1965; Smith, 2000). The version we develop and extend comes from Heathcote and Love (2012), who modeled the finishing times of each accumulator as a lognormal. The lognormal race is not as feature rich as the diffusion process approach or more comprehensive accumulator models (e.g., Brown & Heathcote, 2008) that, for example, provide parameters selectively accounting for speed-vs.-accuracy tradeoffs. The main advantages of the approach are two-fold: First, the approach generalizes seamlessly to any number of response choices including large numbers of choices or even just one choice. Second, the analysis of the process is far more tractable than the diffusion model and more comprehensive accumulator models.. This increase in tractability translates into pragmatic advantages. Here, we show not only how IRT and cell-means models may be placed on accumulation rates, but how autoregressive components may be placed on them as well.

2. Specification of the Base Lognormal Race Model

Consider a single participant who responds on J identical experimental trials. On each trial, the participant is shown a stimulus or is given an item. In responses, the participant endorses one of n possible responses choices. Let x_j and t_j denote the response choice and response time for the jth trial, respectively, with $x_j = 1, \dots, n$; $t_j > 0$; and $j = 1, \dots, J$. In race models, evidence accumulates linearly in competing accumulators, with each accumulator associated to one of the n response. At some point, the evidence in the accumulator crosses the accumulator's bound, and that point of time is referred to as the finishing time. Let y_{ij} denote this finishing time for the ith accumulator on the jth trial. The first accumulator that finishes determines the choice and the response time:

$$x_j = m \quad \Longleftrightarrow \quad y_{mj} = \min_i (y_{ij}), \tag{1}$$

and response time t_j is

$$t_j = \psi + \min_i (y_{ij}), \tag{2}$$

where ψ denotes an irreducible minimum shift that reflects the contribution of nondecision processes such as encoding the stimulus and executing the response (see Dzhafarov, 1992; Luce, 1986; Ratcliff, 1978). The presence of shift parameter ψ complicates analysis to some degree, and its inclusion is necessitated by the fact that empirical RT distributions are often substantially shifted away from zero (Rouder, 2005). Indeed, we show here in our subsequent empirical application that these shifts are not only present, but substantial in size in a word-identification task.

We model each finishing time y_{ij} as log normally distributed:

$$y_{ij} \overset{\text{ind}}{\sim} \text{Lognormal}\big(\mu_i, \sigma_i^2\big).$$

This parametric choice is suitable for response times because the lognormal has support on positive reals, is unimodal, exhibits a characteristic positive skew, and has a soft left-tail rise. Parameters μ and σ^2 serve as scale and shape parameters, respectively. In this version of the race model, finishing times are assumed to be stochastically independent. A generalization of this assumption seems possible (see Heathcote & Love, 2012), but the full ramifications of which have not been systematically explored.

The joint density function of choice m at time t is

$$f(m,t) = g\big(t-\psi;\mu_m,\sigma_m^2\big)\prod_{i\neq m}\big(1-G\big(t-\psi;\mu_i,\sigma_i^2\big)\big), \tag{3}$$

where g and G are the density and cumulative distribution functions, respectively, of the two-parameter lognormal distribution. It is straightforward to expand this joint density to provide the likelihood function across many independent observations.

We have specified different shape parameters σ_i^2 for each accumulator. Although this specification of different shapes is certainly general, it may affect the interpretability of results in practice. In most applications, researchers will construct a contrast between these finishing time distributions. If the homogeneity assumption $\sigma_i^2 = \sigma^2$ is made, then contrasts between finishing time distributions is captured completely by contrasts of μ_i. We describe the analysis of the base model without this homogeneity assumption for generality. In the subsequent applications, we make the homogeneity assumption to increase the interpretability of μ_i. In some cases, especially where researchers wish to account for the data in finer detail, the heterogeneity of shapes may be warranted (see Heathcote & Love, 2012).

3. Bayesian Analysis of the Base Model

3.1. *Prior Specification*

The base lognormal race model is straightforward to analyze in the Bayesian framework. The parameters are ψ, μ_i, σ_i^2. Weakly informative semiconjugate priors may be specified as follows:

$$\pi(\psi) \propto 1, \tag{4}$$

$$\sigma_i^2 \overset{\text{ind}}{\sim} \text{Inverse-Gamma}(a_i, b_i), \tag{5}$$

$$\mu_i \overset{\text{ind}}{\sim} \text{Normal}(c_i, d_i), \tag{6}$$

where the inverse gamma distribution has pdf

$$f\big(\sigma^2;a,b\big) = \frac{b^a}{\Gamma(a)(\sigma^2)^{a+1}}\exp\left(-\frac{b}{\sigma^2}\right), \quad a,b,\sigma^2 > 0.$$

Perhaps the most important specification is that of conditionally independent priors on μ_i. In applications, we place substantively meaningful models as priors on μ_i.

3.2. *Analysis*

Analysis proceeds by MCMC integration of the joint posterior distribution by means of the Gibbs sampler (Gelfand & Smith, 1990). The Gibbs sampler works by alternately sampling from full conditional posterior distributions. These conditional posteriors are conveniently expressed when conditioned on latent finishing times y_{ij}. Let $z_{ij} = \log y_{ij}$, and let $\bar{z}_i = J^{-1} \sum_j z_{ij}$. Conditional posterior distribution for μ_i is

$$\mu_i \mid \cdots \sim \text{Normal}\big(v_i\big[J\bar{z}_i/\sigma_i^2 + c_i/d_i\big], v_i\big) \tag{7}$$

where

$$v_i = \big(J/\sigma_i^2 + 1/d_i\big)^{-1}.$$

The conditional posterior distribution of σ_i^2 is

$$\sigma_i^2 \,\Big|\, \cdots \sim \text{Inverse-Gamma}\left(a_i + J/2, b_i + \sum_j (z_{ij} - \mu_i)^2/2\right).$$

The conditional posterior density of ψ is

$$f(\psi \mid \cdots) \propto \begin{cases} \prod_{ij} g(y_{ij} - \psi; \mu_i, \sigma_i^2), & \psi < \min_{ij}(y_{ij}), \\ 0, & \text{otherwise}, \end{cases}$$

where g is the density of the two-parameter lognormal distribution. The conditional posterior distribution of the finishing times z_{ij} depends on whether the ith accumulator finished first on the jth trial. If it did, then z_{ij} is a direct transform of the observed response time:

$$(z_{ij} \mid x_j = i, \ldots) = \log(t_j - \psi).$$

If the ith accumulator did not win the race, then the relevant information is that $z_{ij} > \log(t_j - \psi)$. The posterior of z_{ij} is simply the prior conditional on $z_{ij} > \log(t_j - \psi)$, e.g.,

$$(z_{ij} \mid x_j \neq i, \ldots) \sim \text{Truncated Normal}\big(\mu_i, \sigma_i^2, \log(t_j - \psi)\big),$$

where the arguments in order are the mean, variance of the corresponding untruncated normal, and the lower bound of support.

To implement the Gibbs sampler, it is necessary to be able to sample from each of these conditional posterior distributions. Such sampling is straightforward for μ_i, σ_i^2 and z_{ij} as the corresponding posterior distributions are well known. The conditional posterior for ψ, however, does not correspond to a known distribution. Because ψ is bounded above, care must be exercised that mass near the bound is sampled. We have found in previous work that adaptive rejection sampling (Wild & Gilks, 1993) works well for sampling from bounded distributions (see Rouder, Sun, Speckman, Lu, & Zhou, 2003), but the approach requires log-concave distributions, i.e., $\frac{\partial^2}{\partial\psi^2} \log f(\psi \mid \cdots) < 0$. Unfortunately, the full conditional posterior distribution for ψ is not generally log-concave. Consequently, we use a Metropolis step with a symmetric normal random walk to sample ψ. This approach seemingly worked well in the presented applications: mixing was acceptable and it was easy to tune the proposal for a desired rejection rate. More details are provided in the subsequent applications.

3.3. DIC Model Comparison

It is natural to perform model comparison with deviance-information-criteria (DIC, Spiegelhalter, Best, Carlin, & van der Linde, 2002) when analysis proceeds through MCMC. DIC is a Bayesian analog to AIC in that goodness of fit statistics are penalized by model complexity. It shares a philosophical commitment with AIC, namely the penalty captures predictive out-of-sample accuracy. Models with covariates that do not increase out-of-sample predictive accuracy are penalized more heavily. Although there are important and compelling critiques of DIC (see, for example, Dawid's and Sahu's discussion in response to Spiegelhalter et al. article in the same issue), we recommend it here over other methods such as model comparison by Bayes factors (Kass & Raftery, 1995) because of tractability. DIC is relatively easy to calculate in MCMC chains for the lognormal race; the marginalization that comprises Bayes factors appears difficult. For some comparisons, however, Laplace approximation (Kass, 1993) or Savage-Dickey density ratio (Gelfand & Smith, 1990; Morey, Rouder, Pratte, & Speckman 2011) computations of Bayes factors may be feasible. DIC for a model is computed as follows:

$$\mathrm{DIC} = \bar{D} + p_D,$$

where $\bar{D}$ is the expected value of deviance with respect to the posterior distribution of the parameters and p_D is known as the effective number of parameters and serves as a penalty term. The computation of $\bar{D}$ is natural in the MCMC chain, and the deviance may be computed with the straightforward extension of (3), provided that the observations are independent. The effective number of parameters takes into account the role of priors in constraining parameters, and models with more constrained priors will be less penalized than models with diffuse priors. The effective number of parameters is

$$p_D = \bar{D} - D(\bar{\theta}),$$

where $D(\bar{\theta})$ is the deviance evaluated at the posterior means of the parameters. DIC across models may be directly compared conditional on the same data, and the model with the lowest DIC value is most preferred. Additional discussion of computing DIC is provided by Gelman, Carlin, Stern, and Rubin (2004), and the foundational assumptions underlying the above expressions is provided by Spiegelhalter et al. (2002). We use DIC here to test the necessity of the parameter ψ, but it could also be used, for instance, to test the assumption of the homogeneity of shapes across accumulators.

4. The Lognormal Race Model Affords Enhanced Statistical Convenience

To extend the base model for real-world applications, more complex models are placed on the log finishing times, z_{ij}. Examples of these more complex models include linear models to account for people, item, and condition effects, autoregressive models to account for trial-by-trial variation, or factor models to capture dimension-reducing structure. We refer to models on z_{ij} as *backend models*. Successful Bayesian analysis comprises the following two computational tasks: (I) the sampling of posterior log finishing times conditional on all parameters including the backend model parameters, and (II) the sampling the posterior backend parameters conditional on log finishing times. The first task, sampling posterior log finishing times conditional on the backend model is straightforward as these quantities are known for the winning accumulator and may be sampled from a truncated normal for the remaining ones. The second task, sampling the parameters of the backend model conditional on log finishing times, is the target of recent developments in Bayesian analysis. These recent developments in analyzing linear models, autoregressive models, and latent-variable models for normally distributed data may be leveraged, and model builders will find standard texts, say Gelman et al. (2004), Jackman (2009), and Lee (2007), to be informative. Included in these texts are efficient sampling algorithms applicable to large dimensional models.

We highlight the statistical convenience of the lognormal race with two applications. The first is for testing, and we develop a simple Rasch-like IRT model that accounts for response time and response choice in a multiple-choice testing setting. The second is for experimental psychology, and we develop an AR(1) model with covariates to account for condition and participant effects while allowing trial-to-trial sequential dependencies. These applications highlight the core attraction of the lognormal race—the ability to construct flexible and diverse models with a psychologically-motivated theoretical orientation.

5. Application I: A Rasch-Like IRT Model for Testing

In this section, we provide a lognormal-race model that is similar in spirit to a Rasch model (1PL) for an n-choice context. For concreteness, we assume that each subject responds to each item; the generalization to other designs is straightforward.

5.1. Model Specification

Let $i = 1, \dots, n$, $j = 1, \dots, J$, and $k = 1, \dots, K$ index choices, items, and participants, respectively. Let x_{jk}, and t_{jk} denote the response ($x_{jk} = 1, \dots, n$) and response time for the kth subject's answer to the jth item. We model these variables as a lognormal race. Let z_{ijk} be the log of finishing times. The race model for this application is

$$x_{jk} = m \quad \Longleftrightarrow \quad z_{mjk} = \min_i (z_{ijk}),$$

$$t_{jk} = \psi_k + \exp\Big(\min_i (z_{ijk})\Big),$$

where ψ_k is a participant-specific shift that models nondecisional latencies such as the time to read the question and execute the motor response.

The key structural part of the model is the decomposition of log finishing times into participant and item effects. We place an additive decomposition that is motivated by the Rasch 1PL IRT model. Let δ_{ij} be an indicator variable that is 1 if i is the correct response for item j and 0 otherwise. The structural component is

$$z_{ijk} = \alpha_{ij} - \delta_{ij}\beta_k + \epsilon_{ijk}, \tag{8}$$

where $\alpha_{\cdot j}$ are a set of difficulties for the jth item, β_k is the kth participant's ability, and ϵ_{ijk} are independent and identically normally distributed zero-centered noise terms with variance σ^2. In (8), increases in β decrease the scale of the finishing time for the correct-item accumulator, resulting in better accuracy and quicker overall finishing times.

The model may be conveniently expressed in matrix notation. Let $\boldsymbol{z}_k$ be the vector of log finishing times for the kth participant, $\boldsymbol{z}_k = (z_{11k}, \dots, z_{n1k}, z_{12k}, \dots, z_{nJk})'$. Let $\boldsymbol{z}$ be the vector of log finishing times across all participants, $\boldsymbol{z} = (\boldsymbol{z}_1', \dots, \boldsymbol{z}_K')'$. Parameters are likewise expressed as vectors: $\boldsymbol{\alpha} = (\alpha_{11}, \dots, \alpha_{n1}, \alpha_{1J}, \dots, \alpha_{nJ})'$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)'$. Let $\boldsymbol{X}_\alpha$, $\boldsymbol{X}_\beta$ be design matrices that map the respective parameters into $\boldsymbol{z}$ such that

$$\boldsymbol{z} = \boldsymbol{X}_\alpha \boldsymbol{\alpha} - \boldsymbol{X}_\beta \boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{9}$$

where $\boldsymbol{\epsilon}$ is the collection of zero-centered noise terms.

In most IRT treatments, participants are modeled as random effects and items are modeled as fixed effects. The random effect specification is achieved with a hierarchical prior on $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} \mid \sigma_\beta^2 \sim \text{Normal}\big(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}\big),$$

$$\sigma_\beta^2 \sim \text{Inverse-Gamma}(a, b).$$

Note that participant parameters are zero-centered, and this specification is all that is needed for identifiability.

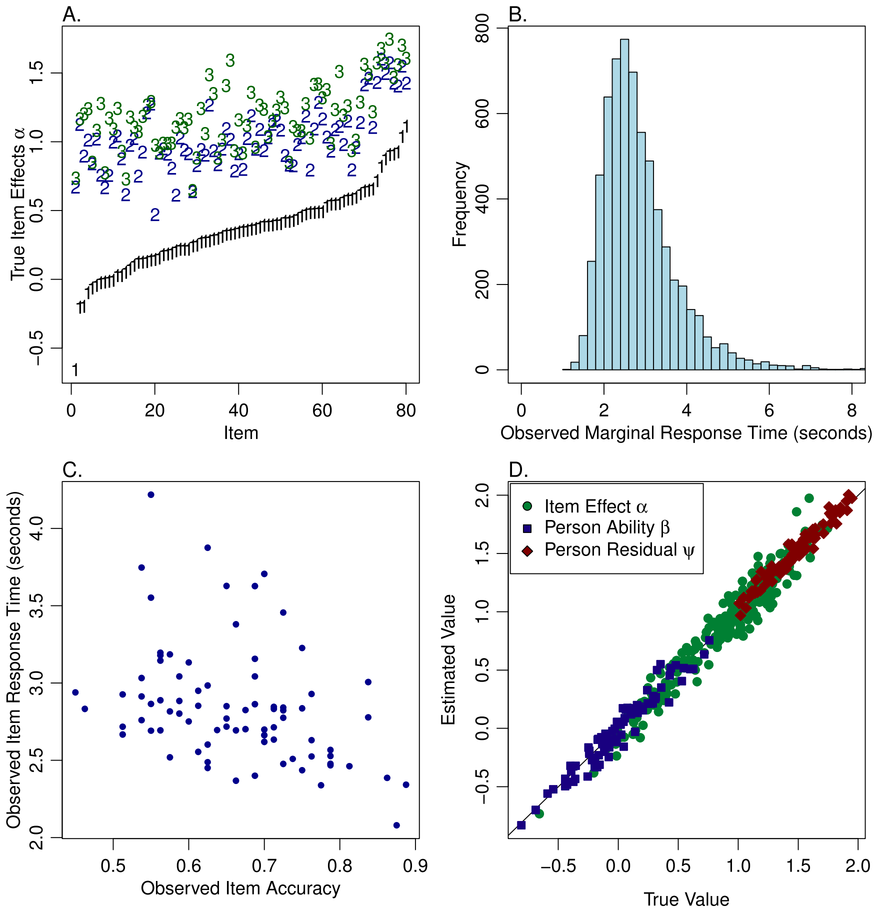

FIGURE 1.
Simulation of a three-choice lognormal race model. (**A**) True item effects. The points labeled "1," "2," and "3" are the true effect for the correct option, the second most likely option, and the third most likely option, respectively. (**B**) Marginal response time distribution for simulated data. (**C**) Accuracy and mean RT for the 80 items from the simulated data. (**D**) Posterior-mean parameter estimates as a function of true value for effects ($\boldsymbol{\theta}$) and shift ($\boldsymbol{\psi}$).

5.2. Simulation and Model Development

Simulated Data Set To assess whether the model can be analyzed efficiently, we performed a small simulation in which 80 people observed 80 items, where each item had three response options. True item effect values for the 80 items are shown in Figure 1A. The points labeled "1" are for the correct response option; those labeled "2" and "3" are for the alternatives, with option "2" being more likely than option "3." These true values were chosen such that easier items had correct options with smaller scales than did the incorrect options. True shift values were $\psi_k \overset{\text{iid}}{\sim} \text{Unif}(1, 2)$, where the parameters are the endpoints of the distribution. Figure 1B show the marginal response time distribution for the simulated data. As can be seen, the location, scale and shape is characteristic of RT data. Figure 1C shows how simulated performance varies by item. Plotted are distributions of item mean accuracy ($J^{-1}\sum_j x_{ij}$) and item mean RT ($J^{-1}\sum_j t_{ij}$), and these plots show that variability in item parameters lead to a reasonable range of performance.

Priors and Posteriors Priors are needed for parameters $\boldsymbol{\alpha}$, σ^2, and $\boldsymbol{\psi}$. Independent normal priors were placed on α_{ij} with a mean of c and a variance of d. To avoid autocorrelation in MCMC sampling, it is useful to treat $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as a single vector, $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ (see Roberts & Sahu, 1997). With this specification, $\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X} = (\boldsymbol{X}_\alpha, \boldsymbol{X}_\beta)$, and the prior on $\boldsymbol{\theta}$ is a conditionally independent, multivariate normal. The conditional posterior on $\boldsymbol{\theta}$ is derived in Gelman et al. (2004):

$$\boldsymbol{\theta} \mid \cdots \sim \text{Normal}(\boldsymbol{\phi}\boldsymbol{q}, \boldsymbol{\phi}),$$

where

$$\boldsymbol{\phi} = \left(\boldsymbol{X}'\boldsymbol{X}/\sigma^2 + \boldsymbol{B}\right)^{-1},$$
$$\boldsymbol{q} = \boldsymbol{X}'\boldsymbol{z}/\sigma^2 + \boldsymbol{B}\boldsymbol{\mu}_0.$$

Vector $\boldsymbol{\mu}_0$ is the prior mean, $\boldsymbol{\mu}_0 = (c\mathbf{1}_{nJ}, \mathbf{0}_K)'$ and matrix $\boldsymbol{B}$ is the prior precision, $\boldsymbol{B} = \text{diag}(\mathbf{1}_{nJ}/d, \mathbf{1}_K/\sigma_\beta^2)$. In application, we set $c = 0$ and $d = 5$, which is a diffuse prior on item effects as these effects describe the log of the scales of finishing times.

We retain the inverse-gamma prior on σ^2. The conditional posteriors for both variance parameters are:

$$\sigma^2 \mid \cdots \sim \text{Inverse-Gamma}\left(a_1 + nJK/2, b_1 + (\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\theta})'(\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\theta})/2\right),$$
$$\sigma_\beta^2 \mid \cdots \sim \text{Inverse-Gamma}\left(a + K/2, b + \boldsymbol{\beta}'\boldsymbol{\beta}/2\right).$$

Prior settings were $a = a_1 = 1$ and $b = b_1 = 0.1$, which are weakly informative. For σ^2, they capture the a priori information that RTs have a noticeably long but not excessive right tail. Reasonable values of σ^2 range from 0.01 to 1.0. With these settings, the prior on σ^2 has mass through this range. The prior settings also struck us as appropriate for σ_β^2, though this was more a matter of coincidence. The β parameter is the log of a scaling parameter, and variances of 1.0 imply an almost three-fold increase in scale across people.

We retain the previous flat prior for ψ, and the conditional posterior is

$$f(\psi_k \mid \cdots) \propto \begin{cases} \prod_{ij} g(\exp(z_{ijk}) - \psi_k; \mu_{ijk}, \sigma^2), & \psi_k < \min_{ij}(\exp(z_{ijk})), \\ 0, & \text{otherwise,} \end{cases}$$

where $\mu_{ijk} = \alpha_{ij} - \delta_{ij}\beta_k$ is the expected value of z_{ijk}.

5.3. Results

MCMC analysis proceeds with Gibbs steps on all parameters except for $\boldsymbol{\psi}$. Each ψ_k is sampled with a separate Metropolis step with normally distributed candidates. We tuned the candidate distribution for each participant with a single standard-deviation parameter, which was set to 0.08 sec for all 80 people. The resulting rejection rates varied from 0.32 to 0.64. We performed 5 runs of 5,000 MCMC iterations with starting values varying randomly in a priori determined reasonable ranges, with the exception of individuals' shift parameters. These parameters need to be less than the minimum RT for the individual, and we simply started chains with shifts that were 0.2 sec below this minimum. MCMC chains for selected parameters are shown Figure 2. Log-rate parameters, $\boldsymbol{\theta}$, show the best mixing, and there is little autocorrelation. There is some autocorrelation in σ^2 and in $\boldsymbol{\psi}$. This autocorrelation reflects the structural properties of the log-normal where shift and shape are only weakly constrained. Detectable autocorrelation, while undesirable, lasted for less 100 iterations in the example, and is thus manageable. Its presence, however, serves as a caveat that analysts must be aware of the potential for poorly mixing chains

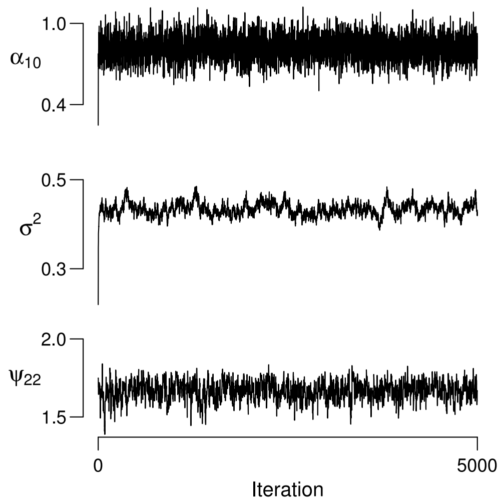

FIGURE 2.
MCMC for selected parameters. Chains are plotted for 5000 iterations from a single run, and the degree of autocorrelation is manageable. Shown are the difficulty for a selected item difficulty, the variance σ^2, and the shift for a selected participants. Chains for other items and participants look similar.

with this model. We set a burn-in period of 500 iterations, and convergence of the chains with burn-in excluded were assessed with Gelman et al.'s (2004) $\hat{R}$ and effective number-of-samples (N_{eff}) statistics. The maximum $\hat{R}$ across all parameters was 1.007, which is well below 1.1, a recommended benchmark for acceptable convergence. The median N_{eff} across all five chains was 6403, which means that targeted posterior quantities are estimated to reasonable high precision. Figure 1D shows parameter estimates (posterior means across all five runs) as a function a true values for log-scale and shift parameters. Here, parameter recovery is excellent for such a richly parameterized model even in such a relatively small set of data.

In this model, the interpretation of person ability, β_k, is the same as in the Rasch model with the exception that ability in the lognormal race model does not have a fixed standard deviation of 1.0. We also estimated the Rasch-model abilities from the dichotomous accuracy data (responses were scored as correct or incorrect) with the `ltm` package in R with the constraint that all item discriminabilities were set to 1.0. The upper triangle of Table 1 shows the Pearson correlations between lognormal-race true abilities, lognormal race estimated abilities, IRT estimated abilities, and overall participant accuracy. As can be seen, all four vectors are highly correlated. For this design and model, overall participant accuracy is highly diagnostic and drives ability estimates. The interpretation of item effects is more detailed in the lognormal race than the dichotomous IRT model because there are separate item effects per response. Figure 1A shows how these item-by-response effects may be organized. Importantly, there is no unique correlate to item difficulty as there are several parameters per item. One ad-hoc approach to constructing item difficulty is to contrast the item effect for the correct response with the minimum across all others. We constructed such a measure, and the lower triangle of Table 1 shows the correlations between true lognormal-race item difficulty, estimated lognormal-race item difficulty, IRT estimated difficulty, and overall item accuracy. As before, overall item accuracy is highly diagnostic and drives difficulty estimates.

Given these high correlations, one could question the need for the lognormal race or even IRT in this case because consideration of overall accuracies provides almost the same information. The appeal of the models is the interpretation of the parameters, as well as other niceties such as the ability to generalize to complex cases where designs are not balanced, and the inter-

TABLE 1.
Correlations between ground truth and estimates of participant and item effects.

	LNR truth	LNR estimate	IRT estimate	Observed accuracy
LNR truth	–	0.96	0.94	0.94
LNR estimate	0.88	–	0.98	0.98
IRT estimate	0.86	0.96	–	0.996
Observed accuracy	0.86	0.96	0.996	–

Upper and lower triangle are correlations for participant and item effects, respectively.

pretation of their parameters. Moreover, IRT models are typically extended to account for item discriminability and possible factor structures among abilities, and we see no reason that the lognormal race model could not be extended similarly. This particular lognormal race IRT model predicts strict constraints among RT and accuracy, and if these constraints do indeed hold in testing data, then IRT researchers may gain confidence that IRT models are not ignoring important variation in RT when modeling accuracy alone. Consequently, it is useful to establish whether the race model is a good description is of extant testing data, though it remains outside the scope of this article.

6. Application II. Assessing Word and Nonword Effects on Finishing Times in Lexical Decision

One of the main advantages of the lognormal race is that it may be applied broadly across domains. Here, we illustrate how it may be used in an experimental setting to understanding the processes underlying reading ability. We analyze a previously unpublished data set from the Word-Nerds-And-Perception Laboratory of DePaul University. Participants performed a lexical-decision task in which they decided if five-letter strings formed valid English-language words. Some strings were valid words, such as *TREE*, while others were not, such as *PREE*. These invalid strings are called *nonwords*. Participants simply classified strings as either words or nonwords by pressing corresponding response buttons.

The data were collected to test predictions of competing process accounts of letter encoding in reading (Davis, 2010; Gomez, Ratcliff, & Perea, 2008; Whitney, Bertrand, & Grainger, 2011). Our goal here is more limited: we seek to characterize structure in the data using the lognormal race. In the data set, some nonwords were formed by replacing a letter in a word. For example, *PREE* is formed by substituting a "P" for the "T" in *TREE*. There are 5 different substitution conditions: either the first, second, third, fourth, or fifth letter in a valid word may be substituted to form a nonword. Other nonwords were formed by transposing letters in valid words. For example, the nonword *JUGDE* is formed by transposing the "D" and "G" in *JUDGE*. There are 7 different transposition conditions depending on which two letters were transposed. (The conditions were transpositions of positions 1 & 2, 1 & 3, 2 & 3, 2 & 4, 3 & 4, 3 & 5, and 4 & 5.) Words were also coded by their frequency of occurrence.[1] There were three categories: words were either considered low, medium, or high in frequency. The 5 nonword substitution conditions, the 7 nonword transposition conditions, and the three word conditions comprise a

[1] Word frequency is the how often a word occurs in natural written discourse. It is simply the frequency of occurrence in a large corpora such as a large collection newspaper and magazine articles (Kucera & Francis, 1967). For instance, *AJAR* occurs less than once per million words of text, *ECHO* occurs 34 times per million words of text, and *CITY* occurs over 200 times per million words of text.

total of 15 conditions on the strings in the data set. One goal of the modeling is to understand these condition effects.

One underappreciated concern in modeling is the correlation of responses times across trials. In most conventional analyses, trial order is not modeled, and there is an implicit assumption that performance on a trial is independent of performance on previous trials. Although it is readily acknowledged that such sequential effects do indeed occur (Bertelson, 1961; Luce, 1986), they are rarely modeled (for notable exceptions, see Falmagne, Cohen, & Dwivedi, 1975; Link, 1992; Craigmile, Peruggia, & Van Zandt, 2010; Peruggia, Van Zandt, & Chen, 2002). One potential problem with this lack of consideration is that the uncertainty in parameters is understated, which can lead to inflated error rates in statistical decision making. We model these correlations as an AR(1) process, which we believe is appropriate in this context—at least to first approximation—because sequential effects are often short lived (Luce, 1986; Peruggia et al., 2002; cf., Craigmile et al., 2010). It is important to emphasize, however, that our main point here is to show *how* a backend model such as the AR(1) model can be added to the lognormal race model, and not to argue that the AR(1) model is the best possible model for this purpose.

The data set consists of 93 participants each performing 720 trials. On half the trials, a nonword was presented, and this nonword was chosen from each of the 12 nonword conditions with equal frequency. On the remaining half, a word was presented, and this word was chosen from each of the 3 word conditions with equal probability. The pool of word and nonword strings was exceedingly large, encompassing over 3000 strings. Each string was presented just a few times throughout the experiment, and consequently, it is not possible to accurately model individual string effects. In the following development, we model condition and participant effects within a lognormal race. About 1 % of trials were discarded because the responses were outside a prescribed window that was set before data collection.

6.1. *Model Specification*

The model for the lexical decision data set is quite similar to the IRT model, with condition playing the role that item played previously. Let $x_{jk} = 1, 2$ and $t_{jk} > 0$ be response choice and response time for the kth participant on the jth trial ($k = 1, \dots, K$, $j = 1, \dots, J_k$), where $x = 1$ and $x = 2$ denote *word* and *nonword* responses, respectively. Response choice and RT are modeled as a lognormal race between a *word*-response accumulator and a *nonword*-response accumulator with log finishing times z_{ijk}, where $i = 1, 2$ denotes the accumulator.

It may seem odd to think that nonword information can accumulate in the same manner as word information, because nonwords may be conceptualized as the *absence* of a word. Our approach to treating the absence of an entity as analogous to an entity itself is not only common, but compatible with extant findings. For example, some word-like nonword strings, such as *neb*, are responded to more quickly than low-frequency words, such as *ajar*. This indicates that nonwords are not simply the absence of words. Moreover, this approach is used in other domains, and a good example is *novelty detection*, where the observer can quickly orient to novel stimuli. The nearly immediate salience of items or stimuli *never experienced* seems to indicate that novelty is psychologically represented as more than the absence of familiarity. We realize notions of word and nonword accumulations may be controversial (Dufau, Grainger, & Ziegler, 2012). Nonetheless, they are theoretically interpretable and frequently used (e.g., Ratcliff, Gomez, & McKoon, 2004), and thus we adopt them here.

In this application, we implement a first order autoregressive component, AR(1), on the log finishing times for both accumulators. We specify the AR(1) components in univariate and multivariate notation as both are convenient in deriving conditional posteriors. Let z_{1k} and z_{2k} denote the vector of J_k latent log finishing times for the kth participant in the word and nonword accumulators, respectively. Let $\boldsymbol{X}_k$ be a design matrix that maps trials into the 15 experimental conditions, and let $\boldsymbol{\mu}_{1k}$ and $\boldsymbol{\mu}_{2k}$ be vectors of 15 participant-specific condition means for the word and nonword accumulators, respectively. Finally, let $\boldsymbol{u}_{ik} = \boldsymbol{z}_{ik} - \boldsymbol{X}_k \boldsymbol{\mu}_{ik}$ denote a vector of J_k residuals. The following AR(1) process is placed on residuals:

$$u_{ijk} = \rho u_{i,j-1,k} + \epsilon_{ijk} \tag{10}$$

where $\epsilon_{ijk} \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$, $|\rho| < 1$, and $u_{i,0,k} = 0$.

Needed is the joint expression of all u_{ijk}. To derive this expression, we note that the vector of noise terms $\boldsymbol{\epsilon}_{ik}$ may be expressed as $\boldsymbol{\epsilon}_{ik} = \boldsymbol{A}_k \boldsymbol{u}_{ik}$, where

$$\boldsymbol{A}_k = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & \ddots & \ddots & & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & & \ddots & \ddots & 1 & 0 \\ 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}.$$

The joint expression is derived by inverting $\boldsymbol{A}_k$. The inverse of $\boldsymbol{A}_k$, denoted $\boldsymbol{L}_k$ is

$$\boldsymbol{L}_k = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \rho & 1 & \ddots & \ddots & & 0 \\ \rho^2 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \rho^{J_k-2} & & \ddots & \ddots & 1 & 0 \\ \rho^{J_k-1} & \rho^{J_k-2} & \cdots & \rho^2 & \rho & 1 \end{bmatrix}.$$

As $\boldsymbol{u}_{ik} = \boldsymbol{L}_k \boldsymbol{\epsilon}_{ik}$, it follows that the joint distribution of $\boldsymbol{u}_{ik}$ is

$$\boldsymbol{u}_{ik} \sim \text{Normal}(\mathbf{0}, \mathbf{\Sigma_k}),$$

where

$$\boldsymbol{\Sigma}_k = \sigma^2 \boldsymbol{L}_k \boldsymbol{L}_k{}'.$$

With these expressions, the model on log finishing times can be expressed in multivariate notation as

$$\boldsymbol{z}_{1k} \sim \text{Normal}(\boldsymbol{X}_k \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{11}$$

$$\boldsymbol{z}_{2k} \sim \text{Normal}(\boldsymbol{X}_k \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{12}$$

The difference between the previous models and the AR(1) is reflected in the nonzero, off-diagonal elements of the covariance matrix $\boldsymbol{\Sigma}_k$. The main computational step is inverting this matrix, but inversion is no more burdensome than the previous models because the inverse of $\boldsymbol{\Sigma}_k = \boldsymbol{A}'\boldsymbol{A}/\sigma^2$, and its determinant is σ^{2J_k}.

For this model, the estimation of parameters in $\boldsymbol{\mu}_{1k}$ and in $\boldsymbol{\mu}_{2k}$ tend to be unstable because of small sample sizes. For over 10 % of the parameters, there is not one observation; for over 80 % , there are less than 20 observations. To stabilize estimates, we used the following additive models as priors. Let $\ell = 1, \dots, 15$ index the condition. Then

$$\mu_{1k\ell} \mid \alpha_\ell, \beta_k, \delta_1 \sim \text{Normal}(\alpha_\ell + \beta_k, \delta_1), \tag{13}$$

$$\mu_{2k\ell} \mid \gamma_\ell, \tau_k, \delta_2 \sim \text{Normal}(\gamma_\ell + \tau_k, \delta_2), \tag{14}$$

where α_ℓ and γ_ℓ are the main effects of the ℓth condition on word and nonword finishing times, respectively; β_k and τ_k are the main effects of the kth participant on word and nonword finishing times, respectively; and δ_1 and δ_2 are the variance of the residual interaction terms.

Priors are needed for $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and $\boldsymbol{\tau}$. As in the testing application, we place zero-centered priors on participant effect vectors $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$. Priors on $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are slightly more complex. They are motivated by the fact that people are relatively accurate in the their lexical decisions. Consequently, finishing times for correct responses, that is word responses to words and nonword responses to non words, must be faster than finishing times for incorrect responses. In the current design, there are 12 nonword conditions and 3 word conditions that are crossed with the nonword and word accumulators. We used 4 hyperpriors to describe the crossing. Let $\boldsymbol{\mu}_\alpha = (\mu_{\alpha_w}, \mu_{\alpha_{\bar{w}}})$ be a 2-item vector of condition means for word and nonword conditions, respectively, let $\boldsymbol{\mu}_\gamma$ be defined analogously, and let $\boldsymbol{X}_0$ be the 15×2 design matrix that maps each condition into whether it is a word or nonword condition. Then the priors on condition means are

$$\boldsymbol{\alpha} \mid \boldsymbol{\mu}_\alpha, \sigma_\alpha^2 \sim \text{Normal}\big(\boldsymbol{X}_0 \boldsymbol{\mu}_\alpha, \sigma_\alpha^2 \boldsymbol{I}\big),$$

$$\boldsymbol{\gamma} \mid \boldsymbol{\mu}_\gamma, \sigma_\gamma^2 \sim \text{Normal}\big(\boldsymbol{X}_0 \boldsymbol{\mu}_\gamma, \sigma_\gamma^2 \boldsymbol{I}\big).$$

The effect of this additional level in the hierarchical prior is to change the direction of shrinkage. In this model, the crossings of word and nonword conditions with word and nonword accumulators serve as unique points of shrinkage, and model different shrinkage points for correct word responses, incorrect word responses, correct nonword responses, and incorrect nonword responses. We place a flat prior on $\boldsymbol{\psi}$, Inverse-Gamma priors on all variance parameters, and diffuse normally-distributed priors on the $\boldsymbol{\mu}_\alpha$ and $\boldsymbol{\mu}_\gamma$ parameters. The prior on ρ was uniformly distributed from -1 to 1.

6.2. Analysis

The inclusion of the AR(1) component helps illustrate the two tasks needed to integrate complex backend models into the lognormal race framework. The first task is sampling the posterior values of log finishing times conditional on all parameters. Let $\bar{i}$ and i denote the winning and losing accumulator. For the winning accumulator, this value is known, and $z_{\bar{i}jk} = \log(t_{jk} - \psi_k)$. For the losing accumulator,

$$z_{ijk} \mid \cdots \sim \text{Truncated-Normal}\big(\text{E}[z_{ijk}] + \rho u_{i,j-1,k}, \sigma^2, z_{\bar{i}jk}\big),$$

where $u_{i,j-1,k}$ is defined in (10), and E is the expected value of the parameter with respect to its prior. The vector of expectation values is simply the product of the relevant design matrices and parameter vectors. Sampling truncated normals is not computationally demanding, though sampling all latent finishing times is time consuming when there are a large number of trials and response alternatives. The second task is sampling the vectors of parameters and hyperparameters given the log finishing times. For all parameters except $\boldsymbol{\psi}$, the conditional posterior distributions remain conjugate; that is, they are multivariate normal and inverse-gamma, and for the dimensions of this application the sampling is convenient. The conditional posterior distribution of ρ is

$$\rho \mid \cdots \sim \text{Truncated-Normal}_{(-1,1)}\big(m'/v', \sigma^2/v'\big),$$

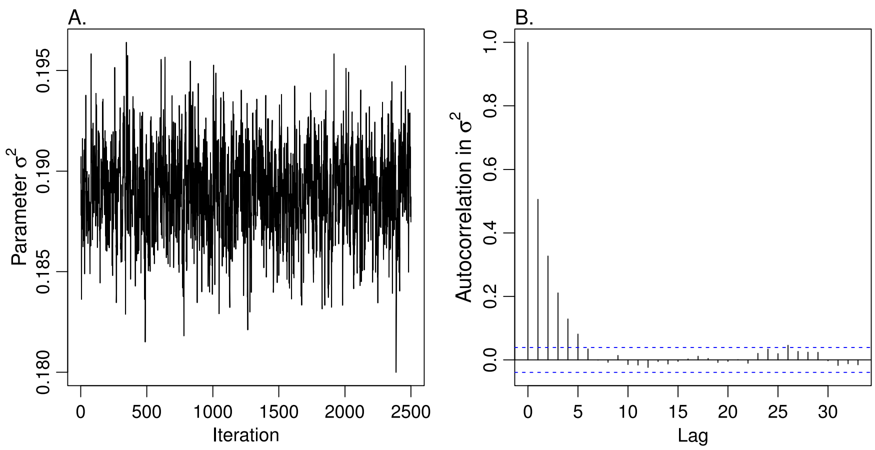

FIGURE 3.
The chain of values for parameter σ^2, the slowest converging parameter. The *plots* show outputs after thinning by a factor of 20. (**A**) MCMC outputs. (**B**) Autocorrelation function of outputs.

where

$$m' = \sum_{i=1}^{2} \sum_{k=1}^{K} \sum_{j=1}^{J_k-1} u_{ijk} u_{i,j+1,k},$$

$$v' = \sum_{i=1}^{2} \sum_{k=1}^{K} \sum_{j=1}^{J_k-1} u_{ijk}^2.$$

Parameters $\boldsymbol{\psi}$ are sampled with Metropolis–Hastings steps, and we have found it quite easy to tune these steps in practice to obtain reasonable rejection rates.

6.3. *Results*

Mixing We performed 5 independent runs of 10,000 MCMC iterations with a burn-in period of 500 iterations. Convergence of the chains with burn-in excluded were assessed in the same manner as in the above IRT application using Gelman et al.'s (2004) $\hat{R}$ and effective number-of-samples (N_{eff}) statistics. The maximum $\hat{R}$ across all parameters was 1.004 indicating acceptable convergence to target posterior distributions. The median N_{eff} across all five chains was 10400 indicating that posterior quantities are estimated to reasonable high precision. Figure 3 shows the resulting MCMC outputs and autocorrelation function for σ^2, the slowest converging parameter, across the 5 runs after thinning by a factor of 20. All other parameters showed equally well or better mixing.

Model Fit The current model was built to be psychometrically useful rather than fit the data precisely. Nonetheless, the model should fit at least coarsely if the parameters are to be interpreted. Figure 4A is a plot of observed mean response time for correct responses as a function of the predicted value. Each point is from a particular participant in a particular condition. There is some misfit for slow RTs, but the overall pattern is reasonable for a measurement model. Figure 4B is the same for the probability of a word response, and each point is from a particular participant-by-condition combination. There is some degree of misfit for nonword conditions. This slight mistfit is not surprising as these conditions have smaller numbers of observations (although there were 360 words and nonwords, the nonwords were divided into 12 conditions while

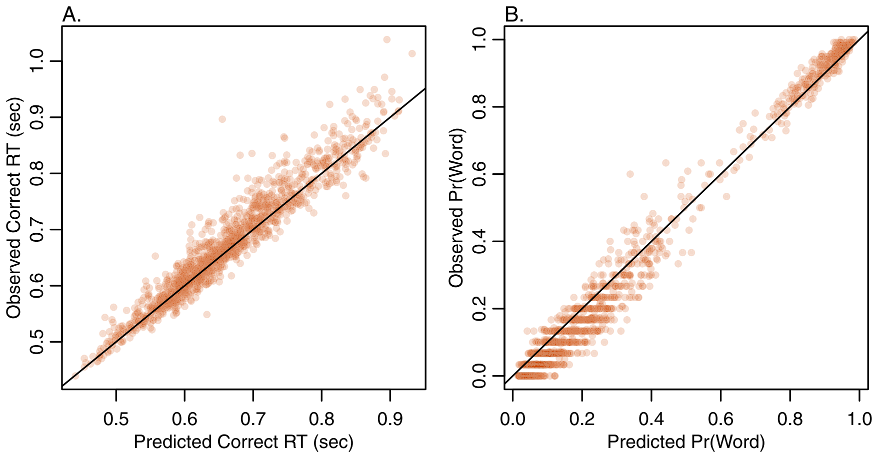

FIGURE 4.
Plots of observed data as a function of predicted values. *Points* denote values from the 1395 condition-by-participant combinations. (**A**) Mean response time for correct responses. (**B**) Probability of a "word" response.

the words were divided into only 3 conditions). With such small numbers at the participant-by-condition level, there is a noticeable effect of the prior pulling very small probabilities away from zero. This type of shrinkage is expected in hierarchical models, and strikes us as reasonable.

Parameter Estimates Figure 5 shows the condition effects on accumulators. The basic pattern is one of mirror symmetry where conditions that result in faster finishing times on one accumulator also result in slower finishing times on the other. The nonword accumulator seemingly reacts to the conditions in an orderly and structured manner, indicating that the notion of nonword accumulation passes a minimal test. Finishing times for nonword errors in word conditions and for word errors in nonword conditions are similar. There are, however, two subtle differences between word and nonword finishing times: (1) The series labeled "a" shows word-accumulator finishing times for word items, and these are faster than the corresponding nonword finishing times for nonword items (series labeled "e" and "f"). (2) The effect of condition is greater for word accumulators than for nonword accumulators (as measured on the log scale): series "a" is more varied than series "d;" series "b" is more varied than series "e;" series "c" is more varied than series "f." Note that this effect of greater variation holds for both word and nonwords. In summary, although nonword accumulation is similar to word accumulation, the later seems more sensitive to manipulations than the former.

Figure 6 shows the parameter estimates of shift ($\boldsymbol{\psi}$) and autocorrelation (ρ). The thick segments in Figure 6A are the posterior interquartile range for each shift estimate, and the thin segments are the corresponding 95 % posterior credible intervals. As can be seen, these shift parameters are well away from zero and vary substantially from one another. Therefore, the use of individual shift parameters seems judicious as a general rule even though it does come with some computational cost. Figure 6B shows the posterior distribution of ρ, and the degree of trial-to-trial correlation is small, about 0.07. Nonetheless, it is reliably positive. Given the small size, models without sequential dependencies will provide for reasonable estimation of condition-effect parameters, the parameters of main interest.

We assessed whether shifts were necessary using DIC as discussed previously. Two models, one with unshifted lognormal distributions, and the above model were compared, though in neither did we include sequential dependencies. The likelihood function is easily found from (3), and may be evaluated on each iteration of the chain. The DIC for the model without shifts is

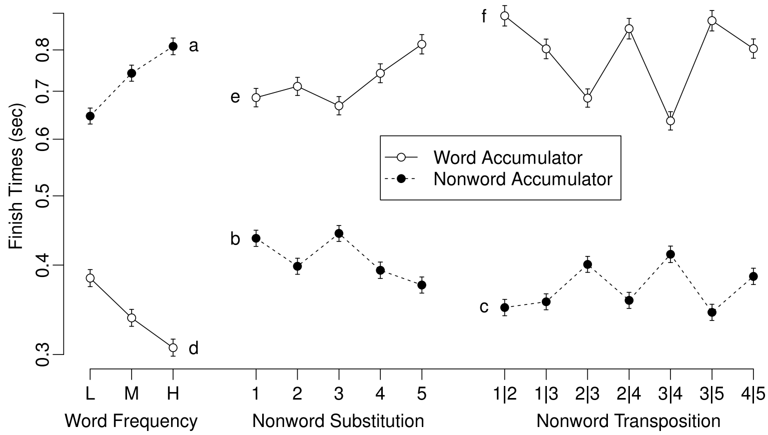

FIGURE 5.
Finishing times for word and nonword accumulators as a function of condition.

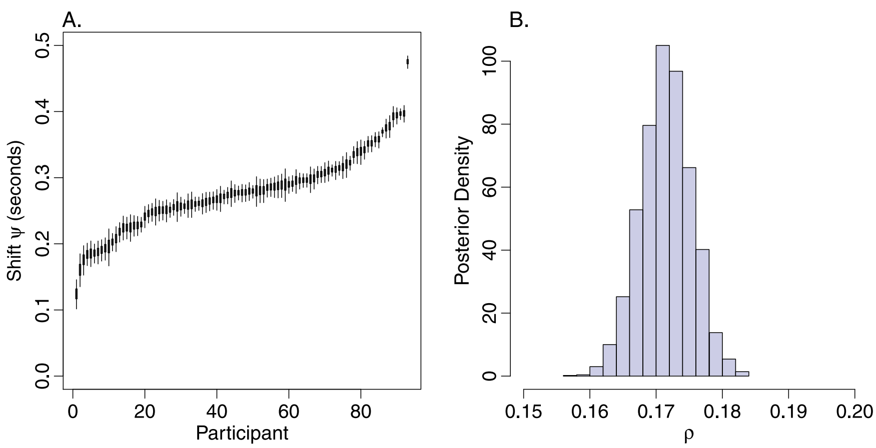

FIGURE 6.
(**A**) Posterior interquartile range (*thick segment*) and 95 % credible intervals (*thin segment*) for each participant's shift parameter ψ. (**B**) Posterior distribution of ρ.

−68,897 and that for the model with shifts is −77,374; hence, the model with shifts is preferred, and the difference of about 8,500 is overwhelming.

7. General Discussion

In this paper, we have provided a simple process model for response time and choice that has desirable statistical properties. In this model, decisions are viewed as arising from a race between competing evidence-accumulation processes. Finishing times of accumulators are log-normal random variables; due to this assumption, the model is tractable when analyzed in the Bayesian framework. It is straightforward to place sophisticated model components, such as autoregressive models, on the finishing times, and it seems equally straightforward to place other sophisticated components such as factor and structural equation models. We view the model as retaining the core strengths of both the psychometric and cognitive-process model traditions: It is plausible from a process perspective, while being useful from a psychometric perspective. Moreover, the model is applicable in a wide range of human performance settings, providing both a theoretical and practical means of uniting analysis in experimental psychology and high-stakes testing.

7.1. Relationship to Other Race Models

The lognormal race is a member of a larger class of racing-accumulator models. In these models, each response has an associated accumulator, and information slowly accumulates. In general, this information is considered stochastic, that is, it has moment-to-moment variation and may be modeled as white noise (McGill, 1963). A decision occurs the first time this noisy information in the accumulators exceeds its bound (Audley & Pike, 1965; Smith, 2000). In application, the rate of information growth and the bound are free parameters estimated from the data. One of the main advantages of the racing accumulator framework is neural plausibility, and the notion of neurons as accumulators has been supported by single-cell neurophysiological recordings. Firing rates in targeted neurons slowly increase with a rate that reflects stimulus strength variables (Hanes & Schall, 1996; Huk & Shadlen, 2005; Roitman & Shadlen, 2002; Ratcliff, Thapar, & McKoon, 2003). This relationship between stimulus strength and firing rates can be seen in several brain regions across a variety of stimuli and tasks (see Gold & Shadlen, 2007; Schall, 2003, for reviews). Not surprisingly, racing accumulator models form the decision-making process in popular neural network approaches (Anderson & Lebiere, 1998; McClelland, 1993; Usher & McClelland, 2001).

The lognormal race follows from Grice (1968), who provided an alternative specification for the racing-accumulator class of models. He assumed there was no moment-to-moment variation in the information in the accumulators. The information accumulated *deterministically* and linearly until it reached a bound. Let ν be the rate of information gain and η be the bound. Then the finishing time, y, is simply η/ν. Grice specified that the bound varied normally from trial-to-trial, and the resulting distribution of y, though incomplete, has been termed the reciprobit distribution (Reddi & Carpenter, 2003). Brown and Heathcote (2008) assumed that both the rate of information accumulation and the bound varied from trial-to-trial. Heathcote and Love (2012) assumed that both ν and η were distributed as lognormals, with $\eta \sim \text{Lognormal}(\mu_\eta, \sigma_\eta^2)$ and $\nu \sim \text{Lognormal}(\mu_\nu, \sigma_\nu^2)$. The distribution of finishing time is $y \sim \text{Lognormal}(\mu_\eta - \mu_\nu, \sigma_\eta^2 + \sigma_\nu^2)$. From this expression, it is clear that with the lognormal distributional assumptions, bounds and accumulation rates cannot be disentangled. Without loss of generality, the bounds may be set to 1.0. This specification results in the base model presented here: $y \sim \text{Lognormal}(\mu, \sigma^2)$ where $\mu = -\mu_\nu$ and $\sigma^2 = \sigma_\nu^2$.

The lognormal race model lacks identifiable decision bounds. The nonidentifiability follows from the lognormal parametric assumption, and with it only the ratio of bounds-to-accumulation-rate is identifiable. This inability to identify bounds separate from accumulation rate is a theoretical disadvantage, at least from a conventional view. We think, however, that it represents an appropriately cautious and humble position. Other race models that distinguish between accumulation rates from decision bounds do so with recourse to specific parametric assumptions, and the estimates of these quantities directly reflects the parametric assumptions. Even the separate estimation of bounds and drift rates in the diffusion model rests on a latent assumption that cannot be tested directly, namely the stationarity of the diffusion process. We would prefer that measurement of these quantities not be conditional on such arbitrary assumptions. As a consequence, we give up the ability to separate bounds from rates. The lognormal parametric specification may be profitably viewed as an elegant way of avoiding fine conceptual distinctions that may not be informed by data.

7.2. *Highly Accurate Responses*

In many areas of experimental psychology, responses are highly accurate and the focus is solely on speed. An example is subliminal priming where the aim is to understand how a weak stimulus affects subsequent processing, and the main paradigm is to see how the presentation of a weak prime speeds the subsequent responses to related materials (e.g., Greenwald, Draine, & Abrams, 1996). In these cases, however, trials still admit multiple responses, even though the correct one is chosen almost always. One advantage of the lognormal race is that it is well suited for these paradigms. In the limit of perfect performance, the response times are lognormally distributed, and this distributional model has been advocated both on practical and theoretical grounds (Ulrich & Miller, 1993). For highly accurate conditions, the analyst must be aware that the latent finishing times on the incorrect accumulators will largely reflect prior assumptions because there are few error responses to inform estimation of these parameters. Nonetheless, the ability to account for highly accurate data is a strong advantage, especially when compared to models that have separate parameters for bounds and accumulation rates. In these models, such as the diffusion model, it is imperative to have some conditions that are far from ceiling to locate the bounds. For the lognormal race, in contrast, such conditions are not necessary, and the model works well even for experiments where there is high accuracy in all conditions.

7.3. *The Need for a Shift Parameter*

We have explicitly included an individual specific shift parameter ψ_k. The inclusion of a shift parameter needs some justification because it is relatively novel in the psychometric literature. Traditionally, psychometric researchers have implemented two-parameter unshifted skewed distributions for RT that do not have a shift parameter (e.g., Thissen, 1983; van Breukelen, 2005; van der Linden et al., 1999). This choice is in contrast to cognitive process modelers who routinely include shift parameters (Luce, 1986). Perhaps the reason psychometricians exclude the shift is convenience—the two-parameter (unshifted) lognormal is easily analyzed with the GLM family while the three-parameter (shifted) model is not. Hence, analysis of mixed two-parameter models may be carried out through familiar means while analysis of the three-parameter model requires additional development. In our Bayesian development, the inclusion of the shift parameter necessitates the inclusion of a Metropolis step, and affects mixing under certain circumstances that are discussed subsequently. Nonetheless, we believe that the empirical data patterns necessitate the inclusion of shift, and models that do not contain a shift parameter will be needlessly misspecified in many domains. Rouder (2005), in response to van Breukelen's two-parameter lognormal model (van Breukelen, 2005) noted that minimum response times were often shifted well away from zero and that this shift varied across people. Indeed, this trend holds here in the lexical-decision data set examined here. Not accounting for shifts will lead to a dramatic overestimation of scale and an equally dramatic underestimation of shape.

7.4. *Caveats*

There are two practical difficulties that arise when analyzing lognormal race models: the potential for poor mixing and the potential for undue influence of the priors.

Poor mixing may occur in the MCMC chain for the shape (σ^2) and shift (ψ) parameters. This poor mixing results from the three-parameter lognormal parametric specification. In this distribution, the left-tail becomes increasingly soft or shallow as the shape becomes more symmetric (that is, as σ^2 decreases). For small true values of σ^2, shift and shape are weakly identified, and this weak identifiability leads to poor mixing. Hence, while the three parameter lognormal is a good choice overall, it becomes more suspect as RT distributions become less skewed. If the observed data does not have noticeable skew then researchers may consider fixing the ψ to zero or abandoning the lognormal parametric specification entirely. Fortunately, RT distributions are typically skewed (Luce, 1986).

A second issue is the potential for undue influence of the priors. This issue is germane for low-probability responses. Consider a lexical decision experiment where all the words are common and easy-to-read (e.g. *dog*, *run*, *bread*) and all the nonwords are clearly not words (e.g., *xqre*, *zvbt*). In this case, we expect almost all words and nonwords to be identified as such, and very few errors to occur. In analysis, however, we estimate finishing times for error responses even though they are rare. In this case, the correct accumulator estimates will largely reflect the data rather than the prior as correct responses are numerous. The difficulty is with the error accumulator estimates, which may largely reflect prior assumptions when there are few if any responses. Hence, researchers need to be aware of the sample size per response option in assessing the influence of the prior and interpreting parameter estimates for error accumulators. In the lexical decision application, we used a hierarchical prior that allows sharing of information across people-by-condition combinations. This approach is suitable even for small effective sample sizes at the people-by-condition level, though researchers should keep in mind that these estimates reflect the structure of the hierarchical prior.

Acknowledgements

This research is supported by National Science Foundation Grants SES-1024080 and BCS-1240359 (JNR) and Australian Research Council Professorial Fellowship (AH).

References

Anderson, J.R., & Lebiere, C. (1998). *The atomic component of thought*. Mahwah: Lawrence Erlbaum Associates.

Audley, R.J., & Pike, A.R. (1965). Some alternative stochastic models of choice. *British Journal of Mathematical and Statistical Psychology*, *18*, 207–225.

Bertelson, P. (1961). Sequential redundancy and speed in a serial two choice responding task. *Quarterly Journal of Experimental Psychology*, *13*, 290–292.

Brown, S.D., & Heathcote, A. (2008). The simplest complete model of choice reaction time: linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Craigmile, P.F., Peruggia, M., & Van Zandt, T. (2010). Hierarchical Bayes models for response time. *Psychometrika*, *75*, 613–632.

Davis, C. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*(3), 713.

Diederich, A., & Busemeyer, J.R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time and simple response time. *Journal of Mathematical Psychology*, *47*, 304–322.

Dufau, S., Grainger, J., & Ziegler, J.C. (2012). How to say "no" to a nonword: a leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1117–1128.

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

Embretson, S.E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, *56*, 495–516.

Falmagne, J.C., Cohen, S.P., & Dwivedi, A. (1975). Two choice reactions as an ordered memory scanning process. In P.M.A. Rabbit & S. Dornic (Eds.), *Attention and performance V* (pp. 296–344). New York: Academic Press.

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. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman and Hall

Gold, J.I., & Shadlen, M.N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: a model of letter position coding. *Psychological Review*, *115*(3), 577.

Greenwald, A.G., Draine, S.C., & Abrams, R.L. (1996). Three cognitive markers of unconscious semantic activation. *Science*, *273*(5282), 1699–1702.

Grice, G.R. (1968). Stimulus intensity and response evocation. *Psychological Review*, *75*, 359–373.

Hanes, D.P., & Schall, J.D. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427–430.

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Cognitive Science*, *3*, 292. http://www.frontiersin.org/Cognitive_Science/10.3389/fpsyg.2012.0029.

Huk, A., & Shadlen, M.N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience*, *25*, 10420–10436.

Jackman, S. (2009). *Bayesian analysis for the social sciences*. Chichester: Wiley.

Kass, R.E. (1993). Bayes factors in practice. *The Statistician*, *42*, 551–560.

Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

Lee, S.-Y. (2007). *Structural equation modelling: a Bayesian approach*. New York: Wiley.

Link, S.W. (1992). *Wave theory of difference and similarity*. Hillsdale: Erlbaum.

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

McClelland, J.L. (1993). Toward a theory of information processing in graded, random, and interactive networks. In D.E. Meyer & S. Kornblum (Eds.), *Attention & performance XIV: synergies in experimental psychology, artificial intelligence and cognitive neuroscience*. Cambridge: MIT Press.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*, 375–407.

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

Morey, R.D., Rouder, J.N., Pratte, M.S., & Speckman, P.L. (2011). Using MCMC chain outputs to efficiently estimate Bayes factors. *Journal of Mathematical Psychology*, *55*, 368–378. http://dx.doi.org/10.1016/j.jmp.2011.06.004.

Peruggia, M., Van Zandt, T., & Chen, M. (2002). Was it a car or a cat I saw? An analysis of response times for word recognition. *Case Studies in Bayesian Statistics*, *6*, 319–334.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

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

Ratcliff, R., & McKoon, G.M. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for decisions between two choices. *Psychological Science*, *9*, 347–356.

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception and Psychophysics*, *65*, 523–535.

Ratcliff, R., Gomez, P., & McKoon, G.M. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182.

Reddi, B., & Carpenter, R. (2003). Accuracy, information and response time in a saccadic decision task. *Journal of Neurophysiology*, *90*, 3538–3546.

Riefer, D.M., Knapp, B.R., Batchelder, W.H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*, 184–201.

Roberts, G.O., & Sahu, S.K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society, Series B, Methodological*, *59*, 291–317.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*, 358–367.

Roitman, J.D., & Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*, 9475–9489.

Rouder, J.N. (2005). Are unshifted distributional models appropriate for response time? *Psychometrika*, *70*, 377–381.

Rouder, J.N., Sun, D., Speckman, P.L., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, *68*, 587–604.

Rumelhart, D.E., & McClelland, J.L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60–94.

Schall, J.D. (2003). Neural correlates of decision processes: neural and mental chronometry. *Current Opinion in Neurobiology*, *13*, 182–186.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. Boca Raton: CRC Press.

Smith, P.L. (2000). Stochastic dynamic models of response time and accuracy: a foundational primer. *Journal of Mathematical Psychology*, *44*, 408–463.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, *64*, 583–639.

Thissen, D. (1983). Timed testing: an approach using item response theory. In D.J. Weiss (Ed.), *New horizons in testing: latent trait test theory and computerized adaptive testing* (pp. 179–203). New York: Academic Press.

Tuerlickx, F., & De Boek, P. (2005). Two interpretations of the discrimination parameter. *Psychometrika*, *70*, 629–650.

Ulrich, R., & Miller, J.O. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, *37*, 513–525.

Usher, M., & McClelland, J.L. (2001). On the time course of perceptual choice: the leaky competing accumulator model. *Psychological Review*, *108*, 550–592.

van Breukelen, G.J.P. (2005). Psychometric modeling of response time and accuracy with mixed and conditional regression. *Psychometrika*.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M.D. (2011). Hierarchical diffusion models for two-choice response time. *Psychological Methods*, *16*, 44–62.

Vandekerckhove, J., Verheyen, S., & Tuerlinckx, F. (2010). A cross random effects diffusion model for speeded semantic categorization decisions. *Acta Psychologica*, *133*, 269–282.
van der Linden, W.J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*, 287–308.
van der Linden, W.J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, *14*, 247–272.
van der Linden, W.J., Scrams, D.J., & Schnipke, D.L. (1999). Using response-time constraints to control for differential speededness in computerized adaptive testing. *Applied Psychological Measurement*, *23*, 195–210.
van der Maas, H.L.J., Molenaar, D., Maris, G., Kievit, R.A., & Borsboom, D. (2001). Cognitive psychology meets psychometric theory: on the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*, *118*, 339–356.
Wagenmakers, E.-J., van der Maas, H.L.J., & Grasman, R.P.P.P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*, 3–22.
Whitney, C., Bertrand, D., & Grainger, J. (2011). On coding the position of letters in words . *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, 1–6.
Wild, P., & Gilks, W.R. (1993). Adaptive rejection sampling from log-concave density functions. *Applied Statistics*, *42*, 701–708.

Manuscript Received: 22 AUG 2012
Published Online Date: 13 FEB 2014