Bayesian Analysis of Simulation-based Models

Brandon M. Turner[a,*], Per B. Sederberg[a], James L. McClelland[b]

[a]*Department of Psychology, The Ohio State University*
[b]*Department of Psychology, Stanford University*

Abstract

Recent advancements in Bayesian modeling have allowed for likelihood-free posterior estimation. Such estimation techniques are crucial to the understanding of simulation-based models, whose likelihood functions may be difficult or even impossible to derive. One particular class of simulation-based models that have not yet benefited from the progression of Bayesian methods is the class of neurologically-plausible models of choice response time, in particular the Leaky, Competing Accumulator (LCA) model and the Feed-Forward Inhibition (FFI) model. These models are unique because their architecture was designed to embody actual neuronal properties such as inhibition, leakage, and competition. Currently, these models have not been formally compared by way of principled statistics such as the Bayes factor. Here, we use a recently developed algorithm – the probability density approximation method – to fit these models to empirical data consisting of a classic speed accuracy trade-off manipulation. Using this approach, we find some discrepancies between an assortment of model fit statistics. In some cases, our results provide modest support for the FFI model, whereas in others substantial support is found for the LCA model. However, when aggregating across all four metrics, clear evidence is gained for one model or another in half of our data.

Keywords: Bayes factor, likelihood-free inference, simulation models, neurologically plausible cognitive models, probability density approximation method, Leaky Competing Accumulator model, Feed Forward Inhibition model

*Corresponding Author
Email address: turner.826@gmail.com (Brandon M. Turner)
This work was funded by NIH award (F32GM103288) to B.M.T.

1. Introduction

The goals of cognitive modeling are to understand complex behaviors within a system of mathematically-specified mechanisms or processes, to assess the adequacy of the model in accounting for experimental data, and to obtain an estimate of the model parameters, which carry valuable information about how the model captures the observed behavior for both individuals and groups. Cognitive models are important because they provide a means with which cognitive theories can be explicitly tested and compared with one another.

Perhaps the greatest strength of many cognitive models is paradoxically the model's greatest weakness. Many cognitive models put forth sophisticated mechanisms meant to capture psychologically plausible processes. While these mechanisms are entirely plausible, they often render the cognitive model intractable, or at least difficult to fully analyze in a principled way such as with Bayesian statistics. The difficulties encountered in deriving the full likelihood function have prevented the application of fully Bayesian analyses for many cognitive models, especially those that attempt to capture neurally-plausible mechanisms.

Consider, for example, the Leaky Competing Accumulator (LCA; Usher and McClelland, 2001) model. The LCA model was proposed as a neurologically plausible model for choice response time in a c-alternative task. The model possesses mechanisms that extend other diffusion-type models (e.g., Ratcliff, 1978) by including leakage and competition by means of lateral inhibition. Because the evidence accumulation process used by the LCA model was designed to mimic actual neuronal activation patterns, one critical assumption is that the signal propagated from one accumulator to another can never be negative. This assumption can be implemented by specifying a floor on each accumulator's activation value, such that if the activation of an accumulator in the model becomes negative, it is reset to zero. The LCA model also assumes a competition among response alternatives that depends on the current state of each of the accumulators. Together, these features of the model sufficiently complicate the equations describing the joint distributions of choice and response time such that the likelihood function for the LCA model has not been derived. As a result, all model evaluations to this point have been performed using either a model simplification or least squares estimation (Usher and McClelland, 2001; Tsetsos et al., 2011; Bogacz et al., 2006; Gao et al., 2011; van Ravenzwaaij et al., 2012; Bogacz et al., 2007; Teodorescu and Usher, 2013), which have been shown to produce less accurate parameter estimates relative to techniques such as maximum likelihood or Bayesian estimation (e.g., Myung, 2003; Rouder et al., 2003; Van Zandt, 2000; Turner et al., 2013a).

Recent advances in likelihood-free techniques have allowed for new insights to simulation-based cognitive models (Turner and Van Zandt, 2012; Turner and Sederberg, 2012; Turner et al., 2013a; Turner and Sederberg, 2014; Turner and Van Zandt, 2014). In particular, the probability density approximation (PDA; Turner and Sederberg, 2014) method now allows for fully Bayesian analyses of computational models exclusively by way of simulation. In this article, we illustrate the importance of our method by comparing two neural network models of choice response time that have never been compared using Bayesian techniques due to their computational complexity: the LCA model (Usher and McClelland, 2001) and the Feed-Forward Inhibition (FFI; Shadlen and Newsome, 2001) model.[2] Both models embody neurologically plausible mechanisms such as "leakage", or the passive decay of evidence during a decision, and competition among alternatives through either lateral inhibition (in the LCA model) or feed-forward inhibition (in the FFI model). However, it remains unclear as to which dynamical system best accounts for empirical data, due to the limitations imposed by intractable likelihoods. Specifically, complexity measures that take into account posterior uncertainty and model complexity have yet to be applied. Here, we will compare the models on the basis of an approximation to the Bayes factor. We begin by describing in greater detail our method for fitting the models to data. We then describe how our posterior estimates are converted into a comparison between the models. Finally, we compare the relative merits of the two models by evaluating the models' fit to the data presented in Forstmann et al. (2011), which consisted of 20 subjects in three speed emphasis conditions.

[2]Although Ratcliff and Smith (2004) used the Bayesian information criteria to compare many simulation-based models, they did not obtain proper Bayesian posteriors, which is the endeavor of the current manuscript.

2. Experiment

The data we will use to test the models were presented in Forstmann et al. (2011), and consist of 20 subjects. The experiment used a moving dots task where subjects were asked to decide whether a cloud of semi-randomly moving dots appeared to move to the left or to the right. Subjects indicated their response by pressing one of two spatially compatible buttons with either their left or right index finger. Before each decision trial, subjects were instructed whether to respond quickly (the speed condition), accurately (the accuracy condition), or at their own pace (the neutral condition). Following the trial, subjects were provided feedback about their performance. In the speed and neutral conditions, subjects were told that their responses were too slow whenever they exceeded a RT of 400 and 750 ms, respectively. In the accuracy condition, subjects were told when their responses were incorrect. Each subject completed 840 trials, equally distributed over the three conditions. These data serve as a benchmark for our metric comparison given that we have some experience in analyzing them in a variety of contexts (Turner et al., 2013c; Turner and Sederberg, 2014; Turner et al., 2013b).

3. Likelihood-free Inference

As the reader of this special issue is no doubt aware, there are many advantages of using Bayesian statistics in cognitive modeling. However, the widespread dissemination of Bayesian statistics can largely be attributed to advanced statistical techniques for approximating the posterior distribution (see, e.g., Robert and Casella, 2004; Gelman et al., 2004; ter Braak, 2006; Gilks and Wild, 1992; Gilks et al., 1995), rather than evaluating it precisely. Approximating any posterior distribution depends on efficient evaluation of two functions: (1) the prior distribution for the model parameters, and (2) the likelihood function relating the model parameters to the observed data. For purely statistical models, evaluating these functions is, generally speaking, straightforward. However, for cognitive models who attempt to provide mechanistic explanations for how data manifest, direct evaluation of the likelihood function can be difficult, if not impossible. We refer to these models as "simulation-based" to indicate that explicit equations for the likelihood function are either (1) intensely difficult to practically evaluate (e.g., Myung et al., 2007; Montenegro et al., 2011; Turner et al., 2013a), or (2) have not yet been derived (e.g., Usher and McClelland, 2001; Shadlen and Newsome, 2001). Recently, a suite of algorithms have been developed specifically for analyzing (simulation-based) cognitive models in a fully (hierarchical) Bayesian context (Turner and Sederberg, 2012, 2014; Turner and Van Zandt, 2014). While combinations of these algorithms can be used to effectively evaluate the joint posterior distribution, we require only one algorithm – the probability density approximation (PDA; Turner and Sederberg, 2014) method – to evaluate the models presented in this article.

3.1. The Probability Density Approximation Method

As discussed in Turner and Sederberg (2014), the PDA method is an alternative likelihood-free algorithm that does not require sufficient statistics for the parameters of interest. Turner and Sederberg (2014) demonstrated the utility of their algorithm by verifying that it could be used to accurately estimate the posterior distribution of the parameters of the Linear Ballistic Accumulator (LBA; Brown and Heathcote, 2008) model, which has a tractable likelihood function and is amenable to Bayesian estimation (Turner et al., 2013c; Donkin et al., 2009a,b). In addition, Turner and Sederberg (2014) showed that the PDA method could be used to estimate the parameters of the LCA model in a fully hierarchical Bayesian context.

Although the details of how to apply the PDA method to various data types are explained in detail in Turner and Sederberg (2014), we will reproduce the relevant details for applying the method to data containing both discrete and continuous measures. For ease of exposition, we consider the common case of data consisting of one discrete measurement (e.g., choice) and one continuous measurement (e.g., response time). For the discrete measurements, suppose there are C options, and for the continuous measurements there are an infinite number of possible values. For the observed data from N trials, we denote the continuous measures as $Y = \{Y_1, Y_2, \ldots, Y_N\}$, the discrete measures as $Z = \{Z_1, Z_2, \ldots, Z_N\}$, and the full set of data as $D = \{D_1, D_2, \ldots, D_N\}$. We assume that the ith data pair $D_i = (Y_i, Z_i)$ arise from a model with parameters θ so that $D \sim \text{Model}(\theta)$. We can then write the density under the assumed model, conditional on the parameters θ, as

$$\text{Model}(D_i = \{Y_i, Z_i\}|\theta). \tag{1}$$

For simulation-based models, the density in Equation 1 is generally what cannot be easily evaluated. For these models, we must instead rely on an approximation. In fact, the accuracy of our estimated joint posterior distribution of the parameters θ depends almost entirely on our ability to accurately approximate Equation 1.

To estimate Equation 1, we begin by generating a proposal parameter value θ^*. We then use θ^* to simulate a set of data $X = \{X^{(1)}, \ldots, X^{(C)}\}$, where $X^{(c)}$ is the set of continuous measurements for the cth discrete alternative. In other words, we separate the continuous measures on the basis of the discrete measures. For example, in a two-alternative choice task where choice response time data are collected, we would divide the simulated data into two bins: $X^{(1)}$ could consist of the response times for choice one (e.g., the correct response), and $X^{(2)}$ could consist of the response times for choice two (e.g., the incorrect response). We then introduce a vector containing the set of the number of observations for each alternative, so that $n = \{n^{(1)}, n^{(2)}, \ldots, n^{(C)}\}$ and $J = \sum_{c=1}^{C} n^{(c)}$ (i.e., J denotes the *total* number of model simulations).

For each response time distribution, we construct a proper kernel density estimate (see Turner and Sederberg, 2014, for details) for the simulated probability density function (SPDF) by evaluating

$$f_{n^{(c)}}\left(x|X^{(c)}\right) = \frac{1}{h^{(c)}J} \sum_{j=1}^{n^{(c)}} K\left(\frac{x - X_j^{(c)}}{h^{(c)}}\right), \tag{2}$$

where $K(\cdot)$ is the kernel and $h^{(c)}$ is a smoothing parameter known as the bandwidth. The kernel is usually chosen to be unimodal and symmetric about zero to place a decreasing weight on observations X_j further from the point where the density is being estimated (i.e., at location x). While the kernel can take many forms, in this article we will only consider the Epanechnikov kernel, given by

$$K(x) = \begin{cases} \frac{3}{4}\left(1 - x^2\right) & \text{if } x \in [-1, 1] \\ 0 & \text{if } x \notin [-1, 1] \end{cases}. \tag{3}$$

The accuracy of kernel density function is measured by the mean integrated squared error (MISE), a measure of divergence between a true and an estimated density function. The Epanechnikov kernel was derived on the basis of minimizing the asymptotic MISE, and so it is optimal in a statistical sense (Epanechnikov, 1969; Silverman, 1986). We denote the set of bandwidth parameters $\mathbf{h} = \{h^{(1)}, h^{(2)}, \ldots, h^{(C)}\}$, so that

$$h^{(c)} = 0.9 \min \left(SD\left(X^{(c)}\right), \frac{IQR\left(X^{(c)}\right)}{1.34} \right) \left(n^{(c)}\right)^{-1/5}, \tag{4}$$

where $SD(\cdot)$ denotes the standard deviation, and $IQR(\cdot)$ denotes the interquartile range. This particular choice of the bandwidth is known as Silverman's rule of thumb (Silverman, 1986), and has been shown to make the kernel density estimate more accurate.

Equation 2 is known as a *deffective* probability density function, which means that if integrated for all values of x, it will integrate to the probability of making a particular response choice. In other words, it is scaled to reflect that for any given choice response time pair, other choices could have been made. Using Equation 2 in our calculations is important so that we our model fits simultaneously capture both aspects of our data (i.e., response choice and response time).

Referring back to Equation 1, the likelihood function can be approximated by way of the following equation:

$$\mathcal{L}(\theta|D) = \prod_{i=1}^{N} \text{Model}(D_i|\theta) = \prod_{i=1}^{N} f_{n^{(Z_i)}}\left(Y_i|X^{(Z_i)}\right). \tag{5}$$

With a suitable approximation of the PDF in hand, we have only to combine the approximated likelihood function with the prior distributions to obtain an approximation of the joint posterior distribution for the model parameters θ:

$$\pi(\theta|D) \propto \pi(\theta)\mathcal{L}(\theta|D).$$

As in conventional Markov chain Monte Carlo, the proposal parameter value θ^* is accepted with Metropolis Hastings probability. Namely, on the tth iteration, the current state of the algorithm is at the previous location θ_{t-1}. We set $\theta_t = \theta^*$ with probability

$$\min\left(1, \frac{\pi(\theta^*|D)q(\theta_{t-1}|\theta^*)}{\pi(\theta_{t-1}|D)q(\theta^*|\theta_{t-1})}\right), \tag{6}$$

otherwise we set $\theta_t = \theta_{t-1}$. In Equation 6, $q(\theta^*|\theta)$ is the probability density function (PDF) of a "proposal distribution" from which θ^* is generated.

The PDA method is surprisingly easy to program and use because many statistical software packages such as R, Python, and MATLAB, already possess density functions that can be modified to use the (popular) Epanechnikov kernel and Silverman's rule of thumb for bandwidth selection. Thus, in practice, implementing the method involves (1) calling the density function for each of the C alternatives, and (2) scaling (i.e., multiplying) the resulting density values obtained by the number of times the corresponding alternative was chosen in the simulation. These scaled densities serve as Equation 2.

4. Comparing the Models

To compare the relative fit of the two models to the data, we will compute a total of four metrics: the Akaike information criterion (AIC; Akaike, 1973), the Bayesian information criterion (BIC; Schwarz, 1978), the Bayesian predictive information criterion (BPIC; Ando, 2007), and the Bayes factor. The AIC measure is obtained by calculating

$$\text{AIC} = -2\log(L(\widehat{\theta}|D)) + 2p, \tag{7}$$

where $L(\widehat{\theta}|D)$ represents the likelihood function evaluated at the best-fitting parameter $\widehat{\theta}$ (i.e., the maximum likelihood value obtained during estimation), and p represents the number of parameters. Lower values of AIC indicate a better model "fit", which is defined by a balance of predictive ability and model complexity.

The BIC is obtained in a similar way as the AIC, specifically by evaluating the following equation:

$$\text{BIC} = -2\log(L(\widehat{\theta}|D)) + \log(N)p, \tag{8}$$

where N represents the number of data points. Equations 7 and 8 differ only in the treatment of the penalization for number of model parameters. For the AIC, the number of parameters are multiplied by two, whereas for the BIC, the natural logarithm of the number of data points is used. Hence, when $N > 7.39$, a stronger penalty is applied for the BIC relative to the AIC. In comparing the two metrics, Kass et al. (2014) noted the following:

> "In practice, BIC is conservative compared to AIC in that it imposes a larger penalty for dimensionality. Thus, BIC is used, rather than AIC, when there is a strong preference for models of lower dimensionality." (p. 297)

The third metric is the BPIC. The BPIC was designed as a correction to the deviance information criterion (DIC; Spiegelhalter et al., 2002) on the grounds that the DIC tends to prefer models that over-fit the data (c.f., Ando, 2007). To compute the BPIC, we first define the "deviance" as $V(\theta) = -2\log(L(\theta|D))$. We then evaluate the expectation $\bar{V}$ of the deviance by taking the mean of V over all sampled values of θ (i.e., $\bar{V} = E(V(\theta))$, where E denotes the expected value). Subtracting from this expectation the best log-likelihood value obtained, $\hat{V} = \min(V)$ (Celeux et al., 2006; Spiegelhalter et al., 2002), we obtain a measure of the effective number of parameters $p_V = \bar{V} - \hat{V}$. The effective number of parameters is based on the difference between the expected deviance and an estimate of the deviance at the most likely value of the parameters (Dempster, 1997).[3] The choice of $\hat{V} = \min(V)$ rather than $\hat{V} = V(E(\theta))$ is justified here because the posterior distributions are non-normal and are not symmetric (Celeux et al., 2006). As p_V increases, the model becomes more flexible, making it easier for the model to fit the data. The BPIC value is obtained by evaluating

$$BPIC = \bar{V} + 2p_V \tag{9}$$

(Ando, 2007). As with the AIC and BIC, models with smaller (i.e., more negative) BPIC values are preferred over models with larger BPIC values.

4.1. Estimating Bayes Factor

The final metric is the Bayes factor. For a given model candidate M_q, model parameters θ_q, and data D, the posterior distribution of the model parameters can be expressed as

$$p(\theta_q|D, M_q) = \frac{L(\theta_q|D, M_q)p(\theta_q|M_q)}{\int L(\theta_q|D, M_q)p(\theta_q|M_q)d\theta_q}, \tag{10}$$

where $p(\theta_q|M_q)$ represents the prior distribution of the parameters θ_q, and $L(\theta_q|D, M_q)$ represents the likelihood function. The denominator of Equation 10 represents the degree of model evidence, or in other words, the probability of observing the data D given a candidate model M_q. The degree of model evidence is often written as $p(D|M_q)$, such that

$$p(D|M_q) = \int L(\theta_q|D, M_q)p(\theta_q|M_q)d\theta_q. \tag{11}$$

[3]Given that this metric is based on the information in the posteriors themselves, a direct comparison between the BPIC, BIC, and AIC is not straightforward.

We can use Bayes rule to evaluate the probability of a particular model M_q among a set of Q models, conditional on the data, given by

$$p(M_q|D) = \frac{p(D|M_q)p(M_q)}{\sum_{j=1}^{Q} p(D|M_j)p(M_j)}. \quad (12)$$

Equation 12 implies that for Models q and r,

$$\frac{p(M_q|D)}{p(M_r|D)} = \frac{p(D|M_q)p(M_q)}{p(D|M_r)p(M_r)}. \quad (13)$$

Within Equation 13, the Bayes factor comparing Models q and r is given by

$$BF_{q,r} = \frac{p(D|M_q)}{p(D|M_r)}.$$

We face two issues at this point. First, Equation 11 is not analytically tractable for the models we will examine in this article, and as a consequence, Equation 11 must be estimated by using numerical integration or approximated asymptotically. Second, because exact equations to calculate the likelihood functions for each model are unavailable, we must resort to an approximation. To approximate the Bayes factor, we rely on a method presented in Kass and Raftery (1995) for estimating the Bayes factor through a comparison of each model's BIC (see Equation 8). Kass and Raftery (1995) show that when comparing Models q and r, the difference in the BIC values $\text{BIC}_q - \text{BIC}_r$ asymptotically approximates $-2\log(BF_{q,r})$ as the sample size increases (i.e., as $N \to \infty$). Hence, we can approximate the Bayes factor by evaluating

$$BF_{q,r} \approx \exp\left[-\frac{1}{2}\left(\text{BIC}_q - \text{BIC}_r\right)\right]. \quad (14)$$

The approximation in Equation 14 does produce more relative error in approximating the Bayes factor than other, Hessian-based methods (e.g., De Bruijn, 1970; Tierney and Kadane, 1986; Kass and Vaidyanathan, 1992), but in large samples the Equation 14 should provide a reasonable indication of model evidence (cf. Kass and Raftery, 1995) For the data we will examine in the present manuscript, the number of data points N is around 400 per subject, which increases the penalty term in the BIC calculation and improves the accuracy of the Bayes factor. Additionally, because the models we investigate in this manuscript have intractable likelihood function, the Hessian matrix is unavailable, making other approximations to the Bayes factor infeasible (e.g., De Bruijn, 1970; Tierney and Kadane, 1986; Kass and Vaidyanathan, 1992). Finally, as noted in Kass and Raftery (1995), in the usual case where the precision of the prior information is small relative to the information provided by the data (i.e., the likelihood function), the Schwarz criterion (Schwarz, 1978) indicates that the model that minimizes the BIC (see Equation 8) is the model with the highest posterior probability. Furthermore, when the prior distribution is a multivariate normal prior with mean at the maximum likelihood estimate and the variance is set equal to the expected information matrix for one observation of data (i.e., a prior called the "unit information prior"), the BIC approximation becomes more accurate (Weakliem, 1999). Specifically, using the Hessian-based method produces an error of order $O(N^{-1})$, using the BIC approximation with the unit information prior the approximation has an error of order $O(N^{-1})$, and using the BIC approximation with no explicit assumptions about the priors the approximation produces an error of order $O(1)$, where $O(x)$ refers to a term bounded in probability to some constant multiplied by x.

5. Models

In this article, we will compare two models inspired by neurophysiology. Both models were designed to embody certain characteristics of actual neuronal functions, such as leakage, lateral and feed-forward inhibition. The first model is the LCA model, and the second is the FFI model. We will now describe each of these models in turn.

5.1. The Leaky Competing Accumulator Model

The LCA model was developed as a neurologically plausible way to describe the dynamics of response competition. For this model, we denote the rate of accumulation for the cth accumulator as ρ_c, the lateral inhibition parameter as β, the leakage parameter as κ, and the degree of noise in the accumulation process as ξ_t, which when simulated is drawn from a normal distribution with a mean of zero and standard deviation η. In other words, at each time step t in the evidence accumulation process, $\xi_t \sim \mathcal{N}(0, \eta)$. The activation of the cth accumulator in the model is represented by the stochastic differential equation

$$\begin{aligned} dx_c &= \left(\rho_c - \kappa x_c - \beta \sum_{j \neq c} x_j \right) \frac{dt}{\Delta_t} + \xi_t \sqrt{\frac{dt}{\Delta_t}} \\ x_c &\rightarrow \max\left(x_c, 0\right), \end{aligned}$$

where Δ_t is a time constant parameter. Once the degree of evidence for any accumulator reaches a threshold α, the process is terminated and a response is elicited. Similar to most models of choice RT, the LCA model assumes a non-decision time parameter, which we will denote τ. Although other choices can certainly be made, we assumed that the accumulation dynamics start at zero by setting $x_c = 0$ for both $c = \{1, 2\}$.

Although in Turner and Sederberg (2014) we fit a hierarchical version of the LCA model to a small subset of the data, here we will fit each subject independently to better assess each model's ability to fit data from different individuals. To satisfy mathematical scaling properties, we constrained the drift rate parameters to sum to one (i.e., $\sum_c \rho^{(c)} = 1$ for each subject). The sum-to-one assumption is a simplifying assumption that is commonly used, but can have an influence on model discriminability (cf. Teodorescu and Usher, 2013). We fixed $dt = 0.01$ (with the unit of seconds), and $\Delta_t = 0.1$. In fitting the model to data, we specified the following uninformative priors:

$$\begin{aligned} \alpha_j^{(k)} &\sim \mathcal{U}(0, 25), \\ \rho_j^{(1)} &\sim \mathcal{U}(0, 1), \\ \eta_j &\sim \mathcal{U}(0, 25), \\ \kappa_j &\sim \mathcal{U}(0, 1), \\ \beta_j &\sim \mathcal{U}(0, 1), \text{ and} \\ \tau_j &\sim \mathcal{U}(0, \min\left[RT_j\right]), \end{aligned}$$

where $k \in \{A, N, S\}$ (i.e., the accuracy (A), neutral (N), and speed conditions (S), respectively), and $\min(RT_j)$ is the minimum of the observed response times for the jth subject. We use the uniform distribution to enforce the constraint that $\beta, \kappa \in [0.0, 1.0]$, which preserves the model's neurological plausibility. Specifically, values of β and κ greater than 1.0 would imply that the effect of lateral inhibition and/or leak would be greater than the activation of the accumulator itself (recall that the drift rates are bound by $\rho \in [0, 1]$), a parameter regime that we felt was at odds with the underlying motivation of the LCA model.

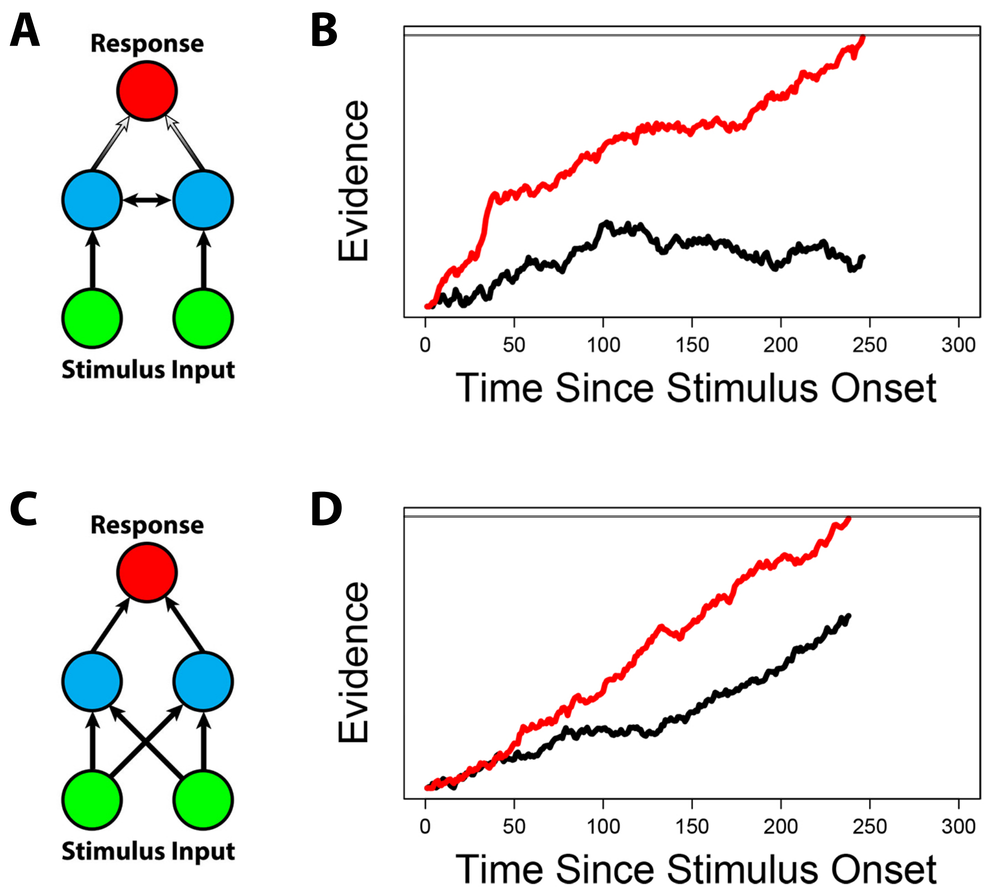

Figure 1: Graphical depiction of models compared. The top row (i.e., Panels A and B) corresponds to the LCA model, whereas the bottom row (i.e., Panels C and D) corresponds to the FFI model. The left column (i.e., Panels A and C) shows a graphical representation of how the stimulus input is mapped to the behavioral response. The right column (i.e., Panels B and D) shows a representative simulation of each corresponding model in a two-alternative decision task.

Panel A in Figure 1 shows a graphical diagram of the LCA model in a two choice decision task. The bottom nodes represent the input of the stimulus, which are connected to the observer's internal belief state (i.e., middle nodes) by the drift rates ρ. In the LCA model, the stimulus input only affects the corresponding belief state. At the belief state level, a competition ensues between the alternatives. The dynamics of the competitive process is dependent on the amount of evidence that has been accumulated as well as the lateral inhibition parameter β. Essentially, as more evidence is accumulated for a particular alternative, the influence of the competition becomes more pronounced, and the leading alternative gains even more of an advantage. In addition, the belief state is "leaky", meaning that some of the accumulated evidence is lost at a rate proportional to κ. Similar to the competition process, the amount of leakage also depends on the current state of accumulated evidence such that a larger amount of evidence is lost as more evidence is accumulated. Finally, the internal belief state level is mapped to an overt response once a threshold amount of evidence α has been accumulated.

Panel B in Figure 1 shows a representative simulation of the LCA model in a two-choice task. At stimulus onset, the evidence for each of the alternatives is equivalent. As the trial continues, one alternative gains an advantage, and due to the competitive process, the leading alternative gains even more of an advantage and accumulates evidence at a faster rate until the leading alternative reaches the threshold.

5.2. The Feed Forward Inhibition Model

The FFI model assumes no leakage and uses a different competitive mechanism where inhibition is based on the average *input* to the other alternatives, such that

$$\begin{aligned} dx_c &= \left(\rho_c - \frac{\nu}{C-1}\sum_{j\neq c}\rho_j\right)\frac{dt}{\Delta_t} + \xi_t\sqrt{\frac{dt}{\Delta_t}} \\ x_c &\rightarrow \max\left(x_c, 0\right), \end{aligned}$$

where ν is the feed-forward inhibition parameter, ρ_c represents the rate of evidence accumulation for the cth alternative, $\xi_t \sim \mathcal{N}(0,\eta)$ represents the within-trial variability, and C represents the number of choice alternatives (i.e., $C = 2$ here). We again constrained the drift rates to sum to one, as in the LCA model, to satisfy mathematical scaling properties. As in the LCA model, we again fixed $dt = 0.01$ (with the unit of seconds), and $\Delta_t = 0.1$. As in the LCA model above, we assumed that the accumulation dynamics start at zero by setting $x_c = 0$ for both $c = \{1, 2\}$.

In fitting the model to data, we specified the following uninformative priors:

$$\begin{aligned}
\alpha_j^{(k)} &\sim \mathcal{U}(0, 25), \\
\rho_j^{(1)} &\sim \mathcal{U}(0, 1), \\
\eta_j &\sim \mathcal{U}(0, 25), \\
\nu_j &\sim \mathcal{U}(0, 1), \text{ and} \\
\tau_j &\sim \mathcal{U}(0, \min[RT_j]).
\end{aligned}$$

As in the LCA model above, we constrained $\nu_j \in [0, 1]$ to preserve the model's neurological plausibility.

Panel C in Figure 1 shows a graphical diagram of the evidence accumulation process in the FFI model. Similar to the LCA model above, the internal belief state is primarily affected by the stimulus input, again regulated by the parameters ρ. Unlike the LCA model, however, the stimulus input for each alternative *also* affects the input for the remaining alternatives (shown in the diagram as the crossing arrows) by way of a feed-forward inhibition process regulated by the parameter ν. At the internal belief state level, there is no internal competition between the alternatives as in the LCA. Finally, the belief state is mapped to the overt response once a threshold amount of evidence α has been reached. In contrast to the LCA model, the FFI model assumes that the mapping to the response state is not subject to imperfections such as leakage. Furthermore, the competitive mechanisms assumed by the FFI are never dependent on the amount of accumulated evidence, as in the LCA model.

Panel D in Figure 1 shows a representative simulation of the FFI model in a two-choice decision task. On stimulus presentation, evidence accumulates for each alternative and they race to the threshold α. In this case, one particular alternative gains a slight advantage and that advantage prevails until it eventually reaches the threshold first. Note that the advantage gained by the (eventual) winning alternative does not increase its win margin as evidence accumulates, as in the LCA model.

5.2.1. A Constrained FFI Model

In addition to the LCA and FFI models, we also examined a constrained version of the FFI model that resembles the popular drift diffusion model

(DDM; Ratcliff, 1978). Specifically, we examined a version of the FFI model that constrained $\nu = C - 1 = 1$, which we will refer to as the constrained FFI (CFFI). In the two-alternative case, this constraint modifies the accumulation process to be completely anticorrelated, turning Equation 15 above to

$$\begin{aligned} dx_c &= (\rho_c - \rho_{-c}) \frac{dt}{\Delta_t} + \xi_t \sqrt{\frac{dt}{\Delta_t}} \\ x_c &\to \max(x_c, 0), \end{aligned}$$

where ρ_{-c} represents the drift rate for the opposing decision alternative with respect to c. In this parameter regime, the CFFI behaves much like the classic DDM with a few exceptions. First, the CFFI does not have trial-to-trial variability in either the nondecision time, drift rate, or starting point. Second, the CFFI still maintains a floor on evidence accumulation such that neither accumulator can ever be negative. Finally, if starting points are manipulated, the two models are not equivalent (Teodorescu and Usher, 2013).

6. Results

6.1. Estimating the Posterior

To estimate the posterior distributions for each model, we used the PDA method for mixed data types (Turner and Sederberg, 2014). For each parameter proposal, we simulated the model $J = 50,000$ times to form a stable approximation of the likelihood function (see Equations 2 and 5). For these models, some parameter combinations lead to model simulations that could, in theory, take an infinitely long time to finish. To avoid this issue, we set a threshold of 10 seconds for the response times. If the model had not crossed a boundary at that point, we recorded the response time as 10 seconds with the choice being randomly selected. Because these model simulations led to poor fits to the data, these particular parameter combinations were never observed in the joint posterior distributions. The bandwidth parameters h were calculated for each proposal by means of Equation 4. To increase the accuracy of the Epanechnikov kernel density approximation, we applied a log transformation to the simulated RTs, which helped produce more normally-distributed data. As described above, we scaled the approximate density functions for each choice by the corresponding proportion of total responses out of the J simulations to determine the defective distribution for each choice.

As shown in Turner et al. (2013c), the parameters of choice RT models can be highly correlated, which makes conventional sampling algorithms such as Markov chain Monte Carlo (MCMC; Robert and Casella, 2004) inefficient to use. As such, we used a genetic algorithm called differential evolution (DE) with MCMC (DE-MCMC; ter Braak, 2006; Turner et al., 2013c; Turner and Sederberg, 2012). DE-MCMC is a population Monte Carlo algorithm that generates proposals on every trial based on the information learned in the current estimate of the posterior. The communication between the "chains" in the algorithm allows DE-MCMC to generate proposals to match the shape of the posterior, regardless of how correlated the parameters may be. Furthermore, the DE-MCMC algorithm is well-designed for high-dimensional parameter spaces (see, e.g., Turner and Sederberg, 2012). For each of the four different likelihood evaluation methods, we implemented our DE-MCMC sampler, with 50 chains for 2,000 sampling iterations following 500 burn-in iterations, producing 100,000 samples of the joint posterior distribution. For each DE proposal, we randomly sampled the scaling factor $\gamma \sim (0.5, 1.0)$. We set the random perturbation parameter b of the uniform distribution equal to 0.001. Convergence was assessed through visual inspection and the R package coda (Plummer et al., 2006). Additional implementation details of the sampler can be found in Turner et al. (2013c).

6.2. Comparing the Models

Once the posteriors had been estimated, we could then evaluate the relative merits of the models by calculating the three model fit statistics discussed above. We calculated the AIC by Equation 7, the BIC by Equation 8, and the BPIC by Equation 9. Table 1 shows these calculations for each of the three models and each of the 20 subjects. The table is arranged so that the three metrics are grouped together to facilitate a comparison across the three models. The last row in the table summarizes the results by calculating for each column, the number of times the model in the corresponding column provided the best fit (i.e., lowest value) in the dataset. Interestingly, the three metrics do not tell the same story. Specifically, while the AIC and BIC measures put the FFI model slightly ahead of the LCA model, the BPIC measure heavily favors the LCA model. The CFFI model clearly performs worse than all of the other models, regardless of the fit statistic.

Table 1: Fit statistics comparing each of the three models.

	AIC			BIC			BPIC		
Subject	CFFI	FFI	LCA	CFFI	FFI	LCA	CFFI	FFI	LCA
1	688.63	**349.60**	351.68	718.57	**384.53**	391.59	708.55	**373.68**	378.85
2	221.60	**135.62**	137.16	251.56	**170.57**	177.11	237.85	163.53	**148.75**
3	419.59	292.29	**267.36**	449.52	327.21	**307.27**	441.66	309.50	**277.29**
4	729.23	**479.79**	480.41	759.18	**514.73**	520.34	745.36	**495.07**	500.09
5	647.98	**498.69**	502.24	677.95	**533.65**	542.19	664.74	521.38	**521.02**
6	-91.45	-87.22	**-147.55**	-61.49	-52.27	**-107.60**	-71.72	-70.10	**-104.17**
7	330.10	**135.13**	137.32	360.07	**170.09**	177.28	348.69	150.70	**150.67**
8	447.04	385.67	**373.75**	477.00	420.63	**413.69**	460.37	411.59	**390.63**
9	504.84	**426.29**	428.49	534.79	**461.23**	468.43	520.64	450.46	**450.44**
10	191.18	152.64	**122.32**	221.12	187.57	**162.24**	210.14	174.20	**138.42**
11	330.83	**186.96**	189.42	360.80	**221.92**	229.38	346.66	202.58	**200.65**
12	316.75	238.60	**191.91**	346.72	273.56	**231.86**	330.96	262.31	**223.14**
13	652.74	479.10	**455.68**	682.66	514.01	**495.58**	670.47	495.19	**475.08**
14	585.37	372.59	**371.70**	615.33	**407.55**	411.65	606.02	**399.22**	402.18
15	559.09	**398.26**	401.50	589.03	**433.19**	441.43	583.58	422.57	**417.26**
16	335.53	**196.47**	204.75	364.32	**230.06**	243.13	365.33	**227.32**	239.98
17	704.30	**396.27**	400.70	734.27	**431.23**	440.65	737.31	425.81	**423.63**
18	373.31	314.61	**309.46**	403.26	349.55	**349.39**	391.31	332.05	**325.70**
19	568.36	403.12	**364.51**	598.28	438.04	**404.41**	592.36	423.10	**381.54**
20	636.17	**437.08**	437.44	666.12	**472.02**	477.38	659.55	**455.59**	460.31
Wins	0	11	9	0	12	8	0	5	15

Table 2: Bayes factors comparing each of the three models.

Subject	FFI/CFFI	FFI/LCA	LCA/CFFI	Winner
1	3.43×10^{72}	34.13	1.01×10^{71}	FFI
2	3.86×10^{17}	26.26	1.47×10^{16}	FFI
3	3.63×10^{26}	4.67×10^{-5}	7.77×10^{30}	LCA
4	1.20×10^{53}	16.54	7.28×10^{51}	FFI
5	2.15×10^{31}	71.62	3.01×10^{29}	FFI
6	9.97×10^{-3}	9.67×10^{-13}	1.03×10^{10}	LCA
7	1.79×10^{41}	36.31	4.93×10^{39}	FFI
8	1.75×10^{12}	0.031	5.59×10^{13}	LCA
9	9.40×10^{15}	36.53	2.57×10^{14}	FFI
10	1.93×10^{7}	3.16×10^{-6}	6.10×10^{12}	LCA
11	1.43×10^{30}	41.54	3.45×10^{28}	FFI
12	7.69×10^{15}	8.81×10^{-10}	8.73×10^{24}	LCA
13	4.19×10^{36}	9.92×10^{-5}	4.22×10^{40}	LCA
14	1.32×10^{45}	7.78	1.69×10^{44}	FFI
15	6.92×10^{33}	61.37	1.13×10^{32}	FFI
16	1.43×10^{29}	690.48	2.07×10^{26}	FFI
17	6.35×10^{65}	110.93	5.73×10^{63}	FFI
18	4.62×10^{11}	0.93	4.98×10^{11}	LCA
19	6.26×10^{34}	5.00×10^{-8}	1.25×10^{42}	LCA
20	1.41×10^{42}	14.56	9.69×10^{40}	FFI

6.3. Bayes Factors

Once an approximation for each of the posterior distributions had been obtained, we evaluated the BIC values according to Equation 8, and subsequently used the BIC values to approximate the Bayes factor for each possible model comparison by evaluating Equation 14 for each individual subject. Table 2 shows the estimated Bayes factors comparing the FFI to the CFFI (second column), the FFI to the LCA (third column), and the LCA to the CFFI (fourth column). Table 2 shows that the FFI provides the best fit for 12 out of the 20 subjects, and the LCA model provides the best fit for the remaining 8 subjects. The constrained FFI model did not provide the best fit to any subject in this particular suite of models.

Figure 2 illustrates a comparison of the FFI model to the LCA model (see column 3 in Table 2). The figure shows the Bayes factor for each subject, ranked according to increasing evidence for the FFI model. The point

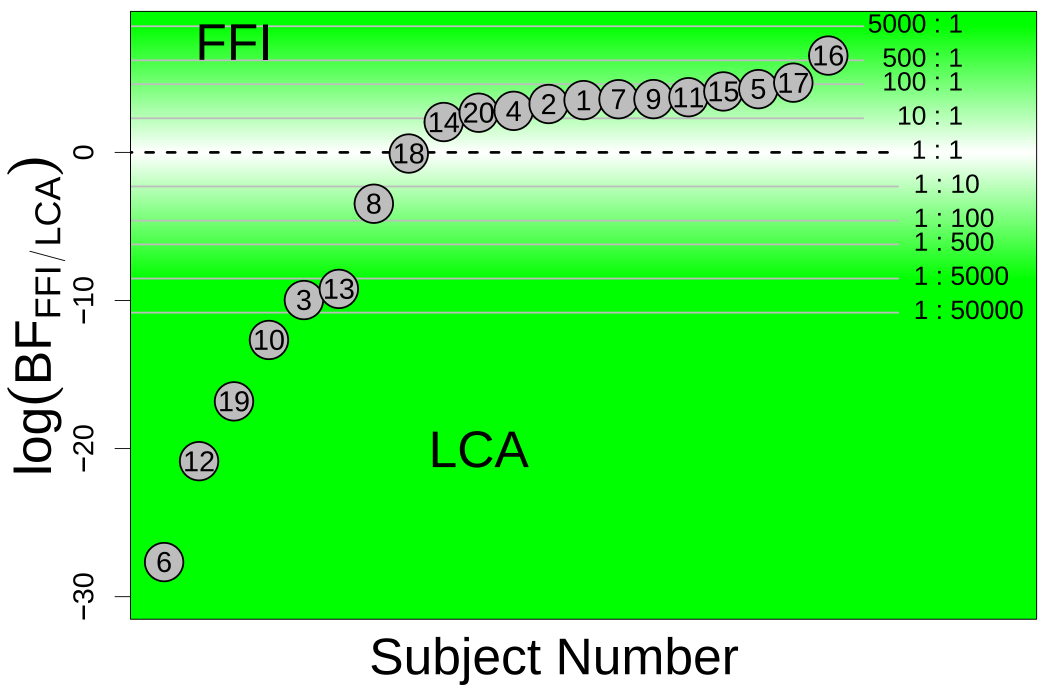

Figure 2: A comparison of the Bayes factors comparing the FFI model to the LCA model for each subject. Subjects have been ranked in increasing order, where a higher Bayes factor corresponds to greater evidence for the FFI model. The point of indifference between the two models is represented as the dashed horizontal line at zero.

of indifference for the two competing models is shown as the dashed black horizontal line at zero. As a reference, other gray lines are plotted to show differing amounts of model evidence. From the point of indifference, regions are color-coded to illustrate greater degrees of evidence for either the FFI model (top) or the LCA model (bottom). Figure 2 suggests that when the LCA model is the preferred model, the evidence greatly outweighs the evidence for the FFI model. However, when the FFI model is the preferred model, there is a smaller degree of evidence for the FFI model over the LCA model.

7. Discussion

In this article, we used the recently developed probability density approximation (PDA) method to fit two neural network models to the data presented in Forstmann et al. (2011). The first model, the Leaky Competing Accumulator (LCA; Usher and McClelland, 2001) uses neurally plausible mechanisms such as competition via lateral inhibition, and leakage. The second model, the Feed-forward Inhibition (FFI; Shadlen and Newsome, 2001) model, assumes that competition between alternatives follows a feed-forward inhibition process, and assumes that leakage is not present in the network. Both models are neurally inspired and have been shown to account for many enriched experimental manipulations (e.g., Usher and McClelland, 2001; Tsetsos et al., 2011; Bogacz et al., 2006; Gao et al., 2011; van Ravenzwaaij et al., 2012; Bogacz et al., 2007; Shadlen and Newsome, 2001; Teodorescu and Usher, 2013)

On fitting the models to data, we then compared the models by calculating several statistics, namely the Akaike information criterion (AIC; Akaike, 1973), the Bayesian information criterion (BIC; Schwarz, 1978), the Bayesian predictive information criterion (BPIC; Ando, 2007), and the Bayes factor. The AIC and BIC measures provided evidence that the FFI model was preferred over the LCA model, but only by two (for the AIC) or four (for the BIC) subjects out of 20. However, when using the BPIC measure, the LCA model provided the best fit to 15 out of 20 subjects, with the FFI model capturing the remaining five. Given the discrepancies among the metrics, it is clear that more extensive analyses are needed to fully differentiate these particular models. We could also compare the models by aggregating across the three metrics. For four subjects (i.e., Subjects 1, 4, 16, and 20) the FFI model provided the best fit on all three metrics, whereas for eight subjects (i.e., Subjects 3, 6, 8, 10, 12, 13, 18, and 19) the LCA model provided the best fit. Examining Table 2 in this way suggests that the decision making processes used by these particular subjects are best described by a particular model.

We also compared the models by approximating the Bayes factor through the Bayesian information criterion (see Equation 14; Schwarz, 1978; Kass and Raftery, 1995). We first determined that the constrained version of the FFI model, which maintained that $\nu = C - 1 = 1$, performed substantially worse than either the full FFI or the LCA models. We then compared the LCA model to the FFI model for each subject. In total, the FFI model outperformed the LCA model for 12 of the 20 subjects. However, we noted that when the LCA model outperformed the FFI model, it did so in an extreme way. This aspect of our results may indicate that there is something unique about the decision processes used by a subset of the subjects in our data. For example, the decision process for these subjects may be prone to a leaky mapping of the internal belief state to the response state, or it may be that the competition between the decision alternatives resembles a time-dependent process (as assumed by the LCA model) rather than a time-invariant one (as assumed by the FFI model). Another possible explanation is that the simplifying assumptions used hindered the LCA model's ability to fit the data for some subjects.

While in this manuscript, we have relied on the BIC approximation to the Bayes factor, there are other choices available in the likelihood-free context. One approach is to treat the model selection problem as a hierarchical modeling problem (Grelaud et al., 2009; Toni and Stumpf, 2010; Turner and Van Zandt, 2014), and estimate the model probabilities using a specific sampling algorithm such as sequential Monte Carlo (Toni and Stumpf, 2010; Toni et al., 2009), Gibbs approximate Bayesian computation (Turner and Van Zandt, 2014), or random forests (Pudlo et al., 2014). However, these methods require certain conditions on the statistics that characterize the observed data for the approximation to hold (Didelot et al., 2011; Robert et al., 2011). Namely, the models must be nested and statistics much be chose that characterize the data in a sufficient manner for the entire collection of models under examination (Didelot et al., 2011). In our case, the problems associated with approximate Bayesian model choice do not apply because we consider the entire set of data, which is guaranteed to be a sufficient statistic (c.f. Toni et al., 2009; Turner and Van Zandt, 2012; Turner and Sederberg, 2014). However, future work could build on our approach by estimating the model evidence explicitly.

In conclusion, for the data tested here (Forstmann et al., 2011), the metrics AIC, BIC, and Bayes factor provided a small amount of evidence to support the FFI model, whereas the BPIC provided a strong amount of evidence in favor of the LCA model. We noted that for some subjects, one model was preferred when corroborating all four metrics. A more extensive analyses of the models would examine other important factors such as the number of decision alternatives, stimulus types (e.g., stationary versus time-varying evidence), and payoff manipulations.

8. References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox, B. N., Caski, F. (Eds.), Second Internation Symposium on Information Theory. pp. 267–281.

Ando, T., 2007. Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models. Biometrika 94, 443–458.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., Cohen, J. D., 2006. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. Philosophical Transactions of the Royal Society, Series B: Biological Sciences 362, 1655–1670.

Bogacz, R., Usher, M., Zhang, J., McClelland, J. L., 2007. Extending a biologically inspired model of choice: Multi-alternatives, nonlinearity and value-based multidimensional choice. Theme issue on modeling natural action selection. Philosophical Transactions of the Royal Society: B. Biological Sciences 362, 1655–1670.

Brown, S., Heathcote, A., 2008. The simplest complete model of choice reaction time: Linear ballistic accumulation. Cognitive Psychology 57, 153–178.

Celeux, G., Forbes, F., Robert, C. P., Titterington, D. M., 2006. Deviance information criteria for missing data models. Bayesian Analysis 1, 651–673.

De Bruijn, N. G., 1970. Asymptotic methods in analysis. Amsterdam: North-Holland.

Dempster, A. P., 1997. The direct use of likelihood for significance testing. Statistics and Computing 7, 247–252.

Didelot, X., Everitt, R. G., Johansen, A. M., Lawson, D. J., 2011. Likelihood-free estimation of model evidence. Bayesian Analysis 6, 49–76.

Donkin, C., Averell, L., Brown, S., Heathcote, A., 2009a. Getting more from accuracy and response time data: Methods for fitting the Linear Ballistic Accumulator. Behavioral Research Methods 41, 1095–1110.

Donkin, C., Heathcote, A., Brown, S., 2009b. Is the Linear Ballistic Accumulator model really the simplest model of choice response times: A Bayesian model complexity analysis. In: Howes, A., Peebles, D., Cooper, R. (Eds.), 9th International Conference on Cognitive Modeling – ICCM2009. Manchester, UK.

Epanechnikov, V. A., 1969. Non-parametric estimation of a multivariate probability density. Theory of probability and its applications 14, 153–158.

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., Brown, S., 2011. The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. Journal of Neuroscience 31, 17242–17249.

Gao, J., Tortell, R., McClelland, J. L., 2011. Dynamic integration of reward and stimulus information in perceptual decision-making. PLoS ONE 6, 1–21.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian Data Analysis. Chapman and Hall, New York, NY.

Gilks, W. R., Best, N. G., Tan, K. K. C., 1995. Adaptive rejection Metropolis sampling withing Gibbs sampling. Applied Statistics 44, 455–472.

Gilks, W. R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. Applied Statistics 41, 337–348.

Grelaud, A., Marin, J.-M., Robert, C., Rodolphe, F., Tally, F., 2009. Likelihood-free methods for model choice in Gibbs random fields. Bayesian Analysis 3, 427–442.

Kass, R. E., Eden, U. T., Brown, E. N., 2014. Analysis of Neural Data, 1st Edition. New York: Springer.

Kass, R. E., Raftery, A. E., 1995. Bayes factors. Journal of the American Statistical Society 90, 773–795.

Kass, R. E., Vaidyanathan, S., 1992. Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. Journal of the Royal Statistical Society, Series B 54, 129–144.

Montenegro, M., Myung, J. I., Pitt, M. A., 2011. REM integral expressions, unpublished manuscript.

Myung, I. J., 2003. Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology 47, 90–100.

Myung, J. I., Montenegro, M., Pitt, M. A., 2007. Analytic expressions for the BCDMEM model of recognition memory. Journal of Mathematical Psychology 51, 198–204.

Plummer, M., Best, N., Cowles, K., Vines, K., March 2006. CODA: Convergence diagnosis and output analysis for MCMC. R News 6 (1), 7–11. URL `http://CRAN.R-project.org/doc/Rnews/`

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., Robert, C. P., Jun. 2014. ABC model choice via random forests. ArXiv e-prints.

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

Ratcliff, R., Smith, P. L., 2004. A comparison of sequential sampling models for two-choice reaction time. Psychological Review 111, 333–367.

Robert, C. P., Casella, G., 2004. Monte Carlo statistical methods. Springer, New York, NY.

Robert, C. P., Cornuet, J.-M., Marin, J.-M., Pillai, N., 2011. Lack of confidence in approximate bayesian computation model choice. Proceedings of the National Academy of Sciences of the United States 108, 15112–15117.

Rouder, J. N., Sun, D., Speckman, P., Lu, J., Zhou, D., 2003. A hierarchical Bayesian statistical framework for response time distributions. Psychometrika 68, 589–606.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Shadlen, M. N., Newsome, W. T., 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. Journal of Neurophysiology 86, 1916–1936.

Silverman, B. W., 1986. Density estimation for statistics and data analysis. London: Chapman & Hall.

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

Teodorescu, A. R., Usher, M., 2013. Disentangling decision models – from independence to competition. Psychological Review 120, 1–38.

ter Braak, C. J. F., 2006. A Markov chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. Statistics and Computing 16, 239–249.

Tierney, L., Kadane, J. B., 1986. Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association 81, 82–86.

Toni, T., Stumpf, M. P. H., 2010. Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics 26, 104–110.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M. P., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface 6, 187–202.

Tsetsos, K., Usher, M., McClelland, J. L., 2011. Testing multi-alternative decision models with non-stationary evidence. Frontiers in Neuroscience 5, 1–18.

Turner, B. M., Dennis, S., Van Zandt, T., 2013a. Bayesian analysis of memory models. Psychological Review 120, 667–678.

Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., Steyvers, M., 2013b. A bayesian framework for simultaneously modeling neural and behavioral data. NeuroImage 72, 193–206.

Turner, B. M., Sederberg, P. B., 2012. Approximate Bayesian computation with Differential Evolution. Journal of Mathematical Psychology 56, 375–385.

Turner, B. M., Sederberg, P. B., 2014. A generalized, likelihood-free method for parameter estimation. Psychonomic Bulletin and Review 21, 227–250.

Turner, B. M., Sederberg, P. B., Brown, S., Steyvers, M., 2013c. A method for efficiently sampling from distributions with correlated dimensions. Psychological Methods 18, 368–384.

Turner, B. M., Van Zandt, T., 2012. A tutorial on approximate Bayesian computation. Journal of Mathematical Psychology 56, 69–85.

Turner, B. M., Van Zandt, T., 2014. Hierarchical approximate Bayesian computation. Psychometrika 79, 185–209.

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

van Ravenzwaaij, D., van der Maas, H. L. J., Wagenmakers, E. J., 2012. Optimal decision making in neural inhibition models. Psychological Review 119, 201–215.

Van Zandt, T., 2000. How to fit a response time distribution. Psychonomic Bulletin and Review 7, 424–465.

Weakliem, D. L., 1999. A critique of the Bayesian Information Criterion for model selection. Sociological Methods and Research 27, 359–397.