Is the Linear Ballistic Accumulator Model Really the Simplest Model of Choice Response Times: A Bayesian Model Complexity Analysis

Chris Donkin (Chris.Donkin@newcastle.edu.au)

Andrew Heathcote (Andrew.Heathcote@newcastle.edu.au)

Scott Brown (Scott.Brown@newcastle.edu.au)
School of Psychology, The University of Newcastle
Callaghan, NSW 2303 Australia

Abstract

Brown and Heathcote (2008) proposed the LBA as the simplest model of choice and response time data. This claim was, in part, based on the LBA requiring fewer parameters to fit most data sets than the leading alternative, the Ratcliff diffusion model (Ratcliff & Tuerlinckx, 2002). However, parameter counts fail to take into account functional form complexity, or how the parameters interact in the model when being estimated from data. We used p_D, or the "effective number of parameters", calculated from Markov Chain Monte Carlo samples, to take these factors into account. We found that in a relatively simple, simulated, data set and on average in a complex, real, data set that the diffusion had fewer effective parameters than the LBA.

Keywords: decision models; response time; Bayesian statistics; model complexity.

A wide range of experimental tasks involve a decision between at least two alternatives. Some believe that the process behind making simple decisions is the same regardless of what the decision is about. The most successful class of theories about simple decision processes are evidence accumulator models. There are many types of evidence accumulator model that differ slightly from one another. However, the central assumption common to all is that, when making a decision about a stimulus, evidence is gradually accumulated for each alternative response. Once there is enough evidence for one particular response that response is made, and the time taken to accumulate that evidence is the decision time. The most frequently applied evidence accumulator model for decisions between two alternatives is the Ratcliff diffusion model (Ratcliff, 1978; Ratcliff & Rouder, 1998; Ratcliff & Tuerlinckx, 2002). For example, Ratcliff and colleagues have used the diffusion model to account for the decision process in lexical decision tasks (Ratcliff, Gomez, & McKoon, 2004), recognition memory tasks (Ratcliff, 1978), to investigate the effects of aging on cognitive performance (e.g. Ratcliff, Thapar, & McKoon, 2004). Ratcliff, Segraves, and Cherian (2003) also present neural evidence consistent with the diffusion model.

Brown and Heathcote (2008) recently proposed an alternative evidence accumulator model of the decision process: the Linear Ballistic Accumulator (LBA) model. The LBA was proposed as a simpler model of decision than the diffusion model. The claim of simplicity was based in part on the fact that the LBA assumes one less source of noise in the decision process. That is, in constrast to the diffusion model, evidence accumulation in the LBA is ballistic (i.e. without moment-to-moment variability). This simplification, enables the derivation of full analytic expressions for the model's full probability density function. Despite this simiplificaiton, Brown and Heathcote (2008) show that the LBA is able to account for benchmark data from two-choice tasks (Ratcliff & Rouder, 1998; Ratcliff, Gomez, & McKoon, 2004) [1]. LBA parameters have also been shown to have neural correlates (Forstmann et al., 2008; Ho, Brown, & Serences, submitted).

Brown and Heathcote (2008) also claimed the LBA is simpler because, when fiting standard two-choice data, it required one less parameter than the most recent version of the diffusion model (Ratcliff & Tuerlinckx, 2002). Myung and Pitt (1997), however, explain that the number of free parameters, k, does not necessarily provide a full indication of model complexity. Specifically, k fails to take into account functional form complexity (i.e., differences in flexability between different mathematical functions), or how the parameters interact when parameters from the model are estimated from data. Spiegelhalter, Best, Carlin, and van der Linde (2002) proposed a method to address these aspects of model complexity using the deviance information criterion (DIC) and an associated estimate, p_D, of the effective number of model parameters. These quantities are estimated using posterior samples obtained by Bayesian Markov Chain Monte Carlo (MCMC) methods. We use these methods to investigate the claim that the LBA is a "simpler" model of the decision process. To begin we provide a brief overview of the diffusion and LBA models.

Overview of Models

Consider the following example – participants are shown a patch of 64x64 pixels, each of which are either white or black, and the asked whether the stimulus is mostly bright or mostly dark. The Ratcliff diffusion model begins by assuming that participants sample information continuously from the stimulus. Each sample of information counts as evidence for one of the two responses and is used to update an evidence total, say x, shown by the irregular line in the left panel of Figure 1.

[1] Brown and Heathcote (2008) also show that the LBA is able to account for decisions between more than two alternatives becasue it allows one accumulator for each choie. As the Ratcliff diffusion model has not been extended to the multiple choice case we will focus on the two choice case.

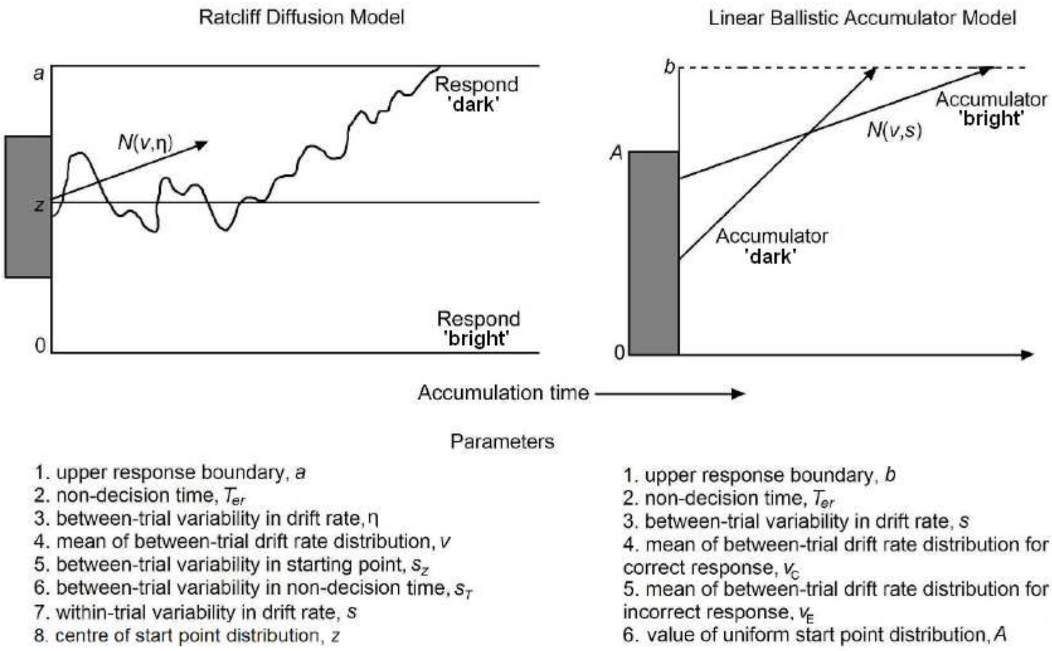

Figure 1: Overview of the diffusion and LBA models (left and right panel, respectively)

Total evidence begins at some starting point, $x = z$, and evidence that favours a "bright" response decreases the value of x and evidence for a "dark" response increases the value of x. Evidence accumumlation continues until x reaches one of the response boundaries, the horizontal lines at 0 or a in Figure 1. The choice made depends upon which boundary was reached, a for "dark" and 0 for "bright" response. The time taken to make the choice is the accumulation time plus a non-decision time component, T_{er}, composed of things such as encoding time and the time taken to make a motor response.

Consider a stimulus composed of almost 100% white pixels. When a participant samples from this stimulus almost all of the evidence will favour a "bright" response, and so the accumulation total will quickly increase towards a. The average rate of this accumulation is called the *drift rate*, v, and variability in moment-to-moment accumulation is assumed to take the value s. Ratcliff (1978) added the additional assumption that drift rate also varies from trial-to-trial according to a normal distribution with mean v and standard deviation η. Ratcliff and Rouder (1998) incorporated between-trial variability in the start point of acccumulation, assuming that z follows a uniform distribution on $[z - \frac{s_z}{2}, z + \frac{s_z}{2}]$. Finally, Ratcliff and Tuerlinckx (2002) included between-trial variability in non-decision time T_{er} in the form of a uniform distribution on $[T_{er} - \frac{s_t}{2}, T_{er} + \frac{s_t}{2}]$.

In the LBA there are separate accumulators gathering evidence for each of the "bright" and "dark" responses. These accumulators are assumed to be linear, ballistic and independent. That means evidence accumulation has a linear increase with no within-trial variability (i.e,. is ballistic rather than stochastic as in the diffusion model), and accumulation in one accumulator has no effect on the other accumulator(s). The amount of evidence an accumulator begins with on each trial is sampled (separately for each accumulator) from the interval $[0, B]$. The evidence in each accumulator increases at a linear rate determined by the drift rate parameters, v_b and v_d, for bright and dark responses, respectively. Accumulation continues until evidence in one accumulator reaches a response boundary, a^2 which is usually assumed to be the same for all accumulators. The accumulator which reaches the boundary first selects its associated response and accumulation time plus non-decision time, T_{er}, gives the reaction time. As in the Ratcliff diffusion model, the drift rate is assumed to vary between-trials according to a normal distribution with mean v and standard deviation η.

To sum up, the diffusion model has the parameters $(a, z, s_z, T_{er}, s_t, v, s, \eta)$ and the LBA has the parameters $(a, B, T_{er}, v_1, v_2, \eta)$, where v_i refers to the mean drift rate in the accumulator for the i_{th} response. The parameterisation for each model, however, differs depending on the design of the data from which the data were obtained. There is, therefore, no fixed difference in the number of parameters between the models. There are, however, parameterisations of these models which are commonly applied. For example, when there is no bias for one response over the other then the z parameter of the diffusion model can be fixed at $\frac{a}{2}$, reducing the number of free parameters in the diffusion model by one. Also, in order to solve a scaling property common to all evidence accumulator models, the s parameter is generally fixed at 0.1. Sim-

[2] In previous applications of the LBA a and B have been labelled b and A, respectively. We adopt this alternative labelling here to facilitate equality in parameter names across models.

ilarly, when fitting the LBA, drift rates for correct and error responses tend to be assumed equal for both choices unless the choice corresponds to an experimental manipulation (e.g., word vs. non-word in a lexical decision task or studied vs. unstudied in a recogniton memory task). Drift rates for error responses are also typically assumed to be fixed at one minus the drift rate for correct response, solving the scaling property for the LBA. This means when the LBA has been applied then usually only one drift rate parameter is estimated– the drift rate for correct responses. Based on these standard parameterisations, Brown and Heathcote (2008) concluded that the LBA uses one less parameter than the diffusion model to account for data typical of two-choice tasks. This finding, combined with some apparently simpler structural assumptions, led Brown and Heathcote (2008) to conclude that LBA was simpler than the diffusion model. We now explore whether the p_D measure of model complexity agrees with the author's conclusions.

Model Complexity

An overly complex model can provide an excellent fit to a given set of data, yet still not be considered to give a satisfying account of the underlying process. In particular, a more complex model can "overfit" the data by fitting the random error specific to a particular sample as well as the structure due to the underlying processes. Becasue only the structre re-occurs in new data, overfitting limits the model's ability in terms of prediction. Myung (2000) suggests that at least two factors contribute to model complexity – the number of parameters in the model and the functional form of the model, which determines how the parameters interact. Functional form complexity can differ between models with the same number of parameters when one model is able to produce a wider range of predictions than the other. In any particular experimental design, the degree to which the effects of functional form complexity are observed depends on the interaction between model and data.

A number of model selection methods take into account functional form complexity. We will focus on one such measure: the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). DIC has been applied across a wide range of fields including psychology (e.g., Myung, Karabatsos, & Iverson, 2005). Vandekerckhove, Tuerlinckx, and Lee (2008) used DIC to compare various instantiations of the diffusion model. The DIC can be considered the Bayesian version of the Akaike Information Criterion (AIC; Akaike, 1973), but with a complexity penalty term which takes into account functional form complexity, rather than simply counting the number of free parameters, as in AIC.

DIC can be computed from MCMC samples of a model's posterior parameter distributions. Let θ represent such a sample. Deviance can be written as $D(\theta) = -2\log L(y|\theta)$, where $L(y|\theta)$ represents the likelihood of data vector y given parameters θ. Then $D(\overline{\theta})$ is the deviance of the estimated posterior mean parameters and $\overline{D}(\theta)$ is the mean of the distribution of posterior samples. DIC can be expressed in two parts as $DIC = D(\overline{\theta}) + 2p_D$, where $p_D = \overline{D}(\theta) - D(\overline{\theta})$, where $D(\overline{\theta})$ is a measure of misfit between data and model predictions, and $2p_D$ is a penalty for the "effective" number of parameters in the model (Spiegelhalter et al., 2002). The p_D measure adjusts the number of parameters in the model to take account of functional form complexity. Larger values of p_D indicate a more complex model able to potentially predict a greater the range of patterns of data. A better model, which achieves a balance between fit and complexity, has a smaller DIC.

Posterior sampling for both the Ratcliff diffusion and LBA models have been implemented using the Bayesian MCMC program WinBUGS (diffusion: Vandekerckhove et al., 2008; LBA: Donkin, Averell, Brown, & Heathcote, 2009). We use these implementations to calculate DIC and p_D, allowing us to compare the functional form complexity between the models. Because DIC and p_D are dependent on the data to which the models are applied we will present the results of fits to two different sets of data: simulated data generated by the diffusion model, and a benchmark data set from Ratcliff and Rouder (1998).

Estimating p_D and DIC for the LBA and Diffusion Models

Simulated Data

The first set of data were generated from a diffusion process with parameters given in Table 1. Our simulated data set was intended to mimick data from a two-choice task with a single experimental factor where stimuli were varied so as to only affect the difficulty of the task. This meant that only the drift rate parameter, v, was allowed to vary across the three conditions. All other parameters $(a, sz, T_{er}, s_t, s, \eta)$ were assumed to be constant across all conditions. We also fixed z to be $\frac{a}{2}$, representing unbiased responding. This parameterisation is standard for fitting data from experiments which have a single within-subjects condition which varies from trial-to-trial (e.g. Ratcliff, Gomez, & McKoon, 2004). The simulated data can be thought of as coming from a single participant who completed 1000 trials in each of the three difficulty conditions.

When fitting both the diffusion model and the LBA model, parameters were fixed to match the assumptions made when generating the data; so only drift rate was allowed to vary between the three difficulty conditions. This means that for the diffusion model we have eight free parameters $(a, s_z, T_{er}, st, \eta, v_1, v_2, v_3)$, and for the LBA seven free parameters $(a, B, T_{er}, \eta, v_1, v_2, v_3)$. Unbiased responding in the LBA corresponds to having the same values of a and B for each response. Posterior samples were obtained for both models using their WinBUGS implementations. For each model three chains each containing 10,000 MCMC samples were collected, with the first 3,000 samples for each chain were discarded as burn-in. Visual inspection of the chains suggested that after burn-in samples collected from each chain were from the same stationary distribution, which we now assume to be the true posterior distribution.

Table 1: Mean of posterior samples for parameters from the diffusion and LBA models for fits to data generated from diffusion model. DIC and p_D are also reported for each model.

Parameter	Data	Diffusion	LBA
a	.125	.128	.252
s_z / B	.044	.034	.432
η	.133	.123	.237
T_{er}	.435	.432	.237
s_t	.196	.196	-
v_1	.1	.103	.609
v_2	.23	.226	.74
v_3	.363	.369	.882
DIC	-	-183.76	-47.55
p_D	-	5.97	6.81

Table 1 contains mean posterior samples for each parameter for both the diffusion and LBA models. The average posterior diffusion model parameter samples are close to the parameters used to generate the data, as expected. The average posterior LBA parameters are close to parameters estimated using non-Bayesian methods of fitting (e.g. maximum likelihood estimation) to the same data set.

DIC and p_D values are also given in Table 1. As one might expect, the DIC for the diffusion model is smaller than the DIC value for the LBA model (-183.76 and -47.55, for diffusion and LBA respectively), suggesting that the diffusion model provides a better account than the LBA of data simulated from a diffusion process. Quite unexpectedly, however, the p_D value for the diffusion model is also smaller than that for the LBA model, p_D equal to 5.97 and 6.81 respectively. This suggests that – despite the diffusion model having more free parameters than the LBA model – when functional form complexity is taken into account, the number of "effective" parameters is actually smaller than that of the LBA model.

At least for these simulated data, from a very simple experimental design, the results seem clear – the diffusion model is less complex than the LBA. As previously stated, however, functional form complexity depends upon the data being modelled. We turn now to actual data, to a data set which has become a benchmark data set for models of choice and response time (Brown & Heathcote, 2008; Vandekerckhove et al., 2008).

Ratcliff and Rouder's (1998) Data

Ratcliff and Rouder (1998) performed a simple brightness discrimination task with two within-subject factors: brightness and instructions. There were 33 levels of brighness used, determined by the proportion of white vs. black pixels in a 64x64 display (brightness was varied randomly from trial-to-trial). Between blocks of trials, participants were given instructions on whether to respond with an emphasis on speed or an emphasis on accuracy.

We fit diffusion and LBA models seperately to data from three individual participants, each of whom completed almost 8000 trials. Both models have previously been fit to the Ratcliff and Rouder (1998) data sets using non-Bayesian estimation techniques (diffusion: Ratcliff & Rouder, 1998; LBA: Brown & Heathcote, 2008). We used very similar parameterisations to that used in the original fits with three exceptions. First, for the diffusion model we included between-trial variability in non-decision time. This variability was included in the diffusion model as it has been standard practice since Ratcliff and Tuerlinckx (2002). Second, for the LBA both the upper bound of the uniform distribution of starting point of accumulation, B, and response threshold, a, were allowed to vary between speed and accuracy conditions. Brown and Heathcote (2008) assumed $B = a$ in the speed-emphasis condition, but we found that fit was greatly improved by removing this constraint. For the diffusion model we followed Ratcliff and Rouder (1998) and assumed that only boundary separation, a was allowed to vary between speed and accuracy conditions. Third, we found that the diffusion gave much better fits to data by estimating between-trial variability in start point of accumulation for speed and accuracy conditions separately. This contrasts with Ratcliff and Rouder (1998) approch, where s_z was fixed at $a/20$ for both speed and accuracy conditions.

For both models, only drift rate was allowed to vary between brightness conditions. Although there were 33 brightness conditions in the original data, the conditions were collapsed to seven since visual inspection suggested that the majority of brightness levels which were either very difficult or very easy were homogenous in RT and accuracy. This meant that for the diffusion model $(a_{acc}, a_{spd}, s_{z_{acc}}, s_{z_{spd}}, T_{er}, s_t, \eta)$ were free parameters, and for the LBA $(a_{acc}, a_{spd}, B_{acc}, B_{spd}, T_{er}, \eta)$ were free parameters. When combined with the seven drift rate parameters common to both models, there were 14 free parameters for the diffusion model, and 13 free parameters for the LBA model.

A single chain of 10,000 samples was collected for each of the LBA and diffusion models, with the first 3,000 samples discarded from analysis as burn-in. Again, visual inspection of the chain confirmed that stationarity after burn-in. Table 2 contains mean posterior parameter values for each model and each participant. Though, for brevity we do not present them here, plots of model predictions and data confirm that the average parameter values provide a good fit to the data. The quality of fit between model and data was greater for the diffusion model than the LBA. This is reflected in DIC and p_D values reported in Table 2: for all participants the diffusion model had a smaller DIC value than the LBA model[3]. As

[3]Donkin, Brown, and Heathcote (2009) have shown that an LBA model where the sum of correct and error drift rates are not over-constrained to be one can provide a large improvement in quality of fit. This comes, however, at the expense of an increase in the number of free parameters. Since we wish the present discussion to be a retrospective look at the claims of Brown and Heathcote (2008) we discuss this no further here.

Table 2: Mean of posterior samples for parameters from the diffusion and LBA models for fits to individual participants from Ratcliff and Rouder (1998). DIC and p_D are also reported for each model.

Participant	Model	a_{acc}	a_{spd}	B_{acc} / $s_{z_{acc}}$	B_{spd} / $s_{z_{spd}}$	η	T_{er}	s_t	DIC	p_D
JF	Diffusion	.256	.061	.066	.006	.155	.245	.181	**-3478**	11.93
	LBA	.603	.215	.373	.116	.263	.107	-	-2293	**11.59**
KR	Diffusion	.249	.065	.023	.015	.153	.227	.152	**-3793**	**10.05**
	LBA	.615	.223	.383	.143	.341	.123	-	-1327	12.79
NH	Diffusion	.246	.086	.078	.003	.213	.259	.172	**-5938**	11.85
	LBA	.479	.251	.27	.121	.307	.129	-	-4870	**11.15**

was the case in the simulated example the decrease between the nominal and effective number of model parameters due to functional form complexity was greater for the diffusion (-2.7 on average) than the LBA (-1.2 on average). Overall, when applied to real data coming from a more complicated design, the diffusion model tended to require fewer "effective parameters" (11.3 on average) than the LBA model (11.8 on average). At the level of individual participants, however, we see that p_D was smaller for the LBA than the diffusion model for two out of three participants.

Discussion

DIC is a model selection criterion which attempts to select the model which is best able to predict new data. DIC, and p_D, a measure of model complexity, can be calculated from MCMC samples from the deviance of posterior parameter distributions. The p_D measure takes into account functional form complexity, and can be thought of as the effective number of parameters used to fit the data. When using data simulated from the diffusion model with a simple experimental design, the diffusion model, perhaps surprisingly, had a smaller p_D value than the LBA model. In other words, for our simulated data set the diffusion model was simpler than the LBA in terms of functional form complexity. When the models were fit to benchmark data from Ratcliff and Rouder (1998) which model was simpler differed between participants. For two out of three participants the LBA required fewer effective parameters. Averaging over participants, however, suggested the diffusion model was simpler.

There are a number of technical details associated with DIC and p_D should be addressed. Spiegelhalter et al. (2002) state that DIC and p_D are appropriate when: the distribution of posterior samples are approximately normal, and the model provides a reasonable account of the data. We have already addressed the second point, i.e. the posterior parameters were providing good predictions of data. In the models presented here the posterior distributions for each parameter closely approximate normal distributions, making it more likely that the joint distribution of these parameters are also approximately normally distributed. DIC and p_D are also dependent on the prior distribution used and the "focus" of our analysis. We have made an attempt to make these factors equivalent across models.

First, we used numerical integration of the Winbugs results for the diffusion model in order to equate the focus of inference for each model. The WinBUGS code given by Vandekerckhove et al. (2008) for the diffusion model implements start point variability and non-decision time variabilty hierachically –that is, by drawing a sample for each of these parameters for each trial performed by a participant on each MCMC iteration. This approch was necessitated because the Ratliff diffuison does not have an analytic likelihood when these sources of between-trial variability are included. In contrast, the WinBUGS code takes advantage of the LBA's mathematical simplicity by using an analytic expression for the likelihood of the LBA model which integrates out all forms of between-trial variability. This difference makes the deviances for each model produced by WinBUGS incommensurate; for the diffusion model this deviance focuses on the particular set of trials observed, whereas for the LBA the deviance is appropriate for the population of possible trials, and hence prediciton of performance by each subject performing new trials. As the latter focus is clearly more appropriate for our purposes we numerically integrated the deviance for each diffusion model posterior sample and used these integrated deviances to calcualte DIC and p_D.

Second, the prior distributions for diffusion model parameters are based on the range of parameter values estimated from all of the published diffusion fits found by Matze and Wagenmakers (submitted). Priors for LBA parameters were obtained from simulations which took the range of diffusion model parameters from Matze and Wagenmakers (submitted) and mapped them onto changes in LBA parameters. This gave a range of LBA parameters to be used as priors which may account for approximately the same range of patterns of data. In both cases the prior distribution of parameters was assumed uniform within these ranges. These prirors are informative not only in excluding parameters outside the allowed range, but also because the width of the range of allowed parameters determines the contribution made by the prior to the posterior deviance. A narrower range reduces posterior deviance and hence improves DIC. The large sample sizes that we examined means that the contribution of the prior is dominated by the likelihood of the data when determining parameter estimates within a model. However, this does not necessarily mean that differences in the prior for each model are not influential on the *difference* in posteior deviance between models, and hence DIC. In ongoing work we are implementing "vague" priors (i.e., priors with approximately equal probability across a very broad range of parameters for both models) in order to test the sensitivity of our results to the prior specification.

In summary, we have provided a relatively preliminary investigation into the complexity of models of choice and response time using a Bayesian model selection criterion. The criterion, DIC, and an associated measure of model complexity that takes into account differences in funcitonal form, p_D, are relatively easy to apply becasue it can be directly calculated based on MCMC samples from posterior model parameter distributions. If we consider simplicity as the range of potential data patterns which a model can predict, our results suggest that it may have been premature to claim that the LBA is the simplest model of choice and response time. Our results suggest that for these models a simple count of parameters will not suffice, and that more investigation is required. Functional form complexity based on prediction, however, is not the only aspect which might define a model's simplicity. For example, the mathematical tractability of the LBA, which enables analytic likelihoods to be derived, make it possible to more estimate parameters from data using even quite basic software, such as Microsoft Excel (Donkin, Averell, et al., 2009).

Although DIC has been found to be reliable (e.g. Myung et al., 2005), there are alternative approches to defining functional form complexity. For example, both DIC and Bayes factors adjust for complexity, but DIC emphasizes posterior prediciton whereas Bayes factors emphasize the selection of a true model. Different approches have different strengths and weaknesses. For example, DIC, like AIC, is inconsistent, so that as sample size increases it tends to select overly complex models. Bayes factors are less attractive in terms of prediciton becasue they asses the degree to which the prior rather than posterior predicts new data (Liu & Aitkin, 2008). As part of a larger project we are investigating the degree to which conclusions about complexity are robust over a range of such model selection measures (Myung & Pitt, 1997; Myung, 2000).

Acknowledgments

We acknowledge support from the Keats Endowment Fund.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

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

Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Manuscript Submitted for Publication*.

Donkin, C., Brown, S., & Heathcote, A. (2009). The overconstraint of response time models. *Manuscript Submitted for Publication*.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., et al. (2008). The striatum facilitates decision-making under time pressure. *Proceedings of the National Academy of Science, 105*, 17538–17542.

Ho, T., Brown, S., & Serences, J. (submitted). Domain general mechanisms of perceptual decision making in human cortex. *Journal of Neuroscience*.

Liu, C. C., & Aitkin, M. (2008). Bayes factors: prior sensitivity and model generalizability. *Journal of Mathematical Psychology, 52*, 362–375.

Matze, D., & Wagenmakers, E. J. (submitted). Psychological interpretation of exgaussian and shifted wald parameters: A diffusion model analysis. *Manuscript submitted for publications*.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*, 190–204.

Myung, I. J., Karabatsos, G., & Iverson, G. J. (2005). A bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology, 49*, 205–225.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4*, 79–95.

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

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

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

Ratcliff, R., Segraves, M., & Cherian, A. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two–choice decisions. *Journal of Neurophysiology, 90*, 1392–1407.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*, 408–424.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*, 438–481.

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.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2008). Hierarchical Bayesian diffusion models for two–choice response times. *Manuscript submitted for publication*.