Title

Does micro-variability make models more complex? A comparison between diffusive and linear evidence accumulation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Donkin, Christopher
Shiffrin, Richard
Brown, Scott
et al.

Publication Date
2010

Peer reviewed

Does micro-variability make models more complex? A comparison between diffusive and linear evidence accumulation

Chris Donkin and Richard M. Shiffrin
Department of Psychological and Brain Sciences, Indiana University
1101 E. 10th St, Bloomington IN 47405 USA

Scott Brown and Andrew Heathcote
School of Psychology, The University of Newcastle,
University Drive, Callaghan NSW 2308 Australia

Abstract

Most theories of how decisions are made assume that the accumulation of evidence from the environment is a noisy process. Recently, models have been proposed which do not have this micro-variability, and as a result are simple in the sense of being analytically tractable. We use a global model analysis method called landscaping to show that in terms of flexibility, simply removing micro-variability does not necessarily make a model more simple. Our landscaping also highlights an experimental design which might be helpful in discriminating between different response models.

Keywords: response time models; complexity; landscaping

A wide range of experimental psychology tasks involve a decision between two alternatives. Which alternative is chosen and the time taken to make that choice has been the subject of intense investigation. The most successful theories for the decision process usually come from a class of evidence accumulation (or sequential sampling) models. Evidence accumulation models assume that participants collect information from the environment to use as evidence as to which potential response is correct. Evidence is accumulated until there is enough to indicate that one of the responses should be given. This response is then made and the time taken for evidence accumulation makes up the decision time component of observed reaction time (RT). Though there are many models which follow this basic framework, the particular assumptions about evidence accumulation that each model makes varies considerably.

Historically, the collection of evidence from the environment has been modeled as a stochastic process (e.g. Ratcliff & Tuerlinckx, 2002; Usher & McClelland, 2001), such that how much evidence there is for a response varies randomly from moment-to-moment. For example, in a random walk process, the amount of evidence accrued between any two moments in time is a sample from a normal distribution.

A small number of recently proposed models, however, have demonstrated that it is not necessary to explicitly model the micro-variability in evidence accumulation (e.g. Reddi & Carpenter, 2000; Reeves, Santhi, & Decaro, 2005). Brown and Heathcote's (2008) Linear Ballistic Accumulator (LBA) model assumes that while a decision is being made, evidence accumulates at a fixed linear rate. Despite this lack of micro-variability the model provides a full account of benchmark choice and response time phenomena.

Brown and Heathcote (2008) proposed the LBA as a simple model of choice and RT because it makes few, and relatively basic, assumptions about how evidence accumulation occurs. Here we investigate a slightly different question, whether or not the LBA, with its lack of micro-variability, is a functionally simpler model. More generally, we aim to examine whether the addition of micro-variability necessarily increases the complexity of a model. Since Occam's Razor says that we should prefer the simplest and complete description of data, and models both with and without micro-variability have been shown to account for empirical data, our investigation may shed light on whether decision models need to necessarily assume micro-variability. We will use a technique called landscaping (Navarro, Pitt, & Myung, 2004) to assess complexity. First, however, we provide an overview of the diffusion and LBA models.

Overview of Models

The Diffusion Model

Consider a recognition memory task in which participants have been asked whether or not a stimulus currently presented was either previously studied, "old", or not studied, "new". A diffusion model account of this choice assumes 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 counter, shown by the irregular line in the right panel of Figure 1. Total evidence begins at some starting point and evidence that favors an "old" response decreases the evidence counter and evidence for a "new" response increases the counter. Evidence accumulation continues until the counter reaches one of the response boundaries, the horizontal lines in Figure 1. The choice made depends upon which boundary was reached, the top barrier for "new" and the bottom barrier for "old". The observed RT is the time taken for accumulation plus a non-decision time component made up of things such as encoding time and the time taken to make a motor response.

A key feature of the diffusion model is its micro-variability, such that the amount of evidence accumulated varies from moment-to-moment according to a normal distribution whose mean we call the *drift rate*. On top of this within-trial variability, there are typically three forms of between-trial variability added to the diffusion model. Drift rate and start point

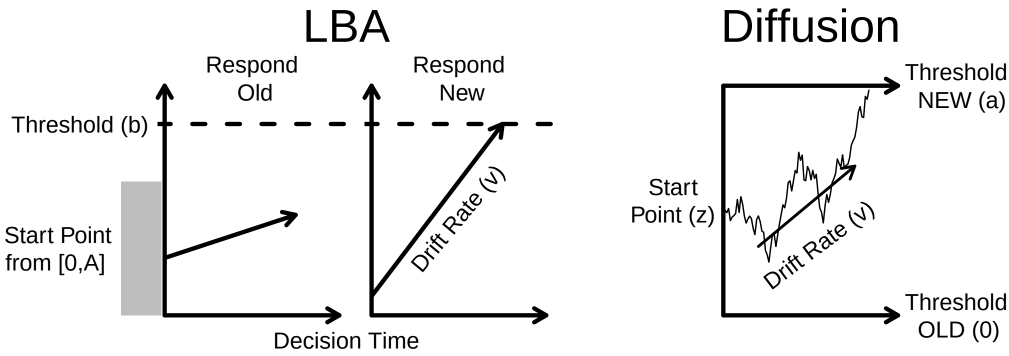

Figure 1: Overview of the diffusion and LBA models (left and right panel, respectively)

are generally assumed to vary from trial-to-trial according to a normal and uniform distribution, respectively. Finally, Ratcliff and Tuerlinckx (2002) included between-trial variability in non-decision time in the form of a uniform distribution.

The LBA Model

In the LBA there are separate accumulators gathering evidence for each of the "new" and "old" responses. As indicated by the straight lines in the left panel of Figure 1, these accumulators accrue evidence linearly and without micro-variability. Accumulation begins at some start point and continues until evidence in one accumulator reaches a response boundary. The accumulator which reaches the boundary first selects its associated response and predicted RT is accumulation time plus non-decision time. As in the diffusion model, the LBA also features between-trial variability. Like the diffusion model, drift rate and start point are assumed to vary between-trials according to normal and uniform distributions, respectively. Unlike the diffusion model, the LBA typically does not require between-trial variability in non-decision time to fit empirical data.

The Complexity of the Models

The LBA was considered by Brown and Heathcote (2008) as a relatively simple model because of its simpler assumptions about variability, and hence fewer parameters. However, recent work has demonstrated that the complexity of a model is not determined simply by the number of parameters in a model, but by how the parameters of the model interact within the model architecture to produce different patterns of predictions – also known as the functional form complexity of a model (e.g. Myung, 2000; Shiffrin, Lee, Wagenmakers, & Kim, 2008). Functional form complexity differs among models when they are able to produce differing ranges of predictions, even when they share the same number of parameters. In this way an overly complex model can provide an excellent fit to data, but not because the model gives a good account of the underlying process, but simply because of the model's flexibility. In particular, a more complex model can "overfit" the data by explaining the noise specific to a particular sample, as well as the structure due to the underlying processes. Because only the structure re-occurs in new data, overfitting limits the model's ability in terms of prediction.

There are many techniques for analyzing the complexity of a model (see Shiffrin et al., 2008 for a review). We will focus on one particular method proposed by Navarro et al. (2004) called landscaping. This method is highly related to parametric bootstrap methods proposed by Wagenmakers, Ratcliff, Gomez, and Iverson (2004). Landscaping, as a means of determining model complexity, is based on the idea that a more flexible model will be better able to mimic the predictions of an alternative model. Landscaping is used to compare the relative flexibility of any two models, and for our purpose these will be models with and without micro-variability (a diffusion and an LBA model, respectively). Note that landscaping tells us about a specific form of local, relative flexibility, rather than the model's general flexibility. In particular, landscaping tells us about how flexible one model is relative to another model, specifically for the regions of the parameter space in which we observe real data. In what follows we will refer exclusively to this local flexibility.

Landscaping

To do landscaping we generate data from one model, say model A, and fit these data with both models, i.e. model A and the alternative model, say model B. We then repeat the process with model B as the data-generating model. How well model B can fit the data generated by model A, and vice versa, gives insight into the relative flexibilities of both models. We will focus on two measures of model flexibility, the first is the difference between how well model B fits model A's data compared to model A, and the second is how often model B can better fit model A's data. The first measure tells us how flexible model B is compared to model A, i.e. if model B gives better fits to data from model A than vice versa, then model B is more flexible. The second measure tells us how distinguishable, or confusable, the two models are, i.e. how often we expect to have model B fit data better than model A when model A is actually the true model.

In all of our landscaping analyses we simulated 3200 data sets from each model. For each data set a random sample of parameters was chosen from uniform distributions whose ranges were determined by previously observed parameters estimated from real data. In particular, Matzke and Wagenmakers (2009) identified the range of parameter values previously estimated across all previous applications of the diffusion model to data. Donkin, Brown, Heathcote, and Wagenmakers (2009) used these values to identify a range of parameters values for the LBA which spanned the same range of data space. Note that parameters are sampled from each distribution independently and so may not reflect the correlations between parameters in real data.

Table 1: Range of parameter values used to generate data sets. Parameters not previously defined are as follows: T_{er} is non-decision time in both models, s and η represent between-trial standard deviation in drift rate in their respective models, and s_z and s_t represent the ranges of between-trial variability in start point and non-decision time in the diffusion model, respectively.

Model		$b-A$	A	T_{er}	s	v	
LBA	Min	0	.15	.1	.15	.5	
	Max	.5	.45	.4	.35	1	
		a	T_{er}	η	s_z	s_t	v
Diffusion	Min	.06	.3	.01	.01	.01	.01
	Max	.25	.6	.25	.08	.3	.5

Landscaping is known to depend on the design of the data simulated. Here we selected two commonly used designs, one in which only the difficulty of the task was manipulated, and one in which both difficulty and response caution were manipulated. To simulate a difficulty manipulation we used three conditions (easy, medium and hard) across which only the drift rate parameter of the model could change. In practice this meant that the distribution of drift rates shown in Table 1 was divided evenly into three smaller distributions, with the ease of the task increasing with drift rate. To simulate a caution manipulation we used the same procedure to create two conditions (speed emphasis and accuracy emphasis) across which only the response boundary parameter could change, i.e. boundary parameter distributions were divided in two and two values were sampled.

Micro-variability

In these first set of analyses we aim to investigate whether the micro-variability of the diffusion model makes it more flexible than a model without micro-variability, the LBA model. The models, however, differ in more ways than just micro-variability. In an attempt to make the models more similar, and hence make the effect of micro-variability more salient, we use a slightly simplified version of the standard diffusion model (cf. Ratcliff & Tuerlinckx, 2002) in which there is no between-trial variability in non-decision time. The models now share the same assumptions about between-trial variability – it is in both drift rate and start point of accumulation (but see the General Discussion for talk of other key differences between the models).

Difficulty Manipulation To create our landscape we first simulated data from both the LBA and the diffusion model. The data were simulated with all parameters except for drift rate fixed across three difficulty conditions, with 200 observations simulated per condition. We used 200 observations per condition because this amount is standard in applications of choice RT models. Both models used seven parameters for both simulating and fitting data – the diffusion model: a, T_{er}, η, s_z, v_{easy}, v_{medium}, v_{hard}, and the LBA: b, T_{er}, s, A, v_e, v_m and v_h. The simulated data were summarized using five quantiles (.1, .3, .5, .7 and .9) and both models were fit using quantile maximum probability estimation (Heathcote, Brown, & Mewhort, 2002) as the objective function and simplex as a search algorithm.

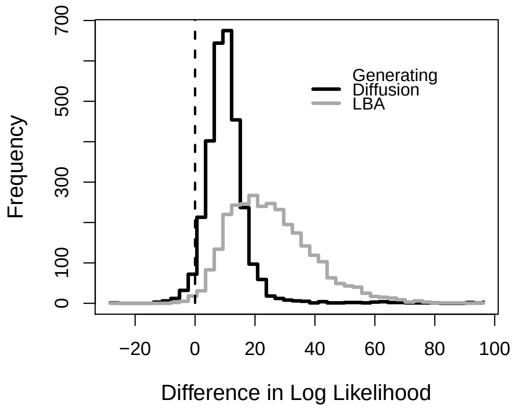

Figure 2: Difference in log-likelihood values between the data-generating model and the alternative model. The black and gray lines represent the diffusion and LBA as the generative models, respectively. The dotted line represents the point at which the data-generating and alternative models give equal quality fits, negative values indicate cases in which the alternative model fits better than the generative model. In this plot the simulated data come from a difficulty manipulation and the models used make the same assumptions about between-trial variability.

Figure 2 shows the difference in quality of fit between the generating and alternative model when the diffusion was the generating model (black histogram) and when the LBA was the generating model (gray histogram). Positive values indicate that the data-generating model fits better than the alternative model, and negative differences indicates that the alternative model is fitting the generating model's data better than the generating model itself. Two things are apparent from the figure: the gray histogram is generally more positive than the black histogram, and neither histogram has much mass in the region of negative differences. The first observation tells us that when the LBA was the generating model the diffusion tended to fit worse than how well the LBA fit when the diffusion was the generating model. In other words, the diffusion model appears to be less flexible than the LBA model in terms of how closely it can resemble the other model's data. The second observation tells us that neither model is very capable of better fitting the other model's data – the LBA fit data generated from a diffusion model better than the generating model in only 3.2% of the 3200 data sets, and the diffusion model better fit data generated from an LBA only 0.8% of the time.

Visual inspection of the fits suggested that the predictions of both models matched the simulated data closely, regardless of which model the data had come from. Indeed, in all but the most extreme cases, the models appeared to be mimicking each other closely. This suggests that the differences in log-likelihood we observe in Figure 2, and in all other figures, are not simply due to the models occupying completely separate data spaces, but reflect differences in the ability of one model to better fit the other model's data (i.e., what we define as model flexibility).

Caution and Difficulty Manipulations To create the landscape for a design in which both caution and difficulty were manipulated we simulated data in which all parameters except for drift rate were fixed across the three difficulty conditions and all parameters except for response boundary were fixed across the two caution conditions. Fits were as in the previous landscape except that each model now had eight parameters – the diffusion: a_{speed}, $a_{accuracy}$, T_{er}, η, s_z, v_e, v_m, v_h, and the LBA: b_s, b_a, T_{er}, s, A, v_e, v_m and v_h. Landscapes were created using both 200 observations per condition (as in the previous landscape), as well as 100 observations per condition (since twice as many conditions meant that total sample size was twice that of the previous landscape). Sample size had little effect on the pattern of results, but the smaller sample size did lead to slightly more confusion between the models. We present, therefore, the results of the landscape using the smaller sample size (i.e. where total sample size was equated across landscapes).

A quick look at Figure 3 suggests that the current landscape is similar to the one where only difficulty was manipulated. Closer inspection, however, reveals two differences: Firstly, the histograms in Figure 3 show a larger mean and variance than those in Figure 2, and secondly, the histograms show even less mass below zero. The first observation suggests that when both caution and difficulty are manipulated that both models are not as good at accounting for the alternative model's data. Note, however, that the relative position of the black and gray histograms continue to suggest that the diffusion model has less flexibility than the LBA. The second observation implies that the models are even more distinguishable when both caution and difficulty manipulations

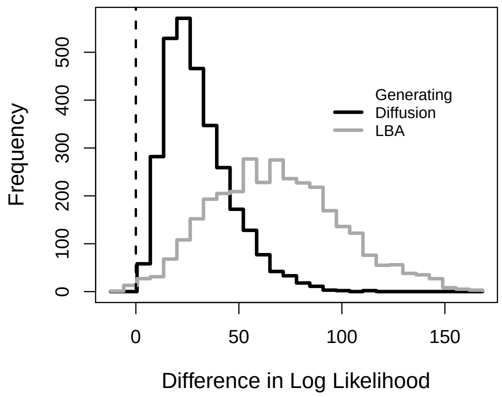

Figure 3: Difference in log-likelihood values between the data-generating model and the alternative model. The data come from a caution and difficulty manipulation, and the models make the same assumptions about between-trial variability.

are made – in 3200 data sets, the LBA never better fit data generated from a diffusion model, while the diffusion model better fit data from an LBA only 0.4% of the time.

Discussion Our first measure of flexibility, the relative shapes and positions of the histograms in our figures, suggest that the LBA is capable of getting better fits of data generated from a diffusion model than vice versa. We take this to mean that the LBA model is more flexible than our simplified version of the diffusion model (i.e. one without non-decision time variability). Since the models were equated on assumptions about between-trial variability, we also take this result as evidence against the idea that the micro-variability in the diffusion model makes the model more flexible than the model without micro-variability, the LBA. Indeed, there may be evidence to suggest the opposite – that micro-variability reduces the functional form complexity of a model. We do not mean our results as conclusive evidence of such a result, however, particularly because micro-variability is not the only difference between the LBA and diffusion models. We direct the reader to our General Discussion for suggestions of how the effects of micro-variability could be more investigated more specifically.

Our second measure of flexibility, how often the alternative model can better fit data from the generating model, gives a less clear result. This is largely because both models seem relatively incapable of better capturing the other model's data, at least for the sample size we use. When we repeated our landscaping analysis with a greatly reduced sample size (just 20 observations per condition) we observed an interesting result, consistent with our first measure of flexibility – the LBA better fit diffusion data in almost one in ten samples, while the diffusion still only better fit LBA data in less than one in two hundred samples. The results reported in Figure 3, however, suggest that the two models are distinguishable based on fit alone for the types of sample sizes typically used. In other words, in the unlikely case that one of the two models was truly responsible for empirical data, then our results suggest that the alternative model would rarely be mistakenly chosen as the best fitting model, provided at least 100 observations were recorded per condition. However, this result is not very useful since we do not believe that a diffusion model without between-trial variability is appropriate. We now repeat our landscaping using a diffusion model *with* between-trial variability in non-decision time, paying particular focus as to whether or not the models remain distinguishable.

Comparing the LBA and the Full Diffusion

The method for creating the following two landscapes was the same as for the previous two landscapes, however, between-trial variability in non-decision time was assumed for the diffusion model (but not the LBA).

Difficulty Manipulation Figure 4 suggests that a full diffusion model may be slightly more flexible than the LBA when only difficulty is manipulated. In particular, though largely overlapping, the grey histogram looks like a slightly left-shifted version of the black histogram, suggesting that the difference between quality of fit for the data-generating and alternative models was smaller when the LBA generated the data. In other words, the diffusion model was slightly better able to fit LBA data than vice versa. When we look at just the cases in which the alternative model fits better than the data-generating model we see that the same pattern continues, the LBA model better fits data simulated from a diffusion model in 6% of simulated data sets, while the diffusion model better fits LBA data 10% of the time.

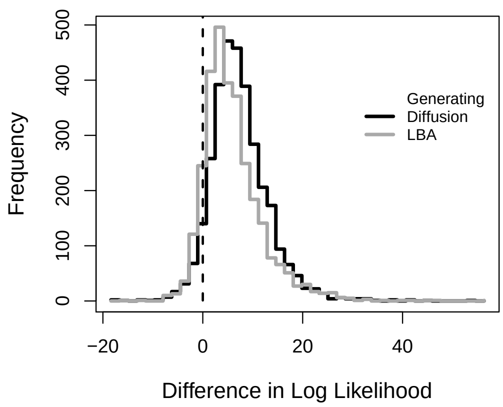

Figure 4: Difference in log-likelihood values between the data-generating model and the alternative model. The data come from a difficulty manipulation, and the diffusion model makes the additional assumption that non-decision time has between-trial variability.

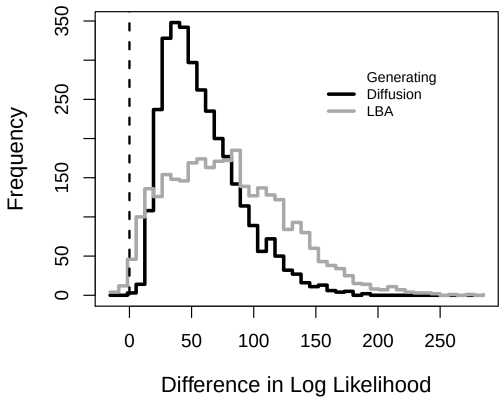

Figure 5: Difference in log-likelihood values between the data-generating model and the alternative model. The data come from a caution and difficulty manipulation, and the diffusion model makes the additional assumption that non-decision time has between-trial variability.

Caution and Difficulty Manipulations The landscape in which both caution and difficulty were manipulated was created using 200 simulated data points in each of the six conditions. From Figure 5 it is not clear which of the full diffusion model or the LBA is more flexible. In particular, the grey histogram has more mass than the black histogram at both very small and very large positive values, suggesting that the diffusion model fit LBA data both very well and very poorly. In terms of how often the alternative model fit better than the data-generating model, when the diffusion model was the generative model then the LBA never fit better, while the diffusion model fit LBA data better in only 0.8% of the simulated data sets.

Discussion The first two landscapes we created suggested that the LBA and the diffusion models were distinguishable, such that each model was relatively incapable of better fitting the other model's data. These second pair of landscapes looked at whether these results extended to the full diffusion model (with between-trial variability in non-decision time). The first landscape we created suggested that this might not be the case. When data came from a design in which only difficulty, i.e. drift rate, varied then both models displayed some reasonable mimicry, such that the LBA looked more like a diffusion model in 6% of the simulated data sets and the diffusion model looked more like an LBA 10% of the time. These proportions are not overly large, but they do suggest that if one of the models actually was the true model, that we would observe the alternative model fitting data better for about one in ten to twenty participants.

The results of our fourth and final landscape suggest that the models become highly distinguishable when both difficulty and caution are manipulated. Indeed, the results suggest that if one of the two models were the true model then the alternative model would be mistaken as the best fitting model for fewer than one in a hundred participants. The difference in distinguishability between these two final landscapes is remarkable, however it is possible that the difference occurs because there are twice as much data under the design with both caution and difficulty manipulations. Equating total sample size using a simplified diffusion model, however, had little effect on distinguishability – doubling sample size meant that the largest confusion occurred 0.4% of the time instead of 0.2%. We expect, therefore, that it is something about the design rather than sample size which causes such a large change in distinguishability. Consistent with this idea, Donkin et al. (2009) showed that the boundary parameters of the LBA and diffusion model do not have a similar effect on model predictions. These results further cement the idea that the key to distinguishing between these two models may lie in the differential effect of manipulating the response boundary parameter in each of the models.

General Discussion

We compared the flexibility of the LBA model, which contains no micro-variability in evidence accumulation, with a simplified version of the diffusion model, which does contain micro-variability. Our results suggest that micro-variability does not necessarily make a model more flexible than one without micro-variability. We can not, however, confidently conclude that micro-variability does not increase flexibility at all. This is because the LBA and the diffusion model, even a simplified version without between-trial variability in non-decision time, do not have identical frameworks. In particular, the LBA has multiple, independent, accumulators while the diffusion has a single accumulator, which implies that evidence for one response is perfectly negatively correlated with evidence for the alternative response. Without further investigation, we can only confidently conclude that a ballistic multiple-accumulator framework gives the LBA more flexibility than a stochastic single-accumulator framework gives the diffusion. Further investigation into the effects of micro-variability might directly compare a multiple accumulator framework with and without micro-variability (e.g. the LBA compared to a simplified version of Usher and McClelland's, 2001, model). Such a study will be more difficult than that carried out here because analytic expressions do not exist for Usher and McClelland's model.

Our final two landscapes compared the relative flexibility of the LBA model and the full Ratcliff diffusion model. Most impressive here was the increase in distinguishability which arose out of the inclusion of a caution manipulation. This increase is quite remarkable, when only difficulty was manipulated the models show the largest overlap of any landscape we analysed, but when caution is added there are almost no cases in which a model fit data better than the data-generating model. The distinction between models may arise because the effect of changing caution (i.e., boundary separation parameters) is different for the two models (Donkin et al., 2009). In particular, micro-variability in processing means that some responses will terminate quickly regardless of the position of response boundaries. Without this micro-variability, however, increasing caution will slow down even the fastest responses. This means that changing caution in the diffusion model effects the speed of the fastest responses much less than in the LBA. Regardless of the cause, our results suggest that whichever model can better account for a combined caution and difficulty manipulation is probably closer to the true model, and unlikely to be due to model mimicry.

References

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

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2009). Diffusion versus linear ballistic accumulation: Different models for response time, same conclusions about psychological mechanisms? *Manuscript submitted for publication*.

Heathcote, A., Brown, S. D., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*, 394–401.

Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of exgaussian and shifted wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*, 798–817.

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

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, *49*, 47-84.

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.

Reddi, B. A. J., & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*, *3*, 827–830.

Reeves, A., Santhi, N., & Decaro, S. (2005). A random–ray model for speed and accuracy in perceptual experiments. *Spatial Vision*, *18*, 73–83.

Shiffrin, R. M., Lee, M. D., Wagenmakers, E.-J., & Kim, W. J. (2008). A survey of model evaluation approaches with a focus on hierarchical bayesian methods. *Cognitive Science*, *32*, 1248–1284.

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

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology*, *48*, 28–50.