A Bayesian Approach to Diffusion Process Models of Decision-Making

Joachim Vandekerckhove (joachim.vandekerckhove@student.kuleuven.be)
Department of Psychology, Tiensestraat 102
University of Leuven, 3000 Leuven, Belgium

Francis Tuerlinckx (francis.tuerlinckx@psy.kuleuven.be)
Department of Psychology, Tiensestraat 102
University of Leuven, 3000 Leuven, Belgium

Michael D. Lee (mdlee@uci.edu)
Department of Cognitive Sciences, 3151 Social Sciences Plaza
University of California, Irvine, CA 92697, USA

Abstract

The Wiener diffusion model, and its extension to the Ratcliff diffusion model, are powerful and well developed process accounts of the time course of human decision-making in two-choice tasks. Typically these models have been applied using standard frequentist statistical methods for relating model parameters to behavioral data. Although this approach has achieved notable successes, we argue that the adoption of Bayesian methods promises to broaden the scope of the psychological problems the models can address. In a Bayesian setting, it is straightforward to include linear, non-linear, and categorical covariates of the basic model parameters, and so provide a much richer characterization of individual differences, the properties of stimuli, the effects of task instructions, and a range of other important issues. In this paper, we provide an example of the Bayesian possibilities by applying the Ratcliff diffusion model to a benchmark data set involving a brightness discrimination task. We simultaneously use a categorical covariate and nonlinear regression to model the psychophysical function in a theoretically satisfying way. We also use Bayesian inference on latent class assignment variables to identify and accommodate contaminant data at the level of individual trials, categorizing them as 'diffusion' trials, 'guesses', and 'delayed startup' trials. Using our application as a concrete example, we discuss the potential benefits of applying the Bayesian framework to process models in the cognitive sciences.

Keywords: Diffusion model; Wiener diffusion; Bayesian inference

Introduction

One area of the cognitive sciences that has many formal models is that of choice reaction time (RT), particularly when the number of choices is restricted to two (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). The practical application of many of the available models, however, has historically been hampered by computational difficulties (e.g., Vandekerckhove & Tuerlinckx, 2007). This is particularly the case for one prominent class of models based on diffusion processes, including the Wiener diffusion model (Link & Heath, 1975) and its popular extension, the Ratcliff diffusion model (Ratcliff, 1978; Wagenmakers, 2008).

For the latter model, several pieces of software have been published to aid in fitting these models to data (Vandekerckhove & Tuerlinckx, in press; Voss & Voss, 2007). Nevertheless, the application of statistical models to the diffusion parameters with these programs is, at present, restricted to the application of linear constraints (such as ANOVA or polynomial regression). The spectrum of possible applications of the diffusion model is much broader than that. In this paper we employ psychophysical curve fitting and latent class assignments in a Bayesian[1] treatment of the diffusion model, as an example of how Bayesian methods can broaden the class of psychological problems diffusion models can address.

The structure of the paper is as follows. We first describe the Wiener and Ratcliff diffusion models as process accounts for two-choice RT. We then report an example diffusion model analysis using Bayesian methods—based on previously studied data relating to a brightness discrimination task—that would be highly challenging to implement in a classical frequentist context. We also demonstrate using this example that the Bayesian approach can be successfully applied to relatively small sample sizes. Finally, we discuss the power and generality of the Bayesian approach for extending the potential of process models in the cognitive sciences.

Diffusion Models

The Wiener Diffusion Model

The Wiener diffusion model as a process for speeded decisions starts from the basic principle of *accumulation of information* (e.g., Link & Heath, 1975). When an individual is asked to make a binary choice on the basis of an available stimulus, the assumption is that evidence from the stimulus is accumulated over (continuous) time and a decision is made as soon as an upper or lower boundary is reached. Which boundary is reached determines which response is given, and the number of accretion steps taken is related to the RT.

Figure 1 depicts the diffusion process, and shows the main parameters of the process. On the vertical axis there are the *boundary separation* a, indicating the level of evidence required to make a response (i.e., speed-accuracy trade-off) and the *starting point* z_0, indicating the a priori status of the evidence counter. The arrow represents the average rate of information uptake or *drift rate* ξ, which indicates the average amount of evidence that the

[1] We want to emphasize that we are using Bayesian methods as a framework for statistical inference, and *not* as a set of theoretical assumptions about how humans make inferences. This means we are not proposing a 'rational' or 'computational-level' model of cognition, despite our reliance on Bayesian methods of inference.

$$\text{Wiener}_{X,T}(x,t|a,t^{er},b,\xi) = \begin{cases} \frac{\pi s^2}{a^2}\exp\left(-\frac{ab\xi}{s^2}\right)\sum_{j=1}^{+\infty} j\sin(bj\pi)\exp\left[-\frac{t-t^{er}}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2 j^2 s^2}{a^2}\right)\right] & \text{if } x=0 \\ \frac{\pi s^2}{a^2}\exp\left(-\frac{(1-b)a\xi}{s^2}\right)\sum_{j=1}^{+\infty} j\sin((1-b)j\pi)\exp\left[-\frac{t-t^{er}}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2 j^2 s^2}{a^2}\right)\right] & \text{if } x=1 \end{cases} \quad (1)$$

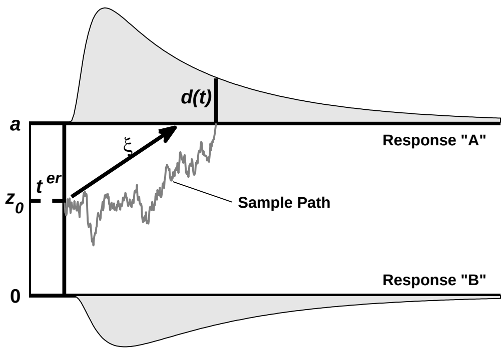

Figure 1: A graphical illustration of the diffusion model. Note that $z_0 = a \times b$. In the Ratcliff diffusion model, b, t^{er}, and ξ vary from trial to trial. The probability density for a correct response given at time t is shown as $d(t)$.

observer receives from the stimulus at each sampling. Finally, the short dashed line indicates the *nondecision time* t^{er}, the time used for everything except making a decision (i.e., encoding the stimulus and physically executing the response).

It is important to note that, considering the Bayesian statistical context of this article, it will be more convenient to use a different parametrization of the process. We will therefore not consider the starting point z_0, but rather use the *initial bias* b, defined as $b = z_0/a$. With these parameters, the joint probability distribution of the RT and accuracy (i.e., the likelihood function) is Equation (1) at the top of the page.

The Ratcliff diffusion model

Despite the elegance of the basic Wiener process as an account of the time course of decision-making, the evolution of diffusion models has involved a series of additional assumptions. These have all been intended to address shortcomings in the ability of the basic model to capture empirical regularities observed in data from human decision-making experiments.

One important change has been the introduction of additional noise processes to capture cross-over effects. 'Cross-over effects' refer to the observation that errors can sometimes be, on average, faster than correct decisions, but other times are as slow or slower. These possibilities are not accommodated by the basic model in Figure 1 without allowing for variation in the parameters. Accordingly, to predict fast errors, the basic model is extended by assuming that the starting point is subject to between-trial variation, and so is convolved with a mixing distribution. Similarly, to predict slow errors, it is assumed that the mean drift rate is also subject to between-trial variation, and so is convolved with a Gaussian distribution.

Additionally, for empirical reasons the nondecision time is assumed to vary from trial to trial, usually according to a uniform distribution. These three noise processes are parameterized with the standard sufficient statistics (mean and variance of a Gaussian or mean and range of a uniform), which become additional parameters of the model. When the Wiener diffusion model is extended with trial-to-trial variabilities such as these, it is often called the Ratcliff diffusion model (Wagenmakers, 2008). This extended model comes with a much greater computational burden, see Tuerlinckx (2004) and Vandekerckhove and Tuerlinckx (2007).

Notation In this paper, we will use X and T to refer to the accuracy and RT variables, and x and t for specific instances of these variables. We will sometimes write T^* and t^* to refer to response vectors (X,T) and (x,t), respectively. We will use indices i ($i = 1,\ldots,I$) and j ($j = 1,\ldots,J$) to indicate conditions and k ($k = 1,\ldots,K$) for trials within conditions. To indicate a vector, we will use a bold font, so that $\mathbf{a}$ is the matrix of boundary separations, in all conditions and all trials. We use the symbol $\sim$ to denote "is distributed according to", so that $t^*_{ijk} \sim \text{Wiener}_{X,T}(a_{ijk}, t^{er}_{ijk}, b_{ijk}, \xi_{ijk})$ and the proportionality symbol $\propto$ to denote "is proportional to".

Application to Benchmark Data

To illustrate the advantages and the potential of approaching diffusion models from a Bayesian perspective, we revisit a benchmark data set (Ratcliff & Rouder, 1998). In addition to fitting five parameters of the Ratcliff diffusion model[2], we also perform a non-linear regression and a latent class assignment.

Data Set

In the experiment by Ratcliff and Rouder (1998), there were two manipulations of interest. First, there was a speed-accuracy instruction (participants were either instructed to be fast or to be accurate) and second, there was a manipulation of brightness. The task was a 2AFC procedure, whereby each participant was shown a stimulus and had to judge whether this stimulus was drawn from a 'bright' distribution or from a 'dark' distribution (the two distributions overlapped significantly, so subjects could not be highly accurate; in total, there were 33

[2] We assume an unbiased diffusion process, so that $b = 0.5$.

different levels of brightness, '1' being completely dark and '17' being completely ambiguous). Feedback was given after each trial. There were three participants (labeled KR, JF, and NH), and the experiment ran over the course of 11 days. After preprocessing[3], there were a varying number of trials in each cell of the design, but the total was around 8,000 for each participant.

From the manipulations, we can expect two things. First, we expect that the speed-accuracy instruction will have an effect on boundary separation. Secondly, we expect that the brightness of the stimulus influences the drift rate. Furthermore, it is likely that the data set will contain at least some contaminant data, which we define as data points that are not generated by the process of interest and are hence not completely germane to the research question. In line with previous work, we will consider two types of contaminants: *guesses* and *delayed startups* (Vandekerckhove & Tuerlinckx, 2007).

Bayesian Modeling

We implemented a Bayesian analysis of the brightness discrimination task data using the graphical model presented in Figure 2. Graphical models (see Griffiths, Kemp, & Tenenbaum, in press; Lee, 2008, for psychological introductions) are a convenient language for describing the probabilistic relationship between parameters and data. In a graphical model, variables of interest are represented by nodes in a graph, with children depending on their parents. Circular nodes represent continuous variables, square nodes discrete variables, shaded nodes observed variables, and unshaded nodes unobserved variables. In addition, plates enclose parts of a graph to denote independent replication.

An important practical advantage of adopting the graphical model formalism is that it allows our modeling to be implemented using WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000). This makes it straightforward to perform full Bayesian inference computationally, using standard MCMC methods to sample from the posterior distribution.

We now explain the graphical model in Figure 2, highlighting the way in which it addresses important psychological problems, including accounting for contaminants in data, relating the physical and psychological properties of stimulus, and allowing for trial-to-trial variability in performance.

Latent Classes The Bayesian approach makes it easy to apply latent predictors to data. In the model in Figure 2, we have assumed that there are three types of experimental trials: (1) Diffusion trials (with probability $1-\pi$), (2) guesses (probability $\pi(1-\gamma)$), and (3) delayed startups (probability $\pi\gamma$). A similar distinction was applied by Vandekerckhove and Tuerlinckx (2007). This

[3]We applied similar preprocessing as Ratcliff and Rouder (1998), removing all trials from the first day, the first 20 trials of the other days, and the first trial of each block. In contrast to their analysis, we did not remove any trials based on an extreme RT.

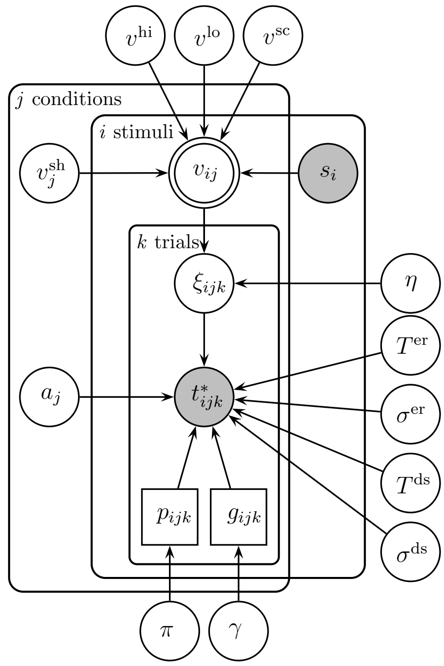

Figure 2: Graphical model representation of our Bayesian analysis of the Ratcliff diffusion model against the benchmark brightness discrimination data.

categorical distinction is latent because we have no direct measures of class membership. Most powerfully, the Bayesian approach allows us to estimate each trial's probability of membership to each of these (mutually exclusive) classes, so that we can identify specific trials that might be contaminants. Class memberships are indicated by two binary variables, $p_{ijk} \sim \text{Bernoulli}(\pi)$ and $g_{ijk} \sim \text{Bernoulli}(\gamma)$.

Applying this latent class assignment implies that we assume that three distinct psychological processes account for the data. The first is a typical diffusion process. The second is a diffusion process devoid of relevant information (a *guess*); That is, the participant has not gained any information from the stimulus and the response is therefore at chance level. In terms of diffusion model parameters, this translates to the assumption that all $\xi_{ijk} = 0$ if $p_{ijk}(1-g_{ijk}) = 1$. The third psychological process is one of *delayed startups*, where trials have a different non-decision time.

The Non-linear Regression of Drift Rate In psychophysics, it is common to perform nonlinear regression to model the effect of stimulus dimensions on accuracy, often using a Weibull link function. Because it seems natural that drift rates have similar asymptotic behavior as a function of stimulus intensity, our model applies a Weibull. For the i^{th} brightness condition ($i = 1, \ldots, 33$) and j^{th} speed/accuracy instruction, then

$$v_{ij} = v^{lo} + \left(v^{hi} - v^{lo}\right) \times \left(1 - \exp\left[-\left(i/v^{sc}\right)^{v_j^{sh}}\right]\right). \quad (2)$$

Note that we allow the shape parameter v^{sh} to be different between the speed/accuracy conditions. This is contrary to Ratcliff and Rouder (1998), who assumed mean drift rates to be equal for equal stimulus intensities.

Variability in Performance In order to extend the Wiener distribution to the Ratcliff diffusion model, the graphical model in Figure 2 implements a mixed-model version of the Wiener distribution. This means that, from trial to trial, some parameters are conditionally independent draws from a mixing distribution. By conceptualizing the Ratcliff diffusion model in this way, we can avoid the computationally intense integrals described in Tuerlinckx (2004), and approximate the integrals using standard MCMC computational methods used to integrate over the posterior. The simplification offered by this approach allows us to choose theoretically plausible mixing distributions, so we choose a Gaussian mixing distribution for drift rate and a truncated Gaussian for nondecision time.

This combination of assumptions in the graphical model can be formally stated as follows:

$$\begin{array}{ll} \text{if } p_{ijk}=0 & \begin{cases} t^*_{ijk} \sim \text{Wiener}(a_j, t^{er}_{ijk}, a_j/2, \xi_{ijk}) \\ \xi_{ijk} \sim \text{N}(v_{ij}, \eta^2) \\ t^{er}_{ijk} \sim \text{TN}_{(0,+\infty)}\left(T^{er}, (\sigma^{er})^2\right) \end{cases} \\ \text{if } \begin{array}{l} p_{ijk}=1 \\ g_{ijk}=1 \end{array} & \begin{cases} t^*_{ijk} \sim \text{Wiener}(a_{(j)}, t^{ds}_{ijk}, a_j/2, \xi_{ijk}) \\ \xi_{ijk} \sim \text{N}(v_{ij}, \eta^2) \\ t^{er}_{ijk} \sim \text{TN}_{(0,+\infty)}\left(T^{ds}, (\sigma^{ds})^2\right) \end{cases} , \\ \text{otherwise} & \begin{cases} t^*_{ijk} \sim \text{Wiener}(a_j, t^{er}_{ijk}, a_j/2, 0) \\ t^{er}_{ijk} \sim \text{TN}_{(0,+\infty)}\left(T^{er}, (\sigma^{er})^2\right) \end{cases} \end{array}$$

where N and TN stand for normal and truncated normal distributions, respectively.

Results

All of our analyses are based on $40,000$ posterior samples collected after a burn-in of $10,000$ samples. First, we investigate recovery of the model by inspecting posterior predictive samples (we limit ourselves to data of participant KR, but results were similar for the others). The two panels in Figure 3 show the proportion of 'bright' responses in the data (open circles) and as recovered by the model (grey dots; the full line connects the mean predictions). Similarly, in Figure 4, we show posterior predictives of the 10^{th}, 30^{th}, 50^{th}, 70^{th}, and 90^{th} percentiles of each RT distribution. In all panels, it is clear that the model recovers the patterns in the data quite well. The exception is the 10^{th} RT percentile in the speed condition, which the model consistently overestimates. This may be due to our restriction that $b = .5$ and does not vary from trial to trial.

Looking at the posterior means and standard deviations for the standard Ratcliff diffusion model parameters (for participant KR) in column A of Table 1, we see that the boundary separation parameter a is much smaller in the speed-instruction condition ($a_1 = 0.05$), as expected. The Weibull asymptote parameters, as well as

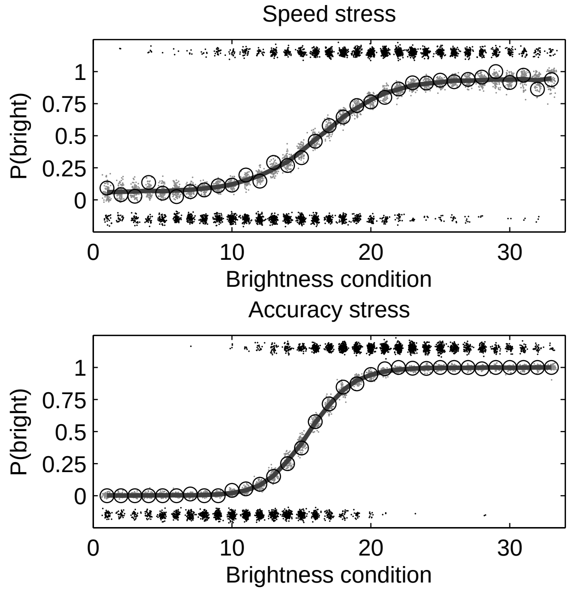

Figure 3: Posterior predictive proportions of 'bright' responses, as a function of stimulus intensity. Grey dots indicate 100 posterior samples, open circles indicate the proportions in the data set. Black dots at top and bottom indicate observed data points (jittered). Thick dark lines connect the posterior mean estimate of the response probabilities in each condition.

the scale parameter, get sensible mean posterior values. Interestingly, the shape parameter is somewhat different between the two instruction conditions, with steeper Weibull functions in the speed-stress condition.

Figure 5 shows posterior distributions of the π, γ, and v^{sh}, for each participant. The difference in v^{sh}s is small in two participants, but large for JF, and it seems consistent between participants.

Participant KR has the highest π parameter—the posterior mean is about .006. Looking at this participant's γ parameter, we see that there is much uncertainty regarding the proportion of guesses (because this parameter pertains to only .6% of the data—43 trials), but there are likely more delayed start-ups (28) than guesses (15). The delayed start-up trials are on average 858 ms slower than regular trials. If we compare the first two columns in Table 1, it appears that accounting for contaminants in this data set makes little difference for the mean estimates of the parameters. The posterior uncertainty of the drift-related parameters, however, is higher in the more complicated contaminant-mixture model.

Sample Size

Typically, applying the Ratcliff diffusion model requires "a fair amount [sic] of data for accurate estimation of its parameters" (Wagenmakers, 2008). By constraining parameters across conditions and using a Bayesian approach with modern computational sampling methods,

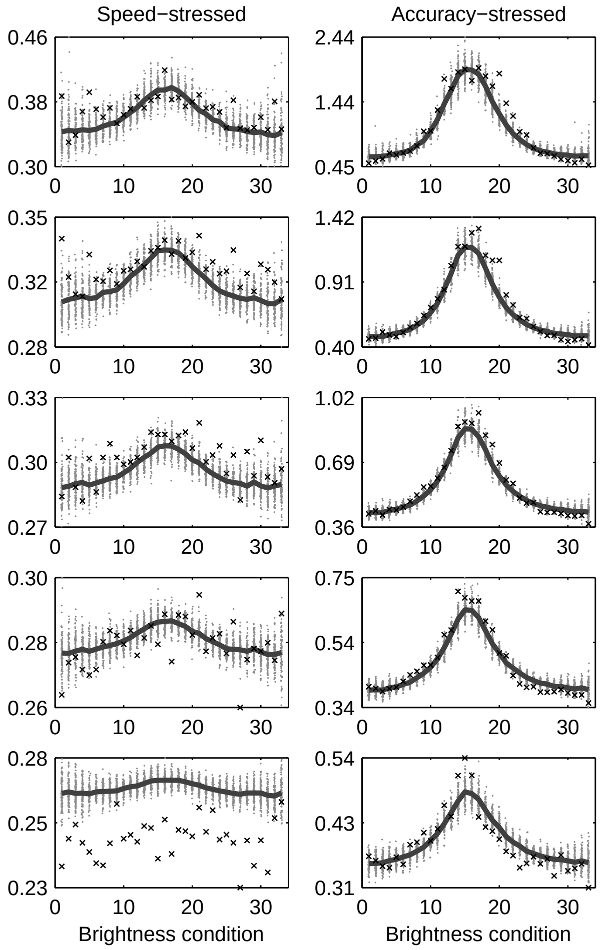

Figure 4: Posterior predictive RT percentiles. Left panels are for the speed-stressed condition. The first through fifth row are for the 90^{th}, 70^{th}, 50^{th}, 30^{th}, and 10^{th} percentiles, respectively. The values of the RT quantiles are shown on the vertical axes. Grey dots indicate 100 posterior samples, the thick dark lines connect the posterior means. Empirical percentiles are shown by crosses.

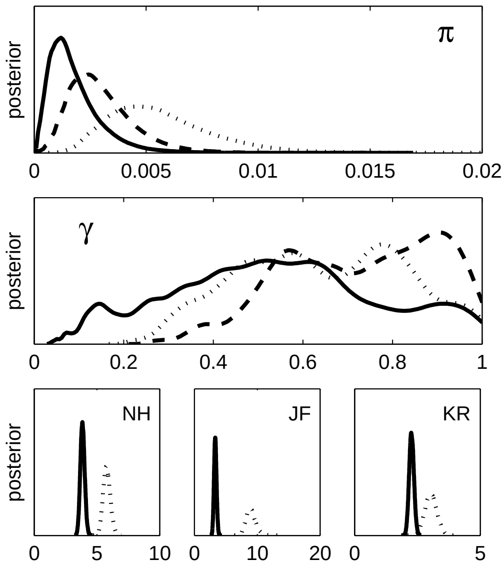

Figure 5: Posterior density plots for parameters π, γ, and v_j^{sh}. In the top two panels (π and γ), line styles indicate participants (KR: dotted, JF: dashed, NH: full). In the bottom panels, v_1^{sh} (dotted, for the speed condition) and v_2^{sh} (full line, for the accuracy condition) are shown for each participant separately. Significant differences are visible.

we expect the need for large data sets to be alleviated. To test this possibility, we conducted analyses based on subsampling from the benchmark data, and comparing the results with the results from the full data set.

To subsample from the original data set, we sampled—without replacement—either 2, 5, 10, or 20% of the data points for participant KR; thus approximately preserving the relative number of data points in each condition. We then applied a model that is similar to the one described in the previous section (see Fig. 2), but we leave out the contaminant modeling ($\pi = 0$) because of the low proportions of contaminants found. We drew 5,000 samples from the joint posterior, after a burn-in of 5,000. For each parameter, we compute the posterior mean. This procedure was repeated 20 times for each proportion, with new subsamples each time. Then, with the mean posterior estimates resulting from each of the smaller data sets, we computed the squared relative bias R^2 for each parameter: $R^2_\theta = \left((\theta - \hat{\theta})/\theta\right)^2$, where θ indicates the parameter as estimated from the full data set (with the same model containing no contaminant component) and $\hat{\theta}$ as estimated from the smaller data set. We summed the R^2s for each parameter set to obtain a measure of how close the recovered parameters of each subsample were to the ones found from the full data set. From each downsampling proportion (2, 5, 10, and 20%), we then chose the results that gave the median recovery under the R^2 criterion, and report those results in Table 1.

As can be seen, most of the estimates from the reduced data sets are very similar to those inferred from the full data set, and they certainly preserve all of the important order relations and trends in the parameter values across conditions. With few data, posterior uncertainty is very large. As expected from statistical theory, the posterior standard deviations scale up with a factor $\sqrt{N_t/N_s}$, where N_t is the total sample size and N_s the size of the subsample.

Conclusions

In this paper, we demonstrated a Bayesian extension of the popular Ratcliff diffusion model. In a single example, we combined a psychophysical link function and latent class assignment to revisit the benchmark data set of Ratcliff and Rouder (1998). As part of the Bayesian method, we employed posterior predictive checks (shown in fig-

Table 1: Some results for participant KR. Posterior means in top half; Posterior standard deviations in bottom half. SDs have been multiplied by 100.

	A^*	B^*	2%	5%	10%	20%
T^{er}	0.25	0.25	0.23	0.26	0.25	0.25
a_1	0.05	0.06	0.07	0.05	0.05	0.05
a_2	0.21	0.21	0.20	0.22	0.24	0.23
η	0.11	0.12	0.08	0.14	0.16	0.16
σ^{er}	0.03	0.03	0.04	0.04	0.03	0.04
v^{hi}	0.59	0.57	0.43	0.67	0.77	0.71
v^{lo}	-0.55	-0.53	-0.63	-0.74	-0.62	-0.60
v^{sc}	0.57	0.56	0.48	0.56	0.63	0.60
v_1^{sh}	3.02	3.07	4.89	2.39	2.20	2.75
v_2^{sh}	2.26	2.33	1.83	1.87	1.95	2.32
T^{er}	0.14	0.14	1.11	0.73	0.42	0.34
a_1	0.08	0.08	0.67	0.41	0.25	0.20
a_2	0.32	0.32	2.10	1.56	1.18	0.89
η	0.69	0.66	5.07	3.40	2.18	1.83
σ^{er}	0.08	0.08	0.70	0.36	0.25	0.19
v^{hi}	3.72	2.88	8.73	12.75	8.71	9.45
v^{lo}	1.95	1.80	12.83	8.66	6.70	4.95
v^{sc}	1.34	1.04	3.22	4.89	3.59	3.09
v_1^{sh}	24.84	22.80	241.43	71.83	40.83	54.36
v_2^{sh}	9.77	9.40	49.30	29.10	25.52	24.62

* *A* is with outlier treatment; *B* is without outlier treatment.

ures 3 and 4) of the model. We found that few of the data points are contaminants. Interestingly—and in contrast to previous analyses—we also found differences in drift rate as an effect of task instruction. In particular, drift rate as a function of stimulus quality increases more steeply under speed-stress than it does under accuracy-stress. In addition, we reported a simple numerical experiment that showed that relatively small samples can yield satisfactory parameter estimates. This suggests that the Ratcliff diffusion model may, using Bayesian methods, be applied to smaller data sets than was previously practicable.

We think many of the specific demonstrations in our example correspond to general points regarding the usefulness of Bayesian statistical methods for understanding process models in the cognitive sciences. At the most general level, the Bayesian framework for scientific inference allows enormous freedom in building process models. All that is required is a formal probabilistic account of how observed data are generated. Once this modeling has been done, and data are available, making inferences is the (conceptually) easy process of reversing the generative process, and inferring which combinations of parameters are likely to have given rise to the data. Important issues like balancing goodness-of-fit with complexity, assessing sensitivity to prior information, conditioning on nuisance variables, and so on, are all dealt with completely and coherently because Bayesian inference has a principled basis in probability theory.

More practically, Bayesian methods, especially through the use of graphical models or other languages that permit the use of modern computational methods for posterior sampling, make it straightforward to undertake analyses that are psychologically rich, but otherwise difficult to implement. For example, mixture models—including especially latent assignment models—allow data in a task to be modeled as having been generated by more than one psychological process. Complex regression structures are straightforward to implement and variability across trials is easily formalized in a Bayesian account.

The ability of Bayesian methods in our example to extend the scope of well-developed and widely-used diffusion process accounts of decision-making is very promising. It suggests that Bayesian methods can be applied widely to process models throughout the cognitive sciences, broadening the set of psychological questions these models can be used to answer.

References

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. *Psychological Review*, *113*, 700–765.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (in press). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modeling.* Cambridge, UK: Cambridge University Press.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, *15*, 1–15.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, *40*, 77–105.

Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.

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

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

Tuerlinckx, F. (2004). The efficient computation of the distribution function of the diffusion process. *Behavior Research Methods, Instruments, & Computers*, *36*, 702-716.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff Diffusion Model to experimental data. *Psychonomic Bulletin & Review*, *14*, 1011–1026.

Vandekerckhove, J., & Tuerlinckx, F. (in press). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*.

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*, 767–782.

Wagenmakers, E.-J. (2008). *Methodological and empirical developments for the ratcliff diffusion model of response times and accuracy.* (Submitted)