Hierarchical Bayesian Modeling of Human Decision-Making Using Wiener Diffusion

Michael D. Lee
Department of Cognitive Sciences
University of California, Irvine

Joachim Vandekerckhove
Department of Psychology
K.U. Leuven

Daniel J. Navarro
School of Psychology
University of Adelaide

Francis Tuerlinckx
Department of Psychology
K.U. Leuven

Abstract

Wiener diffusion accounts of human decision-making are among the most successful and best developed formal models in the psychological sciences. We reconsider these models from a Bayesian perspective, using graphical modeling, and Markov Chain Monte-Carlo methods for posterior sampling. By analyzing seminal data from a brightness discrimination task, we show how the Bayesian approach offers several avenues for extending and improving diffusion models. These possibilities include the hierarchical modeling of stimulus properties, and modeling the role of contaminant processes in generating experimental data. We also argue that the Bayesian approach challenges some basic assumptions of previous diffusion models, involving how variability in decision-making should be interpreted. We conclude that adopting a Bayesian approach to relating diffusion models and human decision-making data will sharpen the theoretical and empirical questions, and improve our understanding of a basic human cognitive ability.

Introduction

In the psychological sciences, one of the most successful and best developed class of formal models addresses how people make simple decisions. In particular, there is a class of 'sequential sampling' models that account for many aspects basic behavioral data, including the accuracy of the decisions that people make, and also for the full distribution of the correct and error response times (Laming, 1968; Link & Heath, 1975; Ratcliff, 1978, 2002; Ratcliff & Rouder, 1998; Ratcliff & Tuerlinckx, 2002; Ratcliff & Smith, 2004; Vickers, 1979).

In sequential sampling models of simple decision-making, it is assumed people sample evidence from some noisy signal, and terminate the sampling process when they have gathered sufficiently strong evidence in favor of one decision or the other (i.e., when the evidence accrued reaches some upper or lower threshold). If it is further assumed that the rate of sampling is sufficiently high, the evidence accumulation process can be modeled using a continuous-time stochastic diffusion process.

One important special class of these sequential sampling models uses Wiener (or Brownian) diffusion, and so assumes that evidence is sampled from a stationary Gaussian distribution. The two-sided first passage distributions of the accumulation process then give the probabilities of the two decisions, and the distribution of response times for each decision.

This basic idea of using Wiener diffusion as a model of human decision-making is illustrated in Figure 1. Five sample paths are shown, all commencing after a fixed offset T^{er} . Each path starts at z, and accumulates evidence from a Gaussian evidence distribution with mean v, until it is absorbed at either the upper boundary a or the lower boundary 0. The boundary at which the path is absorbed corresponds to the decision made, and the time taken for this state to be reaching corresponds to the response time.

The natural psychological intepretation of the four parameters is that the drift rate v measures the evidence provided by the stimulus, the starting point z measures the initial bias in favor of one decision over the other, the boundary separation a measures the level of evidence needed to make a decision, and T^{er} measures the non-decisional time involved in encoding and responding.

Despite the elegance of this framework, the evolution of Wiener diffusion models of decision-making in the psychological sciences 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 occur when errors are faster than correct decisions for easy stimuli under speed instructions, but errors are as slow or slower than correct decisions for hard stimuli under accuracy instructions. These trends 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

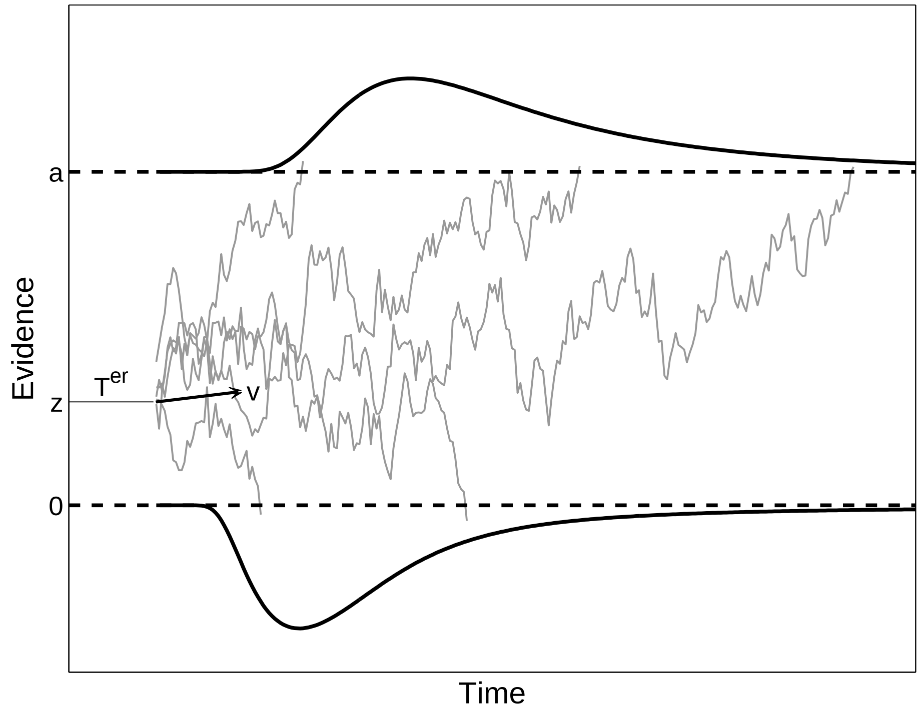

Figure 1. Wiener diffusion account of two-alternative decision-making. On each trial, evidence is accumulated from a mean-v Gaussian from a starting point z until an upper boundary of a or a lower boundary of 0 is reached.

by assuming that the starting point is subject to between-trial variation, and so is convolved with a Gaussian or uniform distribution (Laming, 1968). 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 (Ratcliff, 1978). Both of these noise processes are parameterized with the standard sufficient statistics (typically the mean and variance of a Gaussian), which become additional parameters of the model.

A second important addition has been the mixture modeling of Wiener diffusion distributions with contaminant distributions, to prevent outlier data affecting parameter estimation and model predictions (Ratcliff & Tuerlinckx, 2002; Vandekerckhove & Tuerlinckx, in press). The most recent contaminant model considers two different types of contamination: a delayed-start decision and a random guess. For both of these contaminant types, the decision time is modeled as uniformly distributed over a large range. For the random guesses, it is further assumed both decisions are equally probable. The key idea is that, on any trial, there is a high probability a subject will use a Wiener diffusion process, but a low probability they will use one or other of these contaminating decision-making processes.

With these extensions, Wiener diffusion-based accounts of decision making have proved enormously succesful in psychology (Ratcliff, Van Zandt, & McKoon, 1999). They have been been applied to experiments in memory (Ratcliff, 1978; Spaniol, Madden, & Voss, 2006), letter matching (Ratcliff, 1981), aging (Ratcliff, Spieler, & McKoon, 2004; Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2001, 2003), lexical decision (Ratcliff, Gomez, & McKoon, 2004; Wagenmakers, Ratcliff, Gomez, & McKoon, in press), signal detection (Ratcliff & Rouder, 1998; Ratcliff et al., 2001, 1999), visual search (Strayer & Kramer, 1994), and perceptual judgment (Ratcliff, 2002; Ratcliff & Rouder, 2000; Thapar, Ratcliff, & McKoon, 2003).

A number of methods for fitting extended Wiener diffusion models of decision-making to data have been developed, all relying on frequentist methods of parameter estimation (e.g., Ratcliff & Tuerlinckx, 2002; Tuerlinckx, 2004; Vandekerckhove & Tuerlinckx, in press; Voss & Voss, in press). In this paper, we explore the consequences of adopting a Bayesian approach to relating diffusion models to behavioral data. Rather than deal with the extended versions, we return to the basic diffusion process as an account of decision-making, and show that the coherent Bayesian approach to dealing with uncertainty diminishes the need to assume additional variability processes. We also use hierarchical Bayesian methods as a natural means to model the structure of stimulus information, and employ latent assignment methods to provide a more psychologically complete model of contaminant processes.

Hierarchical Bayesian Modeling

Benchmark Data

To explore the Bayesian approach to diffusion models of decision-making, we reanalyze the benchmark data from Experiment 1 of Ratcliff and Rouder (1998). In this experiment, three observers completed ten 35 minute sessions, each consisting of ten blocks with 102 trials per block. The task of observers was to decide between 'bright' and 'dark' responses for simple visual stimuli with different proportions of white dots, given noisy feedback about the accuracy of the responses. There were 33 different types of stimuli, ranging from 0% to 100% white in equal increments. In addition, the subjects were required to switch between adherence to 'speed' instructions and 'accuracy' instructions every two blocks.

Graphical Model

Our hierarchical Bayesian model is shown in Figure 2, using graphical model notation (see Griffiths, Kemp, & Tenenbaum, in press; Lee, in press; Jordan, 2004, for two psychological and one statistical introduction, respectively). In graphical models, nodes represent variables of interest, and the graph structure is used to indicate dependencies between the variables, with children depending on their parents. We use the conventions of representing continuous variables with circular nodes and discrete variables with square nodes, and unobserved variables without shading and observed variables with shading. For unobserved variables, we distinguish between stochastic variables with single borders and determinstic variables with double borders. We also use plate notation, enclosing with square boundaries subsets of the graph that have independent replications in the model.

In Figure 2, the basic observed data are the decisions d_{ijk} and response times t_{ijk} for the ith trial involving the jth stimulus under the kth (i.e., accuracy or speed) condition. Each trial is classified as having been generated by a Wiener diffusion process, or one of the two contaminant processes, using the binary indicator variables $x_{ijk} \sim \text{Bernoulli}(\pi)$ and $y_{ijk} \sim \text{Bernoulli}(\gamma)$, with rates given priors $\pi \sim \text{Uniform}(0,1)$ and $\gamma \sim \text{Uniform}(0,1)$.

If $x_{ijk} = 1$, the trial corresponds to a Wiener diffusion with drift rate v_{jk}, boundary separation a_k, non-decision time T^{er} and bias a_k. If $x_{ijk} = 0$, the trial is a contaminant, and is a delayed-start decision if $y_{ijk} = 0$ or a random guess if $y_{ijk} = 1$. Delayed-start decisions have decisions corresponding the Wiener diffusion process, but response times distributed uniformly on a wide contaminant interval $(T^{\text{mn}}, T^{\text{mx}})$. Random guesses have equal probability of producing either decision, and have response times distributed uniformly on the same interval.

This means the overall generative model for the bi-variate decision and response time data can be written as

$$d_{ijk} \;\sim\; \begin{cases} \text{Wiener}\big(v_{jk}, a_k, z_k, T^{\text{er}}\big) & \text{if } x_{ijk} = 1 \\ \text{Wiener}\big(v_{jk}, a_k, z_k, T^{\text{er}}\big) & \text{if } x_{ijk} = 0 \text{ and } y_{ijk} = 0 \\ \text{Bernoulli}\big(1/2\big) & \text{if } x_{ijk} = 0 \text{ and } y_{ijk} = 1 \end{cases} \tag{1}$$

for the decisions, and

$$t_{ijk} \;\sim\; \begin{cases} \text{Wiener}\big(v_{jk}, a_k, z_k, T^{\text{er}}\big) & \text{if } x_{ijk} = 1 \\ \text{Uniform}\big(T^{\text{mn}}, T^{\text{mx}}\big) & \text{if } x_{ijk} = 0 \text{ and } y_{ijk} = 0 \\ \text{Uniform}\big(T^{\text{mn}}, T^{\text{mx}}\big) & \text{if } x_{ijk} = 0 \text{ and } y_{ijk} = 1 \end{cases} \tag{2}$$

for the times.

Unlike previous models of these data, we used a psychometric function to relate the physical stimulus properties (i.e., their proportion of white dots) to their psychological properties (i.e., the evidence quantified by their drift rates). We chose a Weibull function form this relationship (e.g., Wichmann & Hill, 2001), so that the drift rate of the jth stimulus under the kth condition is given by

$$v_{jk} \;=\; v^{\text{lo}} + \big(v^{\text{hi}} - v^{\text{lo}}\big)\left(1 - \exp\left(\frac{s_k}{v^{\text{sc}}}\right)^{v_k^{\text{sh}}}\right), \tag{3}$$

where s_j is the proportion of white dots for the jth stimulus, v^{lo} and v^{hi} are the lower and upper asymptotes for the Weibull, v^{sc} is the scale or center of the Weibull, and v_k^{sh} is the shape of the Weibull for the kth condition. We set uniform priors on all these parameter over a range that was observed to cover all of the posterior density.

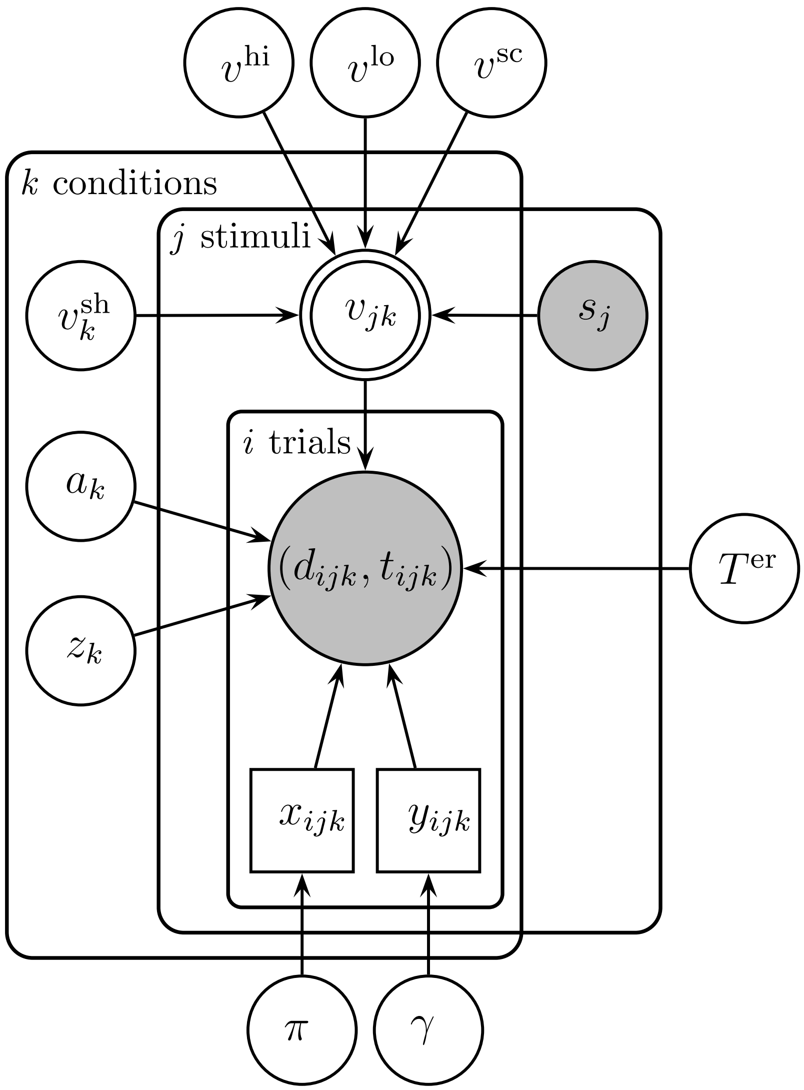

Figure 2. Graphical model for the Wiener diffusion account of the benchmark data. See text for details.

We also set a sufficiently wide uniform prior approach on the boundary separations a_k for the kth condition, and set $z_k = b_k a_k$, with $b_k \sim \text{Uniform}\big(0, 1\big)$ being the proportional position of the starting point relative to boundary separation for the kth condition.

This final model resulted from considering many alternative structures. For example, we considered allowing T^{er}, v^{lo}, v^{hi}, and v^{sc} to vary across task instruction conditions. Our general approach was not to allow parameters to vary across stimuli or conditions unless it was clear that, when they were freely estimated, their values changed significantly.

Results

We implemented the model in Figure 2 using WinBUGS (Spiegelhalter, Thomas, & Best, 2004), which uses a range of Markov Chain Monte-Carlo computational methods, including adaptive rejection sampling, slice sampling, and Metropolis-Hastings (see, for example Chen, Shao, & Ibrahim, 2000; Gilks, Richardson, & Spiegelhalter, 1996; Mackay, 2003) to perform posterior sampling. We applied the model separately to all three subjects in the benchmark data set, drawing 3,000 posterior samples after a burn-in of 5,000 samples.

Parameter Inference. Figure 3 shows the posterior marginal distributions for each parameter for each subject. Many of these posteriors are consistent with previous findings, but have the advantage of providing full distributions, rather than point estimates and standard errors. The boundary separations are greater for the accuracy than the speed condition. The non-decision time posteriors for T^{er} are sensible, as are the asymptotes and scale of the Weibull psychometric function. And the posterior for the contaminant rate π shows that almost all of the trials are inferred to be generated by the Wiener diffusion process.

Psychometric Function. The most interesting result in Figure 3 is that the shape of the Weibull function clearly changes across instruction conditions for all of the participants. The shape parameter is much smaller in the accuracy than in the speed condition, corresponding to a sharper Weibull function. This is made clear in Figure 4 which shows the posterior distribution of the psychometric function for both accuracy and speed conditions for each subject. Superimposed on these curves are points showing the drift rate for each stimulus when it is inferred freely, without the assumption of an underlying psychometric curve. The close alignment of these unconstrained inferences with the curve justifies the assumption of the Weibull, and reinforces that the drift rate for a stimulus changes between accuracy and speed conditions.

Decision Accuracy. Figure 5 shows the ability of the model to account for the accuracy of decisions. These scatterplots show close agreement between the modeled and observed accuracy for each stimulus observed by each subject under both

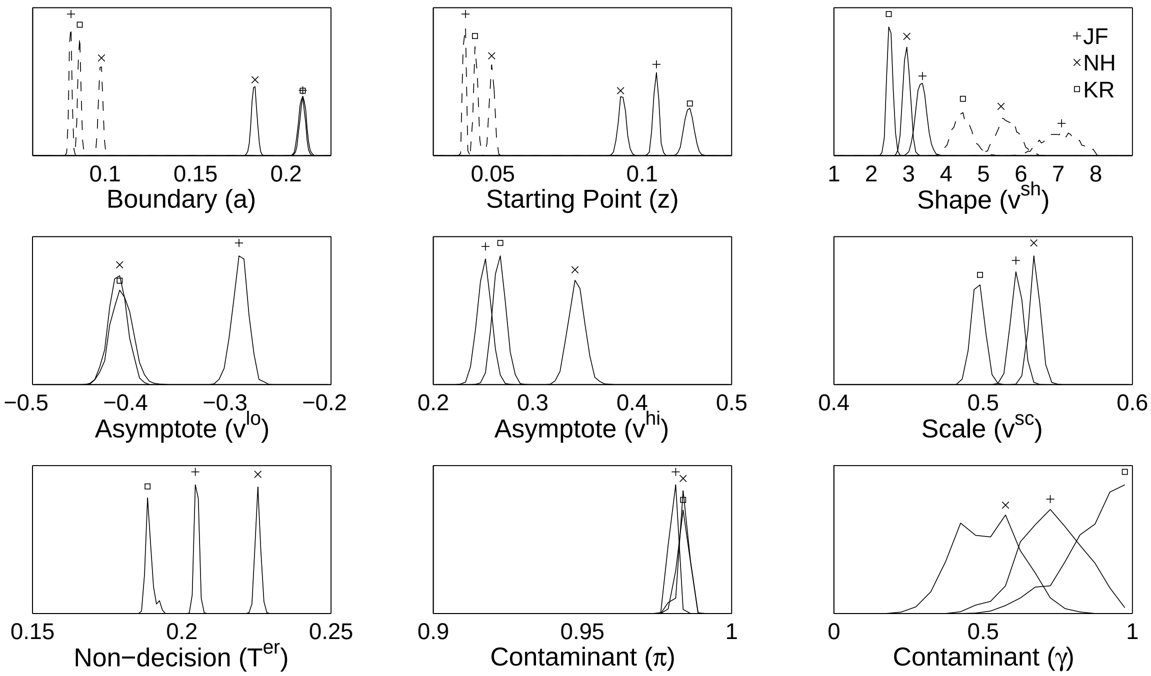

Figure 3. Marginal posterior distribution for each parameter for each subject. For the boundary separation (a) and Weibull shape (v^{sh}) that vary across instruction conditions, the accuracy posteriors are shown by solid lines, and the speed posteriors by broken lines. A marker indicating the subject is shown above the mode for each posterior.

conditions.

Response Time Distributions. Figure 6 shows the posterior predictive response time distributions for one subject (KR) over three different, under both speed and accuracy instructions. The posterior predictive distributions are the expected decisions and response time distributions the model has for future data, based on what is learned from the available data. In this sense, the posterior predictives give an account of how well the model fits in the data, in a way that is automatically sensitive to all aspects of model complexity (Myung, Balasubramanian, & Pitt, 2000). Technically, the posterior predictive is found by averaging across the joint posterior distribution of the parameters, and so is given by

$$p\left(d'_{jk}, t'_{jk} \mid D\right) = \int p\left(d'_{ijk}, t'_{ijk} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta} \mid D\right)\ \mathrm{d}\boldsymbol{\theta}, \tag{4}$$

where d'_{jk} and t'_{jk} denote the predicted decision and response time for the jth stimulus under the kth instruction conditions, $\boldsymbol{\theta} = (v_{jk}, a_k, z_k, T^{\text{er}})$ denotes the model parameters for the prediction, and D denotes the observed data.

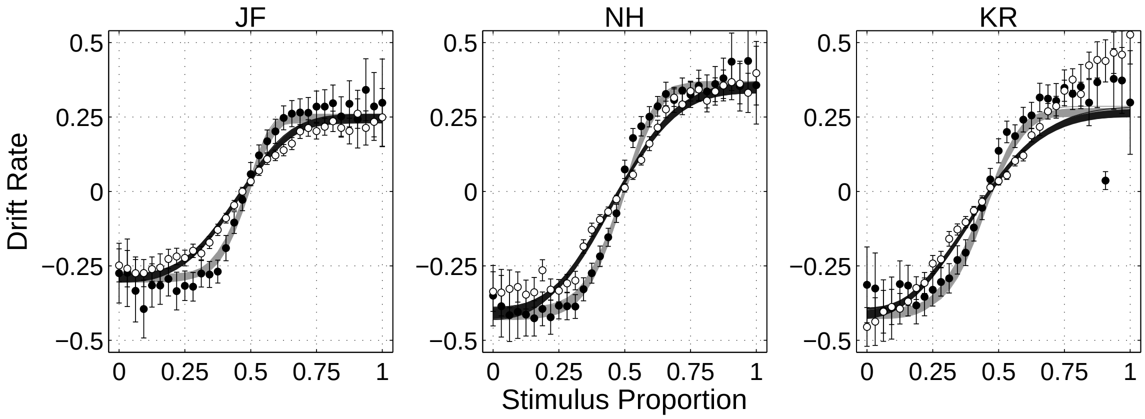

Figure 4. Posterior distributions for the Weibull psychometric functions, relating the proportion of white dots in the stimulus to the drift rate measure of evidence, for the accuracy (grey line) and speed (black line) conditions. Both curves are generated by overlaying the specific Weibull curves corresponding to 100 samples from the posterior. Also shown are the mean and 95% credible intervals for accuracy (black markers) and speed (white markers). Each panel corresponds to a subject.

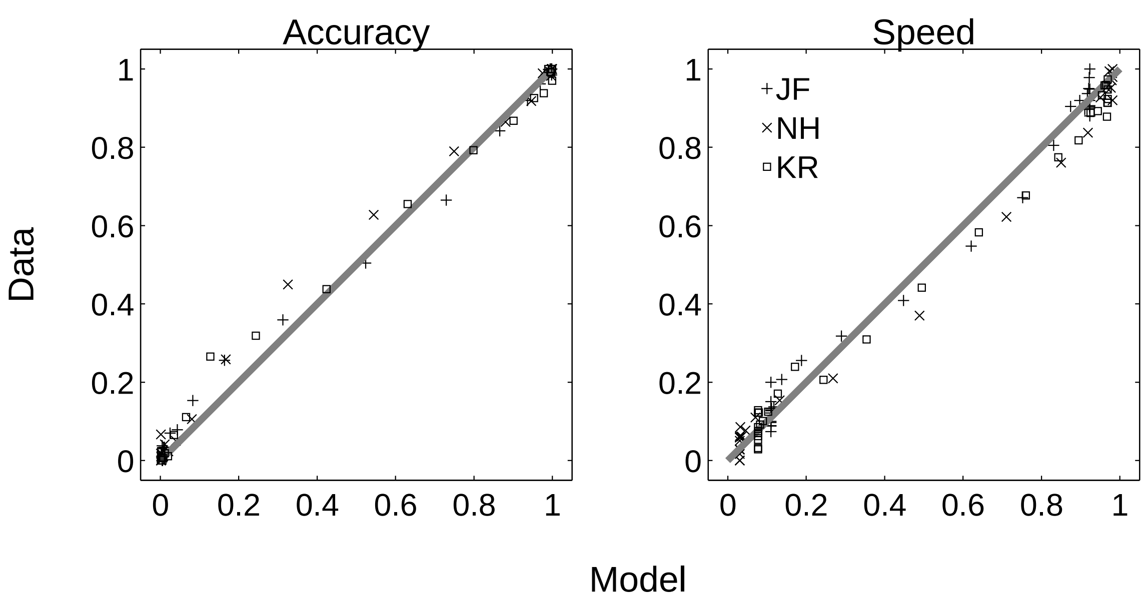

Figure 5. The decision accuracy predicted by the model against the accuracy observed in the data. Each marker corresponds to a stimulus, the type of marker indicates the subject, and the two panels correspond to accuracy and speed instructions.

The proportions of white dots for the three stimuli are 1.0, 0.55 and 0.52. Also shown by short vertical lines are the raw response time data, and filled or unfilled markers are used to indicate response times that are likely (between 50–90% posterior expectation) or very likely (greater than 90% posterior expectation) contaminants. As Figure 6 shows, the model provides a good account across this range of difficulties, for both instructions. Posterior predictive density is given to all the observed data, and outliers are sensibly classified as contaminants.

Latent Contaminant Mixtures. Some sort of approach to outliers is needed for the current data because, for example, subject KR produced a 30 msec response under accuracy instructions, and their mean response time for the same stimulus under accuracy instructions was 1,528 msec. Figure 6 also show the effectiveness of the Bayesian latent assignment approach to contaminant mixture modeling as a method of detecting outliers. The three response times classified as very likely to be contaminants seem to be clear outliers, being much faster than any other the other times.

The current Bayesian approach differs from previous modeling by having latent variables that assign every trial, on each posterior sample, either fully to the Wiener diffusion distribution or to a contaminant distribution. Existing non-Bayesian approaches to contaminant modeling estimate the relative mixture of Wiener diffusion distribution to contaminant distribution across all trials, and then must make post-hoc inferences about which trials are the contaminants. Our Bayesian modeling more closely aligns with the original psychological motivation, by explicitly modeling which trials are driven by the Wiener diffusion, and which result from contaminant processes.

Cross-Over Effects. Figure 7 considers the theoretically important cross-over effect. The posterior predictive response time distributions and raw data, for all three subjects, are shown for a relatively hard stimulus—with a 0.45 proportion of white dots—under accuracy instructions, and for a relatively easy stimulus—with a 0.22 proportion of white dots—under speed instructions. Once again, the posterior predictive distributions provide a good descriptive account of the data.

Importantly, however, the key component of the cross-over effect, in the form of fast errors for easy stimuli under speed instructions, is not observed in the modeling, because of the use of the contaminant distributions. In fact, an examination of the posterior predictives over the entire data set failed to find any example of a cross-over effect. This pattern of results is clearest in Figure 7 for subject KR, who has two fast errors for the easy speed case, but both are classified by the model as highly likely to be contaminants. The thinner curve in this bottom-right panel of Figure 7 shows the posterior predictive distributions that results from modeling these two response times using Wiener diffusion, and makes it clear that the model *can* capture cross-over effects. As pointed out by Lee, Fuss, and Navarro (2007), this capability stems from the posterior predictive distributions being found by integrating over the joint posterior, and so automatically incorporate variability in the starting point and drift

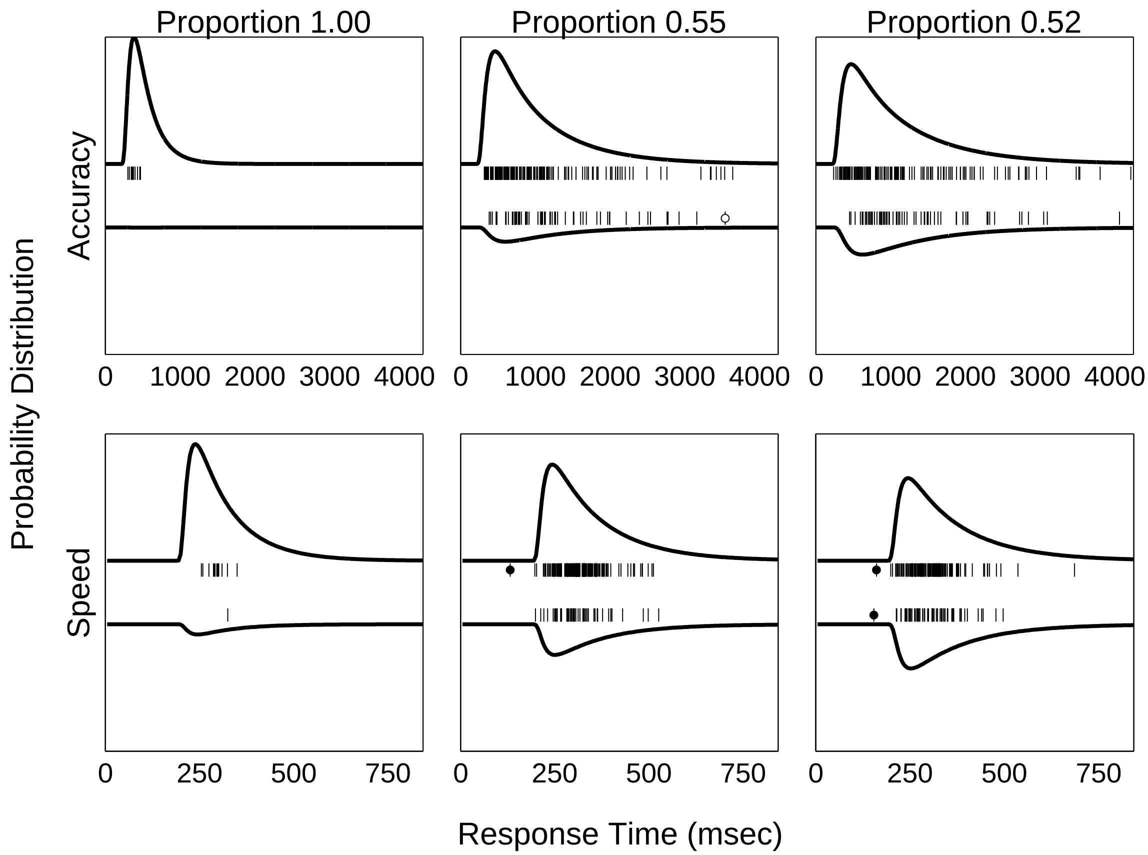

Figure 6. Posterior predictive response time distributions and raw data for subject KR, under accuracy (top row) and speed (bottom row) conditions, for stimuli with proportions of white dots 1.0 (left column), 0.55 (middle column) and 0.52 (right column). The posterior predictive distributions are shown by the bold curves, and the response time data are shown by short vertical lines. Filled markers on a response time indicate a greater than 90% posterior expectation of being a contaminant, while open markers indicate a contaminant posterior expectation between 50–90%.

rate parameters.

These results challenges two basic conclusions from the previous literature. The first is the existence of the cross-over effect, as an empirical regularity in the data. Of course, a model that did not incorporate a contaminant mixture, or perhaps used a different contaminant model, may well include retain data like the two in Figure 7 as the outcome of a Wiener diffusion process. But, as the posterior for the contaminant parameter π in Figure 3 makes clear, the current assumptions retained about 97% of the data. Thus, if there is a cross-over effect in the current data, it arises from performance on a small fraction of the trials.

The second challenge for contemporary Wiener diffusion modeling is that, even if the cross-over effect does exist empirically, it is not clear additional variability processes need to be assumed. In a fully Bayesian approach, the basic Wiener diffusion model naturally accommodates the effect, and it is not clear what advantage is gained from introducing parameterized variability processes. The parameters of the basic Wiener diffusion have all been shown to pass tests of 'selective influence', in which parameters change in interpretable and expected ways under experimental manipulation (Voss, Rothermund, & Voss, 2004). Drift rates change with stimuli, boundary separation with task instructions, starting point with base-rates or payoffs, and times for encoding and responding vary between young and old subjects.

There is not the same body of convincing demonstrations of selective influence for the parameters of the variability processes. The alternative psychological account, suggested by the current Bayesian modeling, is that people are not convolving stimulus information with additional noise sources. Rather, they are uncertain about both the evidence in the stimulus (drift rate), and about the correct strategy (starting point and boundary), and handle this uncertainty in a coherent way, by averaging over it in reaching their decisions.

Required Data. One of the difficulties in estimating parameters of diffusion models using previous techniques is that many data are required. Bayesian methods generally require fewer data, and so we tested the ability of our model to estimate parameters using only 10% of the data. Table 1 shows the expected values of each parameter for each subject, both using the full data set, and using a random 10% sample, selected by having the median value of sum-squared relative error. Most of the estimates for the reduced data set are very similar to those inferred from the full data set, and certainly preserve all of the important order relations and trends in the parameter values across conditions.

Discussion

In this paper, we have presented a fully Bayesian account of the a Wiener diffusion process account of human decision-making, and re-analyzed the seminal data set considered by Ratcliff and Rouder (1998). Our results demonstrate a number of potential advantages of the Bayesian approach.

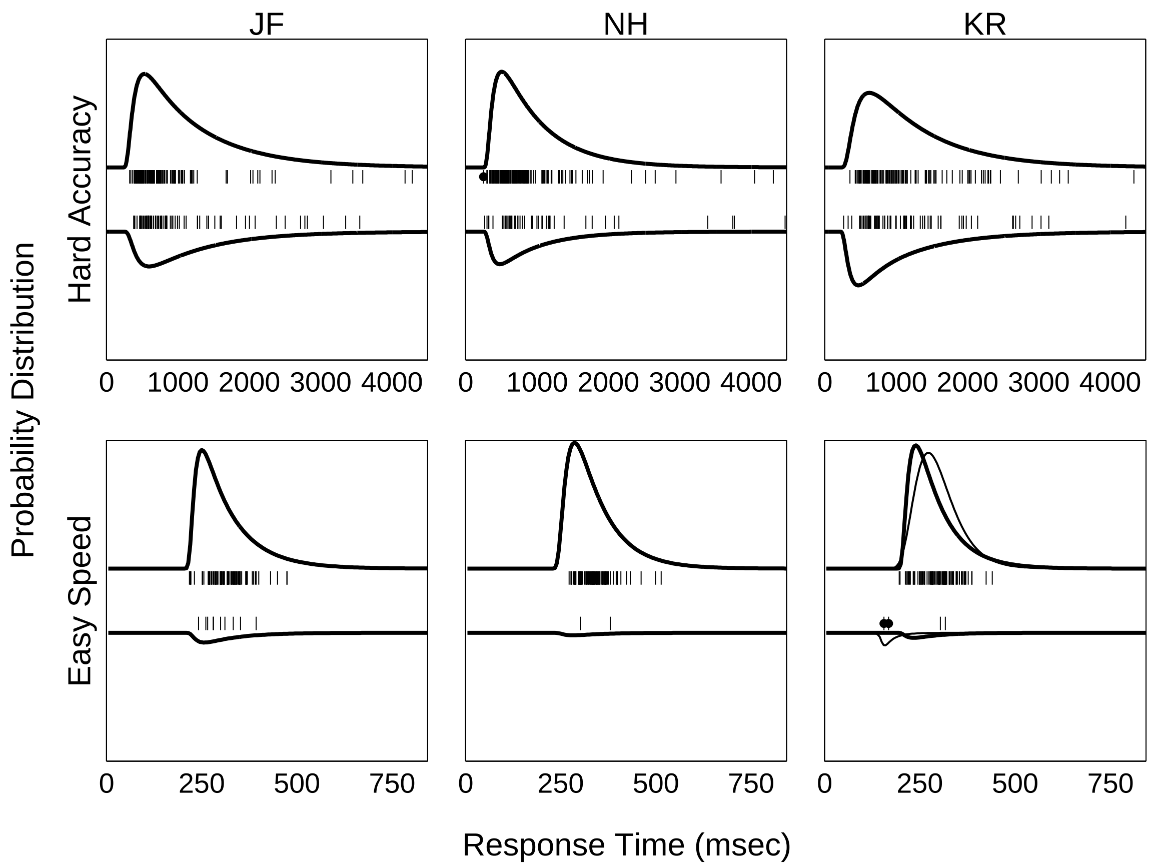

Figure 7. Posterior predictive response time distributions and observed data for all three subjects (one per column), for both a hard stimulus under accuracy instructions (top row), and an easy stimulus under speed instructions (bottom row). The posterior predictive distributions are shown by the bold curves, and the response time data are shown by short vertical lines. Filled markers on a response time indicate a greater than 90% posterior expectation of being a contaminant. For the easy speed condition for subject KR the thinner curve shows the posterior predictive distribution if the two contaminant response times are modeled by the Wiener diffusion process.

Table 1: Expected parameter values for the full data set, and a sub-sample of 10% randomly drawn observations per stimulus per participant. We repeated this analysis for many such sub-samples, but report only the ones that yielded median values with respect to a sum-squared relative error criterion.

	JF		NH		KR	
Parameter	Full	10%	Full	10%	Full	10%
a_{accuracy}	0.21	0.21	0.18	0.20	0.21	0.22
a_{speed}	0.08	0.08	0.10	0.10	0.09	0.08
z_{accuracy}	0.10	0.10	0.09	0.11	0.12	0.12
z_{speed}	0.04	0.04	0.05	0.05	0.05	0.04
T^{er}	0.21	0.22	0.23	0.21	0.19	0.20
v^{hi}	0.25	0.24	0.34	0.33	0.27	0.42
v^{lo}	-0.29	-0.28	-0.41	-0.43	-0.41	-0.46
v^{sc}	0.52	0.53	0.53	0.54	0.50	0.57
$v^{\text{sh}}_{\text{accuracy}}$	3.34	2.97	2.94	2.29	2.50	2.34
$v^{\text{sh}}_{\text{speed}}$	7.04	6.74	5.65	4.79	4.43	2.59

The first, and perhaps most important, advantage is that the Bayesian approach suggests the use of additional variability processes may not be needed. This has practical consequences for simplifying simulation and analysis of the model, and estimation of its parameters. It also has theoretical ramifications, suggesting a different psychological account of how people make simple two-alternative forced-choice decisions under conditions of uncertainty.

A second advantage is the natural application of hierarchical modeling to represent knowledge at different levels of abstraction. Rather than having separate drift rate parameters for all 33 stimuli, we were able to infer a more abstract psychometric function with only four parameters. The same hierarchical ideas could also be applied to subjects to capture individual differences (Lee & Webb, 2005). We plan in the future to investigate the possibility of including an account of individual differences in a Wiener diffusion model of simple decision-making, using data sets that include more than three subjects.

A third advantage of the Bayesian approach is in the handling of contaminant mixture modeling. The Bayesian approach, through the use of latent assignment variables for each trial, more closely corresponds to the psychological conception of using contaminant distributions, which is that most trials are fully the result of a Wiener diffusion process, but some small fraction of trials are fully the result of a contaminating process.

Finally, our results highlight the usefulness of posterior predictive distributions in representing the behavior of a model of the time course of decision-making, and particularly in assessing its descriptive adequacy.

Besides considering individual differences, we plan to focus future work on using more rigorous methods for model development decisions. The graphical model used here arose from exploring the posterior distributions over parameters for a range of plausible models, but relied on subjective judgments, rather than theoretically-driven measures like Bayes Factors (Kass & Raftery, 1995).

Diffusion-based models of simple decision-making are among the most successful, best developed, and most popular formal accounts of human cognition. We have shown that adopting a Bayesian approach to relating these models to data can contribute to their development, and challenge some of their basic assumptions. In this way, adopting the Bayesian approach promises to sharpen the theoretical and empirical questions that need good answers to understanding the basic human cognitive ability to make simple decisions.

References

Chen, M. H., Shao, Q. M., & Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Spinger-Verlag.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov Chain Monte Carlo in Practice*. Boca Raton (FL): Chapman & Hall/CRC.

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 University Press.

Jordan, M. I. (2004). Graphical models. *Statistical Science*, *19*, 140–155.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Laming, D. R. J. (1968). *Information theory of choice–reaction times.* London: Academic Press.

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

Lee, M. D., Fuss, I. G., & Navarro, D. J. (2007). A Bayesian approach to diffusion models of decision-making and response time. In B. Schölkopf, J. C. Platt, & T. Hofmann (Eds.), *Advances in Neural Information Processing Systems 19.* Cambridge, MA: MIT Press.

Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, *12*(4), 605–621.

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

Mackay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms.* Cambridge: Cambirdge University Press.

Myung, I. J., Balasubramanian, V., & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences*, *97*, 11170–11175.

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

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, *88*, 552–572.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review*, *9*, 278–291.

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., & Rouder, J. N. (2000). A diffusion model account of masking in two–choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 127–140.

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

Ratcliff, R., Spieler, D., & McKoon, G. (2004). Analysis of group differences in processing speed: Where are the models of processing. *Psychonomic Bulletin &Review*, *11*, 755–769.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical–decision task. *Psychology and Aging*, *19*, 278–289.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*, 323–341.

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

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.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *102*, 261–300.

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Human Learning and Memory*, *32*, 101–117.

Spiegelhalter, D. J., Thomas, A., & Best, N. G. (2004). *WinBUGS Version 1.4 User Manual*. Cambridge, UK: Medical Research Council Biostatistics Unit.

Strayer, D. L., & Kramer, A. F. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 318–341.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, *18*, 415–429.

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. (in press). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*.

Vickers, D. (1979). *Decision processes in visual perception*. New York, NY: Academic Press.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*, 1206–1220.

Voss, A., & Voss, J. (in press). Fast-dm: A free program for efficient diffusion model analysis. *Behavioral Research Methods*.

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (in press). A diffusion model account of criterion manipulations in the lexical decision task. *Journal of Memory and Language*.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception & Psychophysic*, *63*, 1293–1313.