Cognitive Models and the Wisdom of Crowds: A Case Study Using the Bandit Problem

Shunan Zhang (szhang@uci.edu)
Michael D. Lee (mdlee@uci.edu)
Department of Cognitive Sciences, 3151 Social Sciences Plaza A
University of California, Irvine, CA 92697-5100 USA

Abstract

The “wisdom of the crowds” refers to the idea that the aggregated performance of a group of people on a challenging task may be superior to the performance of any of the individuals. For some tasks, like estimating a single quantity, it is straightforward to aggregate individual behavior. For more complicated multidimensional or sequential tasks, however, it is not so straightforward. Cognitive models of behavior are needed, to infer what people know from how they behave, and allow aggregation to be done on the inferred knowledge. We provide a case study of this role for cognitive modeling in the wisdom of crowds, using a multidimensional sequential optimization problem, known as the bandit problem, for which there are large differences in individual ability. We show that, using some established cognitive models of people’s decision-making on these problems, aggregate performance approaches optimality, and exceeds the performance of the vast majority of individuals.

Keywords: Wisdom of crowds, Cognitive models, Bandit problem, Hierarchical Bayesian modeling

Introduction

An enticing idea in the study of individual and group decision-making is the phenomenon known as the “wisdom of crowds”. The idea is that, by aggregating the behavior of a group of people doing a challenging task, it is possible for group performance to match or exceed the performance of any of the individuals. Surowiecki (2004) provides an extensive survey of wisdom of crowds results over a diverse set of human endeavors and decision-making situations, ranging from guessing the weight of an ox at a county fair, to inferring the location of a missing submarine, to predicting the outcome of sporting events. Recent research in cognitive science has looked at issues including whether it is possible to have a “crowd within”, such that multiple estimates from the same person can be combined to improve their performance (Vul & Pashler, 2008).

While the exact conditions needed for group performance to exceed individual performance are not completely understood, it seems clear that crowds can be wise in any situation where people have some partial knowledge, and the gaps in their knowledge are subject to individual differences. Under these circumstances, aggregation of individual decisions can serve to amplify the common signal and reduce the idiosyncratic noise, leading to superior group performance.

One challenge in producing wisdom of crowds effects arises when tasks are more complicated than estimating a single quantity, or predicting a simple outcome. Many interesting and real-world decision-making situations are inherently multidimensional or sequential. In these situations, it is often not possible to combine the raw behaviors of people, because they are not commensurate. For example, imagine trying to combine the expertise of basketball fans trying to predict the result of an eight-team single elimination tournament, with quarter-finals, semi-finals and a final. Based on their decisions about the quarter-finals, these people may be making decisions about different teams in the semi-finals and final. This makes simple aggregation based on their raw decisions impossible for the later rounds.

For more difficult decision problems like these, we believe cognitive science has a key role to play in wisdom of the crowd research. Rather than aggregating people’s behaviors, it is necessary to aggregate their knowledge, as *inferred* from their behavior. This inference needs models of cognition, accounting for how latent knowledge manifests itself as observed behavior within the constraints of a complicated task. Steyvers, Lee, Miller, and Hemmer (in press) present an example of this approach, using Thurstonian models of judgment to combine people’s ranking decisions for a variety of general-knowledge questions, such as the chronology of the US Presidents.

In this paper, we present a case study of the application of cognitive models for a sequential task known as the bandit problem. By applying a series of existing models of human decision-making on the task to a variety of data sets, we show that it is sometimes possible to produce aggregate performance that is near optimal, and far exceeds the performance of most of the individuals. We discuss what sort of properties cognitive models might need to achieve this sort of useful aggregation of individual knowledge.

Bandit Problems

Bandit problems are a type of sequential decision-making problem widely studied in statistics and machine learning (Gittins, 1979; Kaebling, Littman, & Moore,

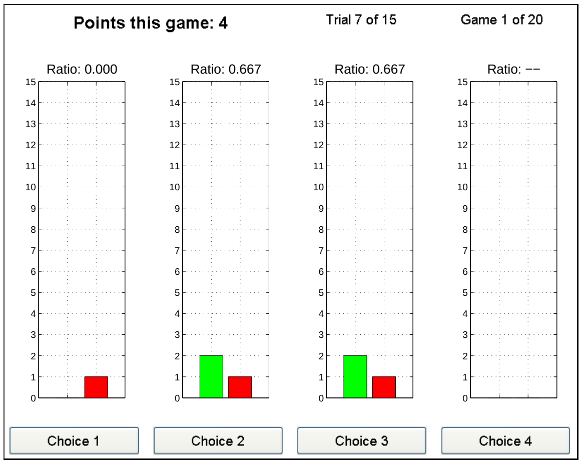

Figure 1: An experimental interface, giving an example of a Bandit problem.

1996; Sutton & Barto, 1998), as well as in cognitive science (Cohen, McClure, & Yu, 2007; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Steyvers, Lee, & Wagenmakers, 2009). In Bandit problems, a decision maker chooses from a set of alternatives with fixed but unknown reward rates, which are drawn from a fixed but unknown environment, with the goals of maximizing the total number of rewards after a fixed number of trials.

A representative experimental interface of Bandit problems is shown in Figure 1. The four large panels contain information of choices and outcomes on four alternatives. On each trial, an alternative is chosen, and either succeeds in giving a reward (green, light) or fails (red, dark). At the top of each panel, the ratio of successes, defined as the ratio of successes to total choices, is shown. The interface provides a count of the total number of rewards obtained up to the current trial. The current game and trial are also shown.

The bandit problem has a well-known optimal decision-making process (e.g., Kaebling et al., 1996, p. 244), calculated by dynamic programming. This allows human decision-making, and plausible psychological models of decision-making, to be assessed in terms of their optimality. In particular, Bandit problems provide a natural task to study the inherent trade-off between exploration (seeking rewarding alternatives among those relatively unexplored) and exploitation (staying with alternatives known to be reasonably good) inherent in many real-world sequential decision-making situations.

Human Data

We use data from three experiments. In the first experiment, reported by Steyvers et al. (2009), a total of 451 participants completed a total of 20 bandit problems, each with 4 alternatives and 15 trials. Reward rates were drawn for each alternative independently from a $\text{Beta}(2,2)$ distribution. The reward rates were drawn only once, but the order of the games was randomized.

The second and third experiments involve new data. A total of 47 and 31 participants, respectively, completed 100 bandit problems, all with 4 alternatives and 16 trials. For the second experiment, the reward rates were drawn independently for each game from $\text{Beta}(8,4)$ (called a "plentiful" environment, because reward rates tend to be high). For the third experiment, reward rates came from a $\text{Beta}(4,8)$ (called a "scarce" environment, because reward rates tend to be low)

Four Decision-Making Models

In this paper, we consider four well-established models of decision-making on bandit problems. These come from the reinforcement- and machine-learning literatures (see Sutton & Barto, 1998), and have previously been examined as models of human decision-making (e.g., Lee, Zhang, Munro, & Steyvers, 2009).

Win-Stay Lose-Shift

Perhaps the simplest reasonable approach for making bandit problem decisions is the Win-Stay Lose-Shift (WSLS) heuristic. In its deterministic form, it assumes that the decision-maker continues to choose an alternative following a reward, but shifts to the other alternative following a failure to reward. In the stochastic form we use, the probability of staying after winning, and the probability of shifting after losing, are both parameterized by the same probability γ.

Extended Win-Stay Lose-Shift

A natural, and psychologically-motivated, extension to the WSLS model is to have different rates for staying after a reward (i.e., reinforcement) and shifting after a lack of reward (i.e., negative reinforcement). Formally, in our extended WSLS model, a decision-maker stays with probability γ^w following a reward, but shifts with probability γ^l following a failure to reward.

ε-Greedy

The ε-greedy model assumes that decision-making is driven by a parameter ε that controls the balance between exploration and exploitation inherent in bandit problems. On each trial, with probability $1-\varepsilon$ the decision-maker chooses the alternative with the greatest estimated reward rate (i.e., the greatest proportion of rewards obtained for previous trials where the alternative was chosen). This can be conceived as an 'exploitation' decision. With probability ε, the decision-maker chooses randomly. This can be conceived as an 'exploration' decision.

ε-Decreasing

The ε-decreasing model is a variant of ε-greedy, in which the probability of an exploration move decreases as trials progress. In its most common form, which we use, the ε-decreasing model starts with an exploration probability ε'

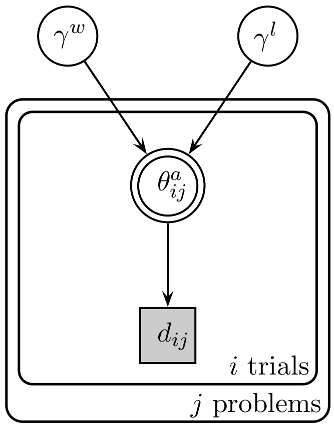

Figure 2: Bayesian graphical model for the extended WSLS decision-making model.

on the first trial, and then uses an exploration probability of ε'/i on the ith trial.

Modeling Analysis

In this section, we implement the four decision-making models in a way that allows for differences in individual behavior to be aggregated, culminating in model-based wisdom of crowds analyses of our experimental data sets.

Bayesian Graphical Model Implementation

We implemented all four decision-making models using the formalism provided by Bayesian graphical models, as widely used in statistics and computer science (e.g., Koller, Friedman, Getoor, & Taskar, 2007). A graphical model is a graph with nodes that represents the probabilistic process by which unobserved parameters generate observed data. Details and tutorials are aimed at cognitive scientists are provided by Lee (2008) and Shiffrin, Lee, Kim, and Wagenmakers (2008). The practical advantage of graphical models is that sophisticated and relatively general-purpose Markov Chain Monte Carlo (MCMC) algorithms exist that can sample from the full joint posterior distribution of the parameters conditional on the observed data. More specifically, for our purposes, graphical models can be specified that naturally combine information across multiple sources, and so can model the individual differences at the heart of the wisdom of crowds phenomenon.

As a concrete example, Figure 2 shows the graphical model implementation of the extended WSLS model. The two model parameters, the probability of win-stay γ^w and lose-shift γ^l, are shown as unshaded (i.e., unobserved) and circular (i.e., continuous) variables. These determine the probability of the ath alternative being

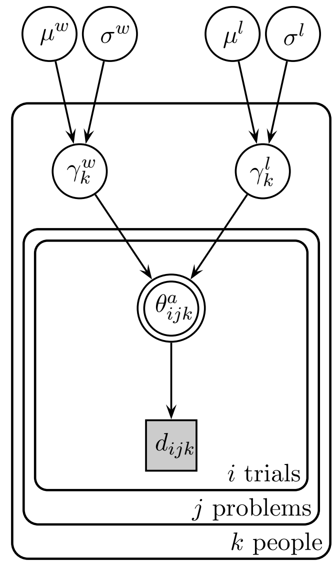

Figure 3: Bayesian graphical model for a hierarchical version of the extended WSLS decision-making model, which allows for individual-level parameter variation.

chosen on the ith trials of the jth game, as

$$\theta^a_{ij} = \begin{cases} \gamma^w & \text{if succeeded on } a \text{ last trial} \\ 1-\gamma^l & \text{if failed on } a \text{ last trial} \\ (1-\gamma^w)/3 & \text{if succeeded on } \bar{a} \text{ last trial} \\ \gamma^l/3 & \text{if failed on } \bar{a} \text{ last trial,} \end{cases}$$

where $\bar{a}$ refers to not choosing the ath alternative. Since θ_{ij} is a deterministic function of γ^w and γ^l, it is shown as a double-bordered node. Given the choice probabilities in θ^a_{ij}, the actual decision made by the ith trial of the jth problem— which is represented by a shaded square node d_{ij}, since it is observed, and discrete—is modeled as $d_{ij} \sim \text{Discrete}(\theta^1_{ij} \ldots, \theta^4_{ij})$.

Parameter Differences

One obvious possibility for individual differences is that two people—even if they are both using, for example, extended WSLS—might not have the same probabilities of wining and staying or losing and shifting. To accommodate variation in these parameters on an individual-by-individual uses, we use a *hierarchical* or *multi-level* approach. The updated graphical model is shown in Figure 3. In this model, the parameters for individual people are drawn from over-arching Gaussian distributions, so that, for the kth person, $\gamma^w_k \sim \text{Gaussian}(\mu^w, \sigma^w)$, and $\gamma^l_n \sim \text{Gaussian}(\mu^l, \sigma^l)$. This allows different people to have different parameter values, while still estimating the mean parameter value of the group as a whole.

We implemented the graphical model in Figure 3, as well as analogous graphical models for the three other decision-making models, in WinBUGS (Spiegelhalter, Thomas, & Best, 2004). This software uses a range

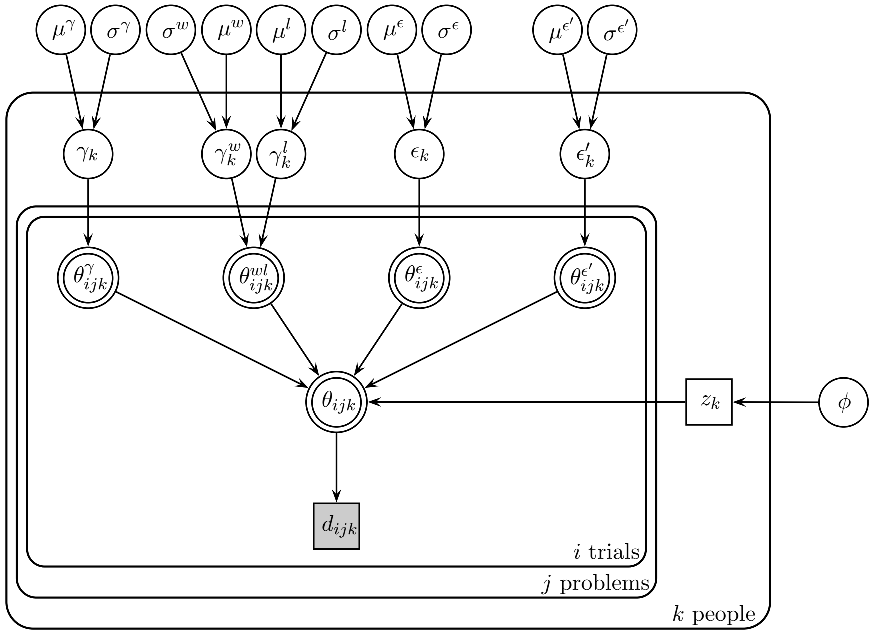

Figure 4: Graphical model using a hierarchical mixture of all four hierarchical decision-making models.

Table 1: Means, and standard deviations in brackets, of the group distributions for each parameter in the four decision-making models.

Parameter	Exp. 1	Exp. 2	Exp. 3
γ	0.71 (.10)	0.70 (0.10)	0.52 (0.10)
γ^w	0.99 (0.27)	0.97 (0.19)	0.81 (0.18)
γ^l	0.59 (0.25)	0.28 (0.23)	0.37 (0.23)
ε	0.24 (0.10)	0.18 (0.11)	0.42 (0.12)
ε'	0.61 (0.11)	0.61 (0.11)	0.90 (0.14)

of MCMC computational methods, including adaptive rejection sampling, splice sampling, and Metropolis-Hastings to perform posterior sampling (e.g., MacKay, 2003). For all four decision-making models, we made inferences about individual- and group-level parameters for all three data sets, using all of the participants. In each analysis, we collected 1,000 samples from 2 chains, collected after a burn-in period of 1,000 samples, and using standard checks for convergence.

Table 1 summarizes individual differences in parameters for each decision-making model, giving the means and standard deviations for each parameter in the hierarchical analysis. Remembering that experiments 1, 2, and 3 correspond to neutral, plentiful and scarce environments, the aggregated group parameters make sense. For example, there is more winning and staying (e.g., in the γ and γ^w parameters) in environments that deliver rewards, and there is more random exploration (e.g., in the ε and ε') in scarce environments that are not delivering rewards. The reasonably large standard deviations for most group distributions also indicate that there are significant individual differences.

Model Differences

An even more fundamental source of individual differences arises when different people use different decision processes. Rather than just varying the parameters of a model, people may differ in terms of which decision-making model they use. We accommodate this type of individual differences using a *mixture* or *latent assignment* model where people are categorized into different model-users.

The graphical model for achieving this mixture of decision models, while retaining the possibility of parameter variation within each model, is shown in Figure 4. Hierarchical versions of all four decision-making models—those used individual to assess parameter variation in the previous section—are all shown.

The key addition, in terms of individual differences, involves the model indicator variable z_k, which indexes which of the four models the kth participant uses. That is, depending on whether z_k is 1, 2, 3 or 4, the kth participant uses WSLS, the extended WSLS, ε-greedy or ε-decreasing to make their bandit problem decisions. The latent indicator variable has prior $z_k \sim \text{Categorical}(\phi)$, where ϕ is a latent base-rate, measuring the proportion of people who follow each model. We use the prior $\phi \sim \text{Dirichlet}(1/4, \ldots, 1/4)$, so that there is no initial bias towards one decision model over another.

Table 2 gives the posterior expectation of the base-rate parameter ϕ, for all three experiments. This provides a natural summary of what proportion of people were us-

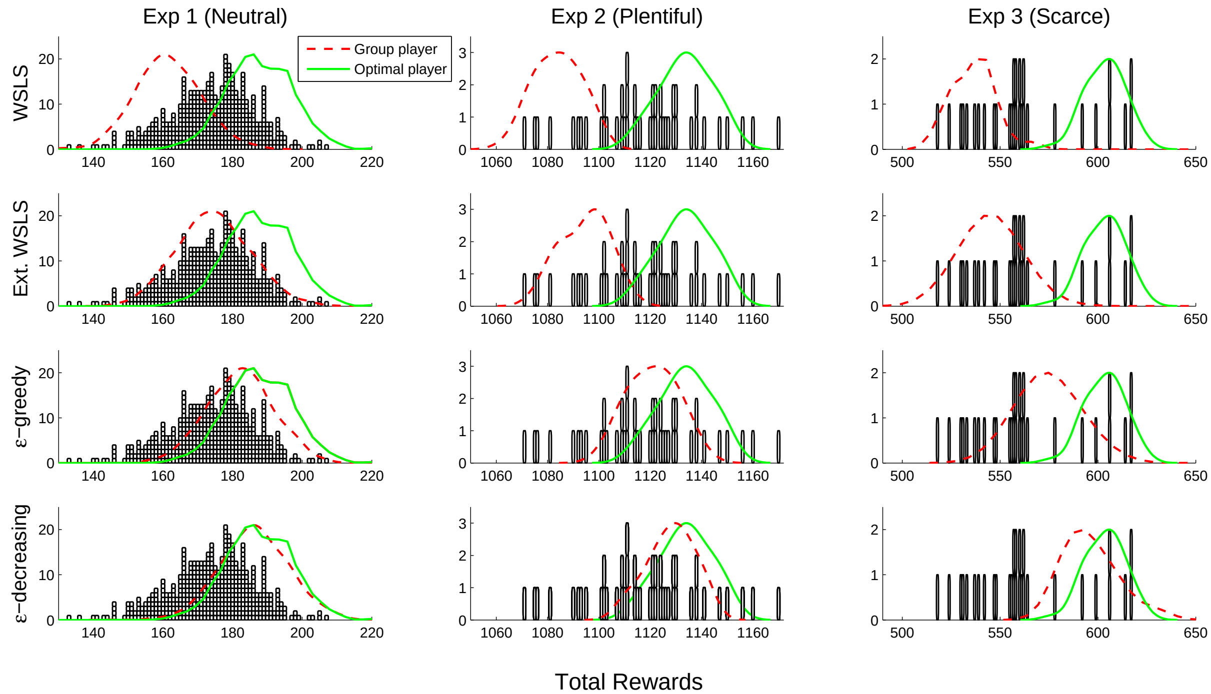

Figure 5: Distribution of rewards for individual participants, the group model, and the optimal decision-making process, for each decision-making model and each experiment. See text for details..

Table 2: Proportion of people using each model, for the three experiments, as measured by the posterior expectation of the ϕ parameter.

Model	Exp. 1	Exp. 2	Exp. 3
WSLS	0%	0%	0%
Extended WSLS	75%	81%	70%
ε-greedy	22%	16%	29%
ε-decreasing	3%	3%	1%

ing each of the four models, and so summarizes individual differences results at this fundamental level. The findings are consistent across all three experiments—even though they have different distributions of reward rates—with the clear majority of the participants inferred to be using the extended WSLS model, and a minority using ε-greedy. The proportion inferred to be using the other two models is negligible.

Wisdom of Crowds Analysis

Our modeling of individual differences in models and parameters immediately allows a range of wisdom of the crowd analyses. The most basic analyses involve taking each of our decision-making models, and using the inferred group mean in the hierarchical analysis, as shown in Table 1 as the aggregate of individual performance.[1] This approach solves the problem of aggregating the knowledge of different people solving different, but related, bandit problems. Rather than aggregating their behavioral choices, we are aggregating the psychology parameter values that lead to those choices.

To complete the model-based wisdom of crowd analyses, we used the group mean parameter values to define a "group model" that used the same decision-process, and completed the same problems given to participants in each of the three experiments. Because the number of rewards obtained is inherently stochastic, we repeated this many times to approximate the distribution of rewards. We also applied the optimal decision-making process to each experiment, to approximate the best possible distribution of rewards for each experiments

The results are shown in Figure 5. The columns correspond to the three experiments. The rows correspond to the WSLS, extended WSLS, ε-greedy and ε-decreasing decision models. Within each panel, the squares piled into histograms show the distribution of performance (i.e., how many rewards were obtained) for the individual participants. The two curves then correspond to the distribution of performance for the group model (red, dotted line) and the optimal decision process (green, solid line).

Figure 5 shows that some of our decision-making

[1]We tried more involved analyses, using the full mixture model in Figure 4 to sample a model, and then parameters, to define a group model. We never found a wisdom of crowd effect comparable to what was achieved with the basic analyses, so we just report those.

models do produce a clear wisdom of the crowds effect, whereas others do not. The distributions of rewards for the group model formed by the WSLS and extended WSLS models does not improve on the distribution of individual performance, and are not close to optimal. For the ε-greedy and ε-decreasing group models, however, there is significant improvement. In particular, the ε-decreasing group model has a distribution of rewards that is extremely close to the optimal distribution for all three experiments.

Discussion

There are some intriguing features of our wisdom of crowd results presented in Figure 5. Most obviously, it is very encouraging that it is possible to take a simple decision-making model like ε-decreasing, take the window it provides onto human decision-making, and produce an aggregate decision-maker that performs near optimally. But, we note that this wisdom of crowd effect is not achieved for all of the cognitive models we tried, and, most particularly, was not achieved for the extended WSLS that provided the best account of the vast majority of individual behavior, as detailed in Table 2.

We think the explanation for this finding is that , the ε-greedy and ε-decreasing models are able to match more closely optimal behavior. Detailed analysis showing this was presented by Lee et al. (2009) and makes intuitive psychological sense. Neither WSLS model is sensitive to which trial in the total sequence is being completed, which is important information in managing the trade-off between early exploration and late exploitation. As a consequence of this sub-optimality, it is not surprising a wisdom of crowd effect was not achieved for these simple models.

What is more surprising is that the effect could be achieved for a decision-making model like ε-decreasing that is not an especially good account of individual behavior. An important topic for future wisdom of crowds research is to identify what properties of cognitive models are important in producing good aggregations of individual knowledge. Being able to mimic optimal behavior is a start, but it is not currently clear how effective models must be able to account for what people do.

More generally, we think our case study with bandit problems demonstrates a very general approach for applying cognitive models to study and use the wisdom of crowds phenomenon. Using graphical models allows hierarchies of parameters, and mixtures of decision processes, to combine the individual differences in people, at the level of their basic knowledge about a task. This leads naturally to a principled sort of aggregation that is applicable to complicated, multidimensional and sequential tasks, which might be among those most needing the pooling of individual capabilities to achieve good performance.

Acknowledgments

This work is was supported by an award from the Air Force Office of Scientific Research (FA9550-07-1-0082).

References

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? Exploration versus exploitation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 933–942.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–879.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society, Series B*, *41*, 148–177.

Kaebling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.

Koller, D., Friedman, N., Getoor, L., & Taskar, B. (2007). Graphical models in a nutshell. In L. Getoor & B. Taskar (Eds.), *Introduction to statistical relational learning.* Cambridge, MA: MIT Press.

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

Lee, M. D., Zhang, S., Munro, M. N., & Steyvers, M. (2009). Using heuristic models to understand human and optimal decision-making on bandit problems. In A. Howes, D. Peebles, & R. Cooper (Eds.), *Proceedings of the Ninth International Conference on Cognitive Modeling — ICCM2009.* Manchester, UK.

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

Shiffrin, R. M., Lee, M. D., Kim, W.-J., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*(8), 1248–1284.

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

Steyvers, M., Lee, M. D., Miller, B., & Hemmer, P. (in press). The wisdom of crowds in the recollection of order information. In J. Lafferty & C. Williams (Eds.), *Advances in neural information processing systems, 23.* Cambridge, MA: MIT Press.

Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, *53*, 168–179.

Surowiecki, J. (2004). *The wisdom of crowds*. New York: Random House.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge (MA): The MIT Press.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*(7), 645–647.