A Simple Sequential Algorithm for Approximating Bayesian Inference

Elizabeth Bonawitz, Stephanie Denison, Annie Chen, Alison Gopnik, & Thomas L. Griffiths
{liz_b, smdeniso, c_annie, gopnik, tom_griffiths}@berkeley.edu
Department of Psychology, 5427 Tolman Hall
Berkeley, CA 94720 USA

Abstract

People can apparently make surprisingly sophisticated inductive inferences, despite the fact that there are constraints on cognitive resources that would make performing exact Bayesian inference computationally intractable. What algorithms could they be using to make this possible? We show that a simple sequential algorithm, Win-Stay, Lose-Shift (WSLS), can be used to approximate Bayesian inference, and is consistent with human behavior on a causal learning task. This algorithm provides a new way to understand people's judgments and a new efficient method for performing Bayesian inference.

Keywords: Bayesian inference; algorithmic level; causal learning

Introduction

In the last five years a growing literature has demonstrated that people often act in ways consistent with optimal Bayesian models (e.g., Griffiths & Tenenbaum, 2005; Goodman, Tenenbaum, Feldman, & Griffiths, 2008). These approaches have provided a precise framework for characterizing intuitive theories and have provided an account of how a learner should update her beliefs as evidence is acquired (Gopnik & Schulz, 2004; Griffiths & Tenenbaum, 2009). The theory-based Bayesian approach has met with much success in describing the inferences made by adults (for a review see Tenenbaum, Griffiths, & Kemp, 2006) and research in cognitive development suggests that children can make similar inferences (Gopnik & Schulz, 2007; Gopnik et al., 2004; Gweon, Schulz, & Tenenbaum, 2010; Kushnir & Gopnik, 2007; Schulz, Bonawitz, & Griffiths, 2007). Taken together, this research strongly suggests that Bayesian statistics provides a productive starting point for understanding human inductive inference.

Theory-based Bayesian approaches have typically been used to give a "computational level" (Marr, 1982) analysis of the inferences people make when solving inductive problems, focusing on the form of the computational problem and its ideal solution. However, it need not be the case that the algorithms people are using to solve these problems actually resemble exact Bayesian inference. Indeed, given the computational complexity of exact Bayesian inference (Russell & Norvig, 2002) and the numerous findings that children and adults alike have difficulty with explicit hypothesis testing (e.g., Kuhn, 1989; Klahr, Fay, & Dunbar, 1993) and sometimes only slowly progress from one belief to the next (Carey, 1991; Wellman, 1990), it becomes interesting to ask how learners might be behaving in a way that is consistent with Bayesian inference.

Here we investigate the algorithms that learners might be using in solving a particular kind of inductive problem – causal learning. These algorithms need to approximate Bayesian inference, but also need to be computationally tractable. One strategy that has proven effective for approximating Bayesian inference in computer science and statistics is using sampling-based approximations, also known as Monte Carlo methods. We introduce a new sequential sampling algorithm based on the Win-Stay, Lose-Shift (WSLS) principle, in which a learner maintains a particular hypothesis until receiving evidence that is inconsistent with that hypothesis. We show that this WSLS algorithm approximates Bayesian inference, and can do so quite efficiently.

Previous work in cognitive psychology has shown that people follow a WSLS strategy in concept learning tasks (Restle, 1962; Levine, 1975). We use this as the starting point for an investigation of whether human behavior that approximates Bayesian inference in causal learning might be explained in terms of a WSLS strategy. We compare the WSLS algorithm to simply sampling from the posterior distribution as an account of human behavior in a simple causal learning task. Both algorithms predict that people should behave in a way that is consistent with Bayesian inference, but WSLS also predicts that there should be a characteristic pattern of dependency between people's successive responses.

The plan of the paper is as follows. First, we introduce the causal learning task that will be the focus of our analysis, and summarize how Bayesian inference can be applied in this task. We then introduce the idea of sequential sampling algorithms, including our new WSLS algorithm. This is followed by a mathematical analysis of the WSLS algorithm, showing that it approximates Bayesian inference. The remainder of the paper focuses on an experiment in which we evaluate how well the WSLS algorithm captures people's judgments in our causal learning task.

Bayesian inference and causal learning

While the algorithms that we present in this paper will apply to any inductive problem with a discrete hypothesis space, we will make our analysis concrete by focusing on a simple causal learning problem. In this problem, there are three categories of objects: red, green, and blue blocks. Each of these kinds of blocks activate a machine with different probability when they are placed on the machine. The red blocks activate the machine on five out of six trials, the green blocks on three out of six trials, and the blue blocks on just once out of six trials. A new block is then presented, which has lost its color, and needs to be classified as either a red, green, or blue block, based on some observations of what happens when it is placed on the machine over a series of trials. The question

is what people will infer about the causal properties of this block, and which class of blocks it belongs to.

Given this hypothesis space, we can consider how an ideal learner should update his or her beliefs in light of the evidence provided by its interaction with the machine. Assume that the learner begins with a prior probability distribution over hypotheses, $P(h)$, where the probability assigned to each hypothesis reflects the degree of belief in each hypothesis being true before seeing any data. Given some observed data d, reflecting whether the block activates the machine on a single trial, the learner obtains a posterior distribution over hypotheses, $P(h|d)$, via Bayes' rule:

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h' \in \mathcal{H}} P(d|h')P(h')} \quad (1)$$

where $P(d|h)$ is the likelihood, indicating the probability of observing d if h were true, and $\mathcal{H}$ is the hypothesis space.

Often, Bayesian inference is performed in a sequential setting, with a series of observations being made one after another, and the posterior distribution being updated after each observation. This is the case with our causal learning problem, where we receive a sequence of observations of the block interacting with the machine on successive trials, rather than a single observation. Letting $d_1, \ldots, d_n$ denote observations after n trials, we are interested in the posterior distribution $P(h|d_1, \ldots, d_n)$. This can be computed via Equation 1, substituting $d_1, \ldots, d_n$ for d. However, it can be simpler to follow a sequential updating rule, which allows us to compute the posterior after observing $d_1, \ldots, d_{n+1}$ from the posterior based on $d_1, \ldots, d_n$. Formally, this is

$$P(h|d_1, \ldots, d_{n+1}) = \frac{p(d_{n+1}|h)p(h|d_1, \ldots, d_n)}{\sum_{h'} p(d_{n+1}|h')p(h'|d_1, \ldots, d_n)} \quad (2)$$

where we assume that the observations d_i are conditionally independent given h (i.e., that a block has an independent chance of activating the machine on each trial, once its color is known).

Sequential sampling algorithms

The Bayesian analysis presented in the previous section provides an abstract, "computational level" (Marr, 1982) characterization of causal induction, identifying the underlying problem and how it might best be solved. We now turn to the problem of how to approximate this optimal solution. Simply implementing Bayesian inference by listing all hypotheses and then updating them following Bayes' rule quickly becomes intractable, as it requires considering each hypothesis after every observation. We thus consider the possibility that people may be approximating Bayesian inference by following a procedure that produces samples from the posterior distribution. This idea is consistent with the prevalence of Monte Carlo methods for approximating Bayesian inference in computer science and statistics (e.g., Robert & Casella, 2004), as well as with behavioral evidence that people select hypotheses in proportion to their posterior probability (Goodman et al., 2008; Denison, Bonawitz, Gopnik, & Griffiths, 2010).

Independent sampling is the simplest kind of Monte Carlo method, and is thus a parsimonious place to start in considering the algorithms learners might use. In particular, the problem of sequentially updating a posterior distribution in light of evidence can be solved approximately using sequential Monte Carlo methods such as particle filters (Doucet, Freitas, & Gordon, 2001). A particle filter approximates the probability distribution over hypotheses at each point in time with a set of samples (or "particles"), and provides a scheme for updating this set to reflect the information provided by new evidence. The behavior of the algorithm depends on the number of particles. With a very large number of particles, each particle is similar to a sample from the posterior. With a small number of particles, there can be strong sequential dependencies in the representation of the posterior distribution. Recent work has explored particle filters as a way to explain patterns of sequential dependency that arise in human inductive inference (Sanborn, Griffiths, & Navarro, 2006; Levy, Reali, & Griffiths, 2009).

Particle filters have many degrees of freedom, with many different schemes for updating particles being possible (Doucet et al., 2001). They also require learners to maintain multiple hypotheses at each point in time. Here, we investigate a simpler algorithm that assumes learners maintain a single hypothesis, resampling from the posterior with a probability dependent on the degree to which the hypothesis is contradicted by data. This is similar to using a particle filter with just a single particle, with a computationally expensive resampling step being more likely to be carried out as that particle becomes inconsistent with the data. Because of its tendency to maintain a hypothesis that makes a successful prediction and change hypotheses when this is not the case, we call this the Win-Stay, Lose-Shift (WSLS) algorithm.

The WSLS principle has a long history both in computer science, where it appears as a heuristic algorithm in reinforcement learning and game theory (Robbins, 1952; Nowak & Sigmund, 1993), and in psychology, where it has been proposed as an account of human concept learning (Restle, 1962). The WSLS strategy has also been shown in children, especially between the ages of three- to five-years-old (Levine, 1975). More recently, WSLS has been analyzed as a simple model of learning that leads to interesting strategies in game theory (Nowak & Sigmund, 1993). In the remainder of the paper, we show that this kind of strategy can yield a simple method for approximating Bayesian inference, and appears to be consistent with human behavior.

Analyzing the Win-Stay, Lose-Shift algorithm

A first step towards exploring the WSLS algorithm is to show that it can be used to approximate Bayesian inference. In this section, we define the WSLS algorithm we will be analyzing, and contrast it to simply sampling from the posterior distribution (which we will term Random Sampling, or RS). Random Sampling assumes that learners draw a new sample from the posterior distribution every time they need to make

a response, which requires evaluating all hypotheses or using a sequential algorithm such as a particle filter with a large set of particles. Taking independent samples from the posterior distribution has two consequences. First, when we consider the behavior of a group of people, the proportion of people selecting each hypothesis will match the posterior probability. Second, successive responses from an individual will be independent of one another. We will show that WSLS shares the first of these properties with RS, but not the second, making it possible to separate these two algorithms empirically.

The simplest version of the WSLS algorithm assumes that learners maintain their current hypothesis provided they see data that are consistent with that hypothesis, and generate a new hypothesis otherwise. This is the version explored by Restle (1962). It is relatively straightforward to show that this can approximate Bayesian inference in cases where the likelihood function $p(d_i|h)$ is deterministic, giving a probability of 1 or 0 to any observation d_i for every h, and observations are independent conditioned on hypotheses. More precisely, the marginal probability of selecting a hypothesis h_n given data $d_1,\ldots,d_n$ is the posterior probability $p(h_n|d_1,\ldots,d_n)$, provided that hypotheses are generated from the posterior distribution whenever the learner chooses to shift hypotheses.

We now turn to a proof of the more general case, in nondeterministic settings. We will do this by considering the conditions required for an argument by induction to apply. First, we assume that $h_n \sim p(h_n|d_1,\ldots,d_n)$. We define the transition kernel of the WSLS algorithm, $q(h_{n+1}|h_n)$, to be:

$$h_{n+1}|h_n \sim \begin{cases} \delta(h_n) & \text{with probability } \phi \\ p(h_{n+1}|d_1,\ldots,d_{n+1}) & \text{with probability } 1-\phi \end{cases}$$

where $\delta(h)$ is the distribution putting all of its mass on h, and ϕ is the probability of staying, which is a function of $d_1,\ldots,d_{n+1}$ and h_n. The distribution over hypotheses after observing d_{n+1} is given by

$$\begin{aligned} & q(h_{n+1}=h|d_1,\ldots,d_{n+1}) \\ &= \sum_{h_n} q(h_{n+1}=h|h_n)p(h_n|d_1,...,d_n) \\ &= \sum_{h_n} \Big(\delta(h_n,j)\phi + (1-\phi)p(h_{n+1}|d_1,...,d_{n+1})\Big)p(h_n|d_1,...,d_n) \\ &= p(h_{n+1}=h|d_1,...,d_{n+1})(1-\mathbb{E}(\phi)) + \phi p(h_{n+1}=h|d_1,...,d_n) \end{aligned}$$

where the expectation, $\mathbb{E}(\phi)$, is with respect to $p(h_n|d_1,...,d_n)$.

We now examine the conditions on ϕ such that $q(h_{n+1}|d_1,\ldots,d_{n+1}) = p(h_{n+1}|d_1,\ldots,d_{n_1})$, corresponding to the conditions required for the marginal distribution under WSLS to match the posterior. If we take

$$\phi = c\frac{p(h_{n+1}=h|d_1,...,d_{n+1})}{p(h_n=h|d_1,...,d_n)} = c\frac{p(d_{n+1}|h_{n+1}=h)}{p(d_{n+1}|d_1,...,d_n)}$$

where c is a constant which can depend on $d_1,\ldots,d_{n+1}$ but is invariant over hypotheses, we obtain

$$\begin{aligned} & q(h_{n+1}=h|d_1,\ldots,d_{n+1}) \\ &= p(h_{n+1}=h|d_1,...,d_{n+1})(1-c) + cp(h_{n+1}=h|d_1,...,d_{n+1}) \end{aligned}$$

which is just $p(h_{n+1}=h|d_1,\ldots,d_{n+1})$. This is because

$$\begin{aligned} \mathbb{E}(\phi) &= \mathbb{E}_{p(h_n|d_1,\ldots,d_n)}\left\{c\frac{p(d_{n+1}|h_{n+1}=h)}{p(d_{n+1}|d_1,...,d_n)}\right\} \\ &= c\frac{\sum_h p(d_{n+1}|h)p(h|d_1,...,d_n)}{p(d_{n+1}|d_1,...,d_n)} = c \end{aligned}$$

This provides us a simple set of conditions under which our criterion is satisfied, with $q(h_{n+1}=h|d_1,...,d_{n+1}) = p(h_{n+1}=h|d_1,...,d_{n+1})$ for any c such that $\phi \in [0,1]$.

There are two interesting special cases to consider. The first arises when we take $c = p(d_{n+1}|d_1,...,d_{n+1})$. In this case, $\phi = p(d_{n+1}|h_{n+1}=h)$. This results in a simple algorithm that makes a choice to resample based on the likelihood associated with the current observation, given the current h. That is, with probability proportional to this likelihood, the learner resamples from the full posterior.

The second special case is the most efficient algorithm of this kind, in the sense that it minimizes the rate at which sampling from the posterior is required. This corresponds to taking $c = \frac{p(d_{n+1}|d_1,\ldots,d_{n+1})}{\max_h p(d_{n+1}|h)}$, resulting in $\phi = \frac{p(d_{n+1}|h_{n+1}=h)}{\max_h p(d_{n+1}|h)}$. For some hypothesis spaces, it may be possible to compute ϕ in advance for all possible data and hypotheses. After this single costly computation is complete, the learner need only look up the values.

This proof shows that the marginal distribution over hypotheses after observing d_n will be the same for any n. However, there are still important differences in what Win-Stay, Lose-Shift predicts for the dependency between guesses for a particular individual as compared to Random Sampling. Namely, there is no dependency between h_n and h_{n+1} in RS, but there is for WSLS: if the data are consistent with h_n, then the learner will retain h_n with probability proportional to $p(d_i|h_n)$ rather than randomly sampling h_{n+1} from the posterior distribution. We can use this difference to attempt to diagnose the algorithm that people are using when they are solving a causal learning problem.

Evaluating inference strategies in people

We now turn to the question of whether people's responses are well captured by the algorithms described above. Namely, we might expect that if participants behave in ways consistent with the WSLS algorithm we should observe dependencies between their responses; specifically, participants should retain hypotheses that are consistent with the evidence, and resample proportional to the likelihood, $p(d|h)$. However, if participants behave in ways consistent with Random Sampling, then responses will be resampled from the posterior distribution regardless of previous guesses, such that there are no dependencies between responses.

Methods

Participants and Design Participants were 65 undergraduates recruited from an introductory psychology course. The participants were split into 2 conditions ($N = 28$ in the "On-

First" condition; $N = 32$ in the "Off-First" condition; 5 participants were excluded for not completing the experiment).

Stimuli Stimuli consisted of 13 white cubic blocks ($1cm^3$). Twelve blocks had custom-fit sleeves made from construction paper of different colors: 4 red, 4 green, and 4 blue. An activator bin large enough for 1 block sat on top of a [15" x 18.25" x 14"] box. Attached to this box was a helicopter toy that lit up when activated. The activator button for the toy was inside the box hidden from view. There was a set of On cards that pictorially represented the toy in the on position, and a set of Off cards that pictorially represented the toy in the off position. Because participants were tested in large groups, a computer slideshow that depicted the color of the blocks and the cards was used to illustrate the evidence shown.

Procedure Participants in each condition were tested on separate days in two large groups. Participants were instructed to record responses using paper and pen and not to change answers provided for previous questions after viewing subsequent evidence. Participants were told that different blocks possess different amounts of "blicketness," a fictitious property. Blocks that possess the most blicketness almost always activate the machine, blocks with very little blicknetness almost never activate the machine, and blocks with medium blicketness activate the machine half of the time. A red block was chosen at random and placed in the activator bin. The helicopter toy either turned on or remained in the off position. The experimenter explained that a corresponding On or Off card was placed on the table to depict the event and the computer slidehow slide showed the same evidence. The card remained on the table and the computer slideshow remained on the screen throughout the experiment. After 5 more repetitions using the same red block for a total of 6 demonstrations, participants were told that red blocks have the most blicketness (they activated 5/6 times). The same procedure was repeated for the blue and green blocks with the blue blocks having very little blicketness (activating 1/6 times), and green blocks having medium blicketness (activating 3/6 times). All evidence remained visible on the computer slideshow. To ensure that participants were paying attention, they were asked to match each color to the proper degree of blicketness (most, very little, medium) by writing down their responses.

After the memory check, a novel white block that lost its fitted-sleeve was presented and participants were asked to write down an initial guess about what color fitted-sleeve the white block should have (red, green, or blue). The white block was then placed into the activator bin four times and each time the participant saw whether or not the toy activated. After each demonstration, the appropriate On or Off card was chosen and the slideshow was advanced to represent the state of the toy. Participants were asked to record their best guess about what color they believed the block to be after each demonstration, but before each guess was made

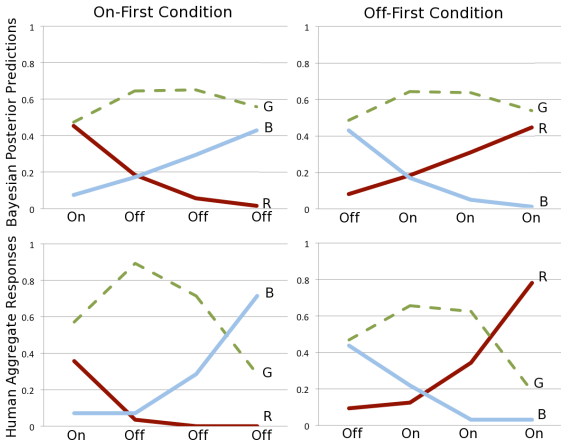

Figure 1: Bayesian posterior probability and human data for each block, red (R), green (B), and blue (B) after observing each new instance of evidence, using parameters estimated from fitting the Bayesian model to the data.

the participants were told, "It's okay if you keep thinking it is the same color and it is also okay if you change your mind." In the On-first condition the toy turned on for the first trial, and did not activate on the three subsequent trials. In the Off-first condition the toy did not activate on the first trial, but turned on for the three subsequent trials.

Results

Comparison to Bayesian inference Responses were uniquely and unambiguously categorized as "red", "green", and "blue". There was a slight bias to favor green blocks (60%), with red (25%) and blue (15%) blocks being less favored.[1] We determined the parameters for the prior distribution and likelihood in two ways. For the first way ("initial responses") priors were determined by the participants' initial block color predictions and the likelihood of block activation was determined by the initial observations of block activations during the demonstration phase ($5/6$ red, $1/2$ green, $1/6$ blue). For the second way ("maximized") we searched for the set of priors and the likelihood activation weights that would maximize the log-likelihood for the model.[2] Using either set of parameters, participant responses were well captured by the posterior probability (initial responses: $r(22) = .76, p < .001$; maximized: $r(22) = .85, p < .0001$, see Figure 1). The primary difference between the model and data is that people seem to change their beliefs more strongly than the model predicts. This may be a consequence of pedagogical reasoning, a point we return to in the Discussion.

[1] Such a bias is consistent with people's interest in non-determinism; the green blocks were the most stochastic in that they activated on exactly half the trials.

[2] The maximized priors were .27 red, .48 green, .25 blue; these priors correspond strongly to the priors represented by participants. The maximized likelihood was .85 red, .5 green, .16 blue which also corresponds strongly to the likelihood given by the initial activation observations.

Comparison to WSLS and RS To compare people's responses to the WSLS and RS algorithms, we first calculated the "switch" probabilities under each model in the two ways previously described: using the parameters from the initial responses and using the previously estimated maximized parameters. Calculating switch probabilities for RS is relatively easy: because each sample is independently drawn from the posterior, the switch probability is simply calculated from the posterior probability of each hypothesis after observing each piece of evidence. Switch probabilities for WSLS were calculated such that resampling is based only on the likelihood associated with the current observation, given the current h. That is, with probability equal to this likelihood, the learner resamples from the full posterior. Responses were much better captured by the WSLS algorithm using the maximized parameters ($r(15) = .81, p < .0001$) and the parameters given by participant initial responses ($r(15) = .78, p < .001$) as compared to the RS algorithm (maximized: $r(15) = .58, p = .02$; initial responses: $r(15) = .39, p = ns$). See Figure 2. We also computed the log-likelihood scores for both models. The WSLS model better fit the data than the RS model (initial responses: $p(d|WSLS) = -221$, $p(d|RS) = -262$; maximized: $p(d|WSLS) = -215$, $p(d|RS) = -251$). These results suggest that the pattern of dependencies between people's responses are better captured by the WSLS algorithm than by an algorithm such as RS that produces independent samples.

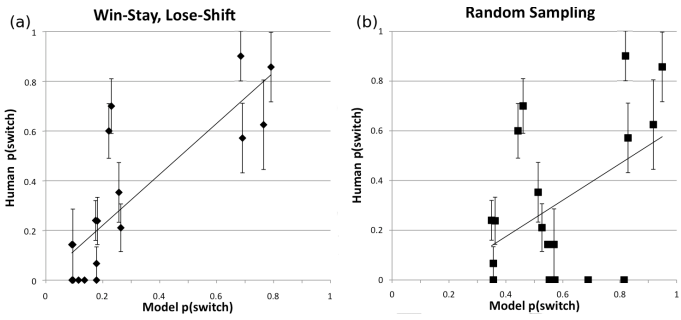

Figure 2: Correlations between the probability of switching hypotheses in the models given the maximized parameters and the human data, for (a) the Win-Stay Lose-Shift algorithm and (b) Random Sampling.

Discussion

Our results show how tracking learning at the level of the individual can help us understand the specific algorithms that learners might be using to approximate Bayesian inference. First we introduced an algorithm, Win-Stay, Lose-Shift that approximates Bayesian inference by maintaining a single hypothesis over time, and proved that the marginal distribution over hypotheses after observing data will always be the same for this algorithm as for sampling from the posterior (Random Sampling). That is, both algorithms return a distribution over responses consistent with the posterior distribution obtained from Bayesian inference. We provided an analysis of WSLS with two special cases. The first case resulted in a simple algorithm that makes a choice to resample based on the likelihood associated with the current observation, given the current hypothesis. The second is the most efficient algorithm of this kind in that it minimizes the rate at which sampling from the posterior is required, and may thus be of interest for approximating Bayesian inference in other settings.

Our analysis also made it clear that there are important differences in what WSLS and RS predict for the dependency between guesses, making it possible to separate these algorithms empirically. We explored the algorithms that people use for solving inductive inference problems through an experiment using a simple causal learning task. In this experiment, people's overall responses are consistent with Bayesian inference, but people show dependencies between responses characteristic of the WSLS algorithm, rather than independently sampling responses each time from the posterior. These results extend previous work exploring WSLS strategies, showing that at least one strategy of this kind provides a viable way to approximate Bayesian inference, demonstrating that causal induction contains problems from this class, and providing evidence that WSLS is an appropriate algorithm for describing people's inferences.

Connecting the computational and algorithmic levels is a significant challenge for Bayesian models of cognition and this is only a beginning step in understanding the psychological processes at work in causal inference. We believe that there are several important directions for future research in this area. First, it would be interesting to test the algorithm's predictions across various psychological experiments that have relied purely on a Bayesian inference approach; this would allow for a better assessment of the WSLS algorithm's efficiency. Second, both algorithms can be seen as extreme versions of particle filters: Random Sampling in cases where there are a large set of particles drawn from the posterior and randomly drawing one member of the set at random for each query; and, Win-Stay Lose-Shift, which is similar to using a single particle that is resampled from the posterior when the particle becomes inconsistent with the data. There may be some value in exploring algorithms that lie between these extremes, with a more moderate number of particles as well as exploring algorithms that shift from one hypothesis to the next by modifying the current hypothesis in a principled and structured manner. Considering intermediate models would also allow future work to examine the degree to which fewer or greater numbers of particles capture inference and to what degree these constraints change with age and experience. Third, we constrained our space to a modest number of hypotheses, but other work has begun to examine how hypothesis spaces may be learned and simultaneously searched; this should be jointly developed with approaches taken here that explore the space of plausible algorithms that capture people's causal inferences. Fourth, in this particular task, the aggregate distribution of adult responses shifted more dramatically than the Bayesian model presented here predicted.

It is likely, given the context of showing participants a predetermined computer slideshow, that adults were making a pedagogical assumption (Shafto & Goodman, 2008) which would better capture the data. Future work may investigate this possibility.

Young children have particularly limited cognitive resources (e.g., German & Nichols, 2003; Gerstadt, Hong, & Diamond, 1994; Siegler, 1975), but are nonetheless capable of behaving in a way that is consistent with optimal Bayesian models. Children must thus be especially adept at managing limited resources to approximate Bayesian inference. Arguably, many of the most interesting cases of belief revision happen in the first few years of life (Wellman, 1990; Bullock, Gelman, & Baillargeon, 1982; Carey, 1985; Gopnik & Meltzoff, 1997). Understanding more precisely how specific algorithms shape children's learning may provide a potential solution to the problem of how limited cognitive resources and Bayesian frameworks of children's cognition can be reconciled. We are currently investigating these questions.

While there is still important work to be done, connecting the algorithmic level to the computational level is a first step in understanding the algorithms that learners may be using to approximate Bayesian inference. We have demonstrated that the WSLS algorithm, previously provided as a model of human hypothesis testing, can be used to approximate Bayesian inference. This provides a way to perform sequential Bayesian inference while maintaining only a single hypothesis at a time, and leads to an efficient approximation scheme that might be useful in computer science and statistics. We have also shown that a WSLS algorithm seems to capture people's judgments in a simple causal learning task. Our results add to the growing literature suggesting that even responses by an individual that may appear non-optimal may in fact represent an approximation to a rational process.

Acknowledgments. This research was supported by the McDonnell Foundation Causal Learning Collaborative, grant IIS-0845410 from the National Science Foundation, and grant FA-9550-10-1-0232 from the Air Force Office of Scientific Research.

References

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (p. 209-254). New York: Academic Press.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & S. Gelman (Eds.), *Epigenesis of mind: Essays on biology and cognition.* Hillsdale, NJ: Erlbaum.

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. (2010). Preschoolers rationally sample hypotheses. In *Proceedings of the 32nd annual conference of the cognitive science society.*

Doucet, A., Freitas, N. de, & Gordon, N. (2001). *Sequential Monte Carlo methods in practice.* New York: Springer.

German, T., & Nichols, S. (2003). Children's inferences about long and short causal chains. *Developmental Science*, *6*, 514-523.

Gerstadt, C., Hong, Y., & Diamond, A. (1994). The relationship between cognition and action: performance of children 3 - 7 years old on a stroop-like day-night test. *Cognition*, *53*, 129-153.

Goodman, N., Tenenbaum, J., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32:1*, 108-154.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and bayes nets. *Psychological Review*, *111*, 1-31.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Science*, *8*, 371-377.

Gopnik, A., & Schulz, L. (Eds.). (2007). *Causal learning: Psychology, philosophy, and computation*. Oxford: Oxford University Press.

Griffiths, T., & Tenenbaum, J. (2009). Theory-based causal induction. *Psychological Review*, *116*, 661-716.

Griffiths, T., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 354-384.

Gweon, H., Schulz, L., & Tenenbaum, J. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Science*, *107*(20), 9066-9071.

Klahr, D., Fay, A., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, *25*, 111-146.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674-689.

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *44*, 186-196.

Levine, M. (1975). *A cognitive theory of learning: Research on hypothesis testing*. Hillsdale, NJ: Lawrence Erlbaum.

Levy, R., Reali, F., & Griffiths, T. L. (2009). Modeling the effects of memory on human online sentence processing with particle filters. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 21* (pp. 937–944).

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, *364*, 56-58.

Restle, F. (1962). The selection of strategies in cue learning. *Psychological Review*, *69*, 329-343.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *58*, 527-535.

Robert, C., & Casella, G. (2004). *Monte Carlo statistical methods (2nd ed.)*. New York: Springer.

Russell, S. J., & Norvig, P. (2002). *Artificial intelligence: A modern approach* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared make your tummy ache? naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, *43*, 1124-1139.

Shafto, P., & Goodman, N. (2008). Teaching games: Statistical sampling assumptions for pedagogical situations. In *Proceedings of the 30th annual conference of the cognitive science society.*

Siegler, R. (1975). Defining the locus of developmental differences in children's causal reasoning. *Journal of Experimental Child Psychology*, *20*, 512-525.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Science*, *10*, 309-318.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.