Exploring the Influence of Particle Filter Parameters on Order Effects in Causal Learning

Joshua T. Abbott (joshua.abbott@berkeley.edu)
Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, University of California at Berkeley, Berkeley, CA 94720 USA

Abstract

The order in which people observe data has an effect on their subsequent judgments and inferences. While Bayesian models of cognition have had some success in predicting human inferences, most of these models do not produce order effects, being unaffected by the order in which data are observed. Recent work has explored approximations to Bayesian inference that make the underlying computations tractable, and also produce order effects in a way that seems consistent with human behavior. One of the most popular approximations of this kind is a sequential Monte Carlo method known as a particle filter. However, there has not been a systematic investigation of how the parameters of a particle filter influence its predictions, or what kinds of order effects (such as primacy or recency effects) these models can produce. In this paper, we use a simple causal learning task as the basis for an investigation of these issues. Both primacy and recency effects are seen in this task, and we demonstrate that both kinds of effects can result from different settings of the parameters of a particle filter.

Keywords: particle filters; order effects; causal learning; rational process models

Introduction

How do people make such rapid inferences from the constrained available data in the world and with limited cognitive resources? Previous research has provided a great deal of evidence that human inductive inference can be successfully analyzed as Bayesian inference, using rational models of cognition (Anderson, 1990; Oaksford & Chater, 1998; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). Rational models answer questions at Marr's (1982) computational level of analysis, producing solutions to *why* humans behave as they do, whereas traditional models from cognitive psychology tend to analyze cognition on Marr's level of algorithm and representation, focusing instead on *how* cognitive processes support these behaviors. Although Bayesian models have become quite popular in recent years, it remains unclear what psychological mechanisms could be responsible for carrying out these computations. Of particular concern is that the amount of computation required in these models becomes intractable in real-world scenarios with many variables, yet people make rather accurate inferences effortlessly in their everyday lives. Are people implicitly approximating these probabilistic computations?

Monte Carlo methods have become a primary candidate for connecting the computational and algorithmic levels of analysis (Sanborn, Griffiths, & Navarro, 2006; Levy, Reali, & Griffiths, 2009; Shi, Feldman, & Griffiths, 2008). The basic principle underlying Monte Carlo methods is to approximate a probability distribution using only a finite set of samples from that distribution. Recent work has focused on two Monte Carlo methods in particular: importance sampling and particle filtering. Importance sampling draws samples from a known proposal distribution and weights these samples to correct for the difference from the desired target distribution. Particle filters are a sequential Monte Carlo method that uses importance sampling recursively. When approximating Bayesian inference, the posterior distribution is represented using a set of discrete samples, known as *particles*, that are updated over time as more data are observed. These methods can be shown to be formally related to existing psychological process models such as exemplar models (Shi et al., 2008), and can be used to explain behavioral data inconsistent with standard Bayesian models in categorization (Sanborn et al., 2006), sentence parsing (Levy et al., 2009), and classical conditioning experiments (Daw & Courville, 2008). However, there has not previously been a systematic investigation of how the parameters of these Monte Carlo methods affect the predictions they make.

In this paper we explore how the parameters of particle filters affect the predictions that they make about order effects, using a simple causal learning task to provide a context for this exploration. It is a common finding that the order in which people receive information has an effect on their subsequent judgments and inferences (Dennis & Ahn, 2001; Collins & Shanks, 2002). This poses a problem for rational models based on Bayesian inference as the process of updating hypotheses in these models is typically invariant to the order in which the data are presented. Previous work has shown that particle filters can produce order effects similar to those seen in human learners (e.g., Sanborn et al., 2006). However, this work has focused on primacy effects, in which initial observations have an overly strong influence on people's conclusions. In other settings, people produce recency effects, being more influenced by more recent observations. Causal learning tasks can result in both primacy and recency effects, with surprisingly subtle differences in the task leading to one or the other (Dennis & Ahn, 2001; Collins & Shanks, 2002). Causal learning thus provides an ideal domain in which to examine how the parameters of particle filters influence their predictions, and what kinds of order effects these models can produce.

The plan of the paper is as follows. In the next section we discuss previous empirical and theoretical work on human causal learning, showing different kinds of observed order effects and providing the Bayesian framework we will be working in. We then formally introduce particle filters, followed by our investigation of how varying certain particle filter parameters controls order effects. After this, we use our newfound understanding of the parameters to model different order effects in previous experiments. Finally, we discuss the implications of our work and future directions for research.

Order Effects in Causal Learning

We focus our investigation of order effects in causal learning on a pair of studies based on sequences of covarying events. Dennis and Ahn (2001) presented participants with a series of trials indicating whether or not a plant was ingested and whether or not this resulted in an allergic reaction. The sequence of trials was split into two equal blocks of covarying events; one primarily indicating a generative causal relationship between plant and reaction, and the other primarily indicating a preventative relationship. The overall contingency of the combined blocks was 0. The blocks were presented one after the other, with the initial block chosen randomly, and after observing all trials participants were asked to make a strength judgement (-100 to 100) on the causal relationship they thought existed between plant and reaction. After answering, the blocks were presented again in reverse order and the subjects were asked to make another judgement. If the generative relationship block was presented first, followed by the preventative block, participants responded with a preference for a generative relationship (M=17.67, SD=25.66). However, if the preventative block was presented first, participants responded that only a weak preventive causal relationship existed (M=-5.50, SD=22.27). These results indicate a primacy effect in favor of generative causal relationships.

The primacy effect found by Dennis and Ahn (2001) was contradictory to previous models of associative strength that showed recency effects, and a subsequent follow-up study was conducted that examined the role of judgment frequency in producing different kinds of order effects (Collins & Shanks, 2002). In this study, the authors used exactly the same trial sequence as Dennis and Ahn (2001), but asked the participants for a causal strength judgment after every 10 trials rather than just one final strength judgement. Making frequent judgments resulted in recency effects, with participants producing more generative estimates when they saw the generative data most recently (M=58.4, SD=34.2), and more preventative estimates when they saw the preventative data most recently (M=-23.3, SD=39.3).

Causal Learning as Bayesian Inference

We can formulate this causal learning problem as a problem of Bayesian inference through the use of a causal graphical model framework similar to that used in Griffiths and Tenenbaum (2005). In this framework, we assume a single directed graph structure defined over three binary variables: an effect E, a potential cause C, and a background cause B which captures all other causes of E that are not C. Additionally, there are strengths, s_0 and s_1, that indicate how strongly B and C influence the presence, or lack thereof, of E. This graphical model is shown in Figure 1.

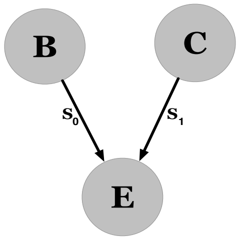

Figure 1: Directed graph involving three binary variables – (B)ackground, (C)ause, and (E)ffect – relevant to causal induction, and two edge weights – s_0 and s_1 – indicating how strongly B and C influence E, respectively.

Using this graphical model we are interested in the conditional probability $P(E|B,C)$ and want to evaluate how well the strength weights predict the observed data. Motivated by the models used in Griffiths and Tenenbaum (2005) and Cheng (1997), we define the conditional probability using the noisy-OR and noisy-AND-NOT functions for generative and preventative causes respectively. Assuming a background cause is always present (ie. $B = 1$), we will get a sequence of events in the form "Cause was ($C = 1$) or was not ($C = 0$) present and an effect did ($E = 1$) or did not ($E = 0$) occur". This gives us four possible conditions to evaluate. Depending on the sign of s_1 – the strength of C causing/preventing E, we compute either the noisy-OR ($s_1 \geq 0$, a generative cause) or the noisy-AND-NOT ($s_1 < 0$, a preventative cause). Table 1 presents the resulting probabilities.

Table 1: The noisy-OR (s_1 is positive) and noisy-AND-NOT (s_1 is negative) functions

C	E	s_1 is positive	s_1 is negative
1	1	$s_0 + s_1 - s_0 s_1$	$s_0(1 + s_1)$
1	0	$1 - (s_0 + s_1 - s_0 s_1)$	$1 - [s_0(1 + s_1)]$
0	1	s_0	s_0
0	0	$1 - s_0$	$1 - s_0$

To complete the definition of this Bayesian model, we need to specify the prior distribution that is assumed on the strengths s_0 and s_1. Our starting assumption is that these weights are each drawn from a uniform prior over their entire range, with $s_0 \sim \text{Uniform}(0,1)$ and $s_1 \sim \text{Uniform}(-1,1)$, where negative values of s_1 imply a preventative cause, as detailed above. This basic model assumes that s_0 and s_1 remain constant over time. A slightly more complex model would allow the strength of causes to drift, taking on a value that is close to the value on the previous trial but with some stochastic variation. We can do this by assuming that s_0 and s_1 have a prior on each trial that is based on their value on the previous trial. We assume that s_0 follows a Beta distribution, $\text{Beta}(\lambda s + 1, \lambda(1 - s) + 1)$ where s is the value on the previous trial and λ controls the rate of drift, with large values indicating a slow drift. We assume that s_1 preserves its sign (and is thus fixed as generative or preventative), but its absolute value is drawn from a Beta distribution in the same way. For the first trial, s_0 and s_1 are assumed to follow the uniform prior distributions given above.

With our Bayesian model of causal learning defined, we now turn to the problem of inference. In the next two sections we introduce the general schema for a particle filter and then indicate how it can be applied to the specific model of causal learning we have outlined in this section.

Particle Filters

A particle filter is a sequential Monte Carlo method that can be used to approximate a sequence of posterior distributions, as is necessary when performing Bayesian inference repeatedly in response to a sequence of observations (Doucet, Freitas, & Gordan, 2001). When using a particle filter, it is typically assumed we have a sequence of unobserved latent variables $z_1, \ldots, z_t$, where $\mathbf{z}_{0:t}$ is modeled as a Markov process with prior distribution $P(z_0)$ and transition probability $P(z_t|z_{t-1})$. We then have a sequence of observed variables $y_1, \ldots, y_t$, and are attempting to estimate the posterior distribution $P(\mathbf{z}_{0:t}|\mathbf{y}_{1:t})$. The posterior is given by Bayes' rule, for any time t, as:

$$P(\mathbf{z}_{0:t}|\mathbf{y}_{1:t}) \propto P(\mathbf{y}_{1:t}|\mathbf{z}_{0:t})P(\mathbf{z}_{0:t}).$$

We can obtain a recursive formula for this as:

$$P(\mathbf{z}_{0:t+1}|\mathbf{y}_{1:t+1}) \propto P(\mathbf{z}_{0:t}|\mathbf{y}_{1:t})P(y_{t+1}|z_{t+1})P(z_{t+1}|z_t).$$

Assuming we have a set of samples from $P(\mathbf{z}_{0:t}|\mathbf{y}_{1:t})$, importance sampling can be used to approximate this posterior distribution by sampling from $P(z_{t+1}|z_t)$ for each value of z_t in our sample, weighting each value of z_{t+1} by $P(y_{t+1}|z_{t+1})$, and then resampling from this weighted distribution. The result will be a set of samples that approximate $P(\mathbf{z}_{0:t+1}|\mathbf{y}_{1:t+1})$. The recursive nature of this approximation, where we can obtain samples from $P(\mathbf{z}_{0:t}|\mathbf{y}_{1:t})$ given samples from $P(\mathbf{z}_{0:t-1}|\mathbf{y}_{1:t-1})$, leads to a natural algorithm. This algorithm, in which a set of samples is constantly updated to reflect the information provided by each observation, is known as a particle filter. The samples are referred to as particles.

Particle Filter Template

We will examine variants of a particle filter based on the bootstrap filter presented in Doucet et al. (2001). There are three steps to this filter:

Initialization ($t = 0$)**:** A set of N particles and associated importance weights are initialized,

$$z_0^{(i)} \sim P(z_0) \qquad\qquad w_0^{(i)} = 1/N$$

for $i = (1, \ldots, N)$.

Importance Sampling: After each observation, a new set of particles is proposed based on the previous set of particles and the importance weights are computed,

$$\tilde{z}_t^{(i)} \sim P(z_t|z_{t-1}) \qquad\qquad w_t^{(i)} = w_{t-1}^{(i)}P(y_t|\tilde{z}_t^{(i)})$$

for $i = (1, \ldots, N)$.

Selection: A new set of particles is sampled with replacement from a distribution based on the normalized importance weights, and the weights are reset,

$$z_t^{(i)} \sim \sum_j w_t^{(j)}\delta(\tilde{z}_t^{(j)}) \qquad\qquad w_t^{(i)} = 1/N$$

for $i = (1, \ldots, N)$, where where $\delta(\tilde{z})$ is a distribution that puts all its mass on $\tilde{z}$.

At the final time T we have an approximation to the posterior $P(\mathbf{z}_{0:T}|\mathbf{y}_{1:T})$, corresponding to the discrete distribution obtained by assigning each particle its normalized weight.

Parameters of Particle Filters

We can introduce variation into this algorithm by exploring different methods of particle selection. Resampling with replacement after every observation quickly reduces the diversity in the set of particles, as a few highly weighted particles can take over the population. Thus, a common addition to the bootstrap filter is to vary how often resampling takes place, using some measure of the amount of variability seen in the weights of the particles. Resampling with replacement can also result in identical copies of particles. Markov chain Monte Carlo (MCMC) is often used in conjunction with resampling as a *rejuvenation* step to restore diversity into the set of particles (Chopin, 2002).

These choices about how to implement a particle filter have implications for the way that it behaves, but the consequences of manipulating these parameters on sensitivity to trial order have not been systematically explored. In the following section, we set up a particle filter for our Bayesian model of causal learning, and use it to investigate how different resampling methods affect our predictions. This investigation seems particularly interesting given the potential psychological interpretation of each of these parameters: resampling and rejuvenation require greater computation than simply continuing to update the weights on particles, and might thus be used strategically as a form of more deliberative reasoning that is triggered by some aspect of the state of the learner, or the task they are performing.

Particle Filter Parameters and Order Effects

With our Bayesian model of causal learning defined and an algorithm for a general purpose particle filter proposed, we now turn to exploring the effects of varying the parameters of the particle filter. We first modify the template given above to fit our problem:

Initialization ($t = 0$)**:** A set of N particles, where each particle $z_0^{(i)}$ holds a pair of strength estimates (s_0, s_1), and associated importance weights is initialized. s_0 and s_1 are drawn from the prior defined above, and the weights are set to be uniform, $w_0^{(i)} = 1/N$.

Importance Sampling: After each observation, a new set of particles is proposed from the Beta distribution defined above, and the importance weights are computed using Table 1.

Selection: This step is where the four models we analyze diverge. In Model 1, we never resample, simply letting the importance weights determine the strength estimates. In the other models, we resample particles based on a multinomial distribution defined on the importance weights. In Model 2, we resample at each trial t. In Models 3 and 4, we resample only if the variance of the weights is too large as defined by the Effective Sample Size (ESS). The ESS is $\approx \| \mathbf{w}_t \|^{-2}$, and we set a threshold at $0.10N$, ten percent of the number of particles. Model 4 has an extra step after resampling where we perform rejuvenation on the particles. We perform 10 iterations of Metropolis-Hastings with new values for (s_0, s_1) drawn from a from a Normal distribution centered on (s_0^{old}, s_1^{old}) with a standard deviation of 0.10 and accept the proposed (s_0^{new}, s_1^{new}) pairs following the Metropolis-Hastings acceptance rule.

Results on the Causal Learning Task

We applied all four models to a simulated version of the causal learning task of Dennis and Ahn (2001), using the same contingencies they listed for Experiment 3. For each of the four different resampling methods, we averaged performance over 500 runs, varied the number of particles from 1 to 1000, and set $\lambda = 10,000$. We presented the generative-preventative sequence first, and then re-initialized the particle filters and ran the preventative-generative sequence. The results are depicted in Figure 2.

Model 1 - Never Resample: In Figure 2 (a), we see that this model predicts a strength of 0 because the importance weights will average out over the trials. At first the particles with positive strengths will have higher weights but will drop when the negative trials begin. The opposite effect occurs for particles with negative strengths. At the end of the simulation the average of the weights goes to 0 since the overall contingency between C and E in the combined sequence is 0.

Model 2 - Always Resample: As shown in Figure 2 (b), this model exhibits a strong primacy effect, with strength estimates ending up at values consistent with the first block presented (positive for generative, negative for preventative) across a wide range of numbers of particles. This happens because particles with opposite strength to the current trial are replaced very early in the sequence. This destroys diversity in the particle set and only particles with strengths in common with the first few trials of the particular sequence remain.

Model 3 - ESS Resample: Figure 2 (c) shows that when we resample the particle set only once the variance in particle weights becomes high, we see primacy effects for smaller numbers of particles and then a convergence to 0 with larger numbers of particles. Since the ESS threshold is based on a percentage of the number of particles, smaller numbers of particles are more likely to lead to frequent resampling because it is less likely they will contain a good set of candidate strength values. Larger populations take a longer time to meet the ESS threshold, producing behavior that is more similar to Model 1. We get a better understanding of this model's behavior by focusing on the predictions of 50 particles at each trial. Figure 3 (a) shows that after resampling, the diversity of the particle set narrows. This results in a primacy bias, albeit a smaller effect than Model 2 due to infrequent resampling.

Model 4 - ESS Resample with Rejuvenation: The results in Figure 2 (d) show that this model rises and falls in its sensitivity to order as a function of the number of particles, as in Model 3. However, the Metropolis-Hastings rejuvenation step produces a wider range of particle strengths after resampling. This is illustrated in Figure 3 (b) where we focus on this model's behavior at each trial for a set of 50 particles. After resampling and rejuvenation, the diversity of particles is much broader than simply resampling; resulting in the predictions following the most recent trials.

Modeling Order Effects in Causal Learning

Now that we have established the effects of different parameters of the particle filter, we can consider what settings are required to reproduce the empirical data from the causal learning experiments discussed earlier in the paper. We show that both primacy and recency effects can be produced using the sequence of 80 trials presented in both Dennis and Ahn (2001) and Collins and Shanks (2002), if the particle filter parameters are selected to reflect the difference in the procedures used in these two experiments.

Dennis and Ahn (2001): Primacy Effect

Using the ESS resampling model (Model 3) with a stronger prior for positive s_1 strengths, we can observe that the performance of the particle filter with a small number of particles predicts similar primacy effects to those found in Dennis and Ahn (2001). Our use of a stronger prior was motivated by Schustack and Sternberg (1981), where it was found that people weight generative evidence higher than inhibitory evidence. Figure 4 (a) presents these results, where it is apparent that the generative-block first performance shows more of a primacy effect than than preventative-block first.

Collins and Shanks (2002): Recency Effects

We observed recency effects with our ESS resampling with rejuvenation model (Model 4), however, to maintain consistency in our modeling efforts we use the modified ESS resampling model defined above with the addition that every 10 trials we resample with rejuvenation. This resampling scheme is motivated by the procedure in the Collins and Shanks (2002) experiment, where participants were asked to estimate the strength of the relationship every 10 trials. Making such a judgment could trigger the kind of deliberative reasoning that resampling and rejuvenation reflect. Figure 4 (b) shows that the performance of this model more accurately predicts the recency effects with a small number of particles. We observe too that the recency effect occurs much sooner under this resampling scheme than in our original simulation with Model 4.

General Discussion and Conclusions

While Bayesian models have become quite popular as rational explanations of human inductive inference, a number of significant criticisms remain unresolved. In particular, it is not clear how these computational level analyses connect to

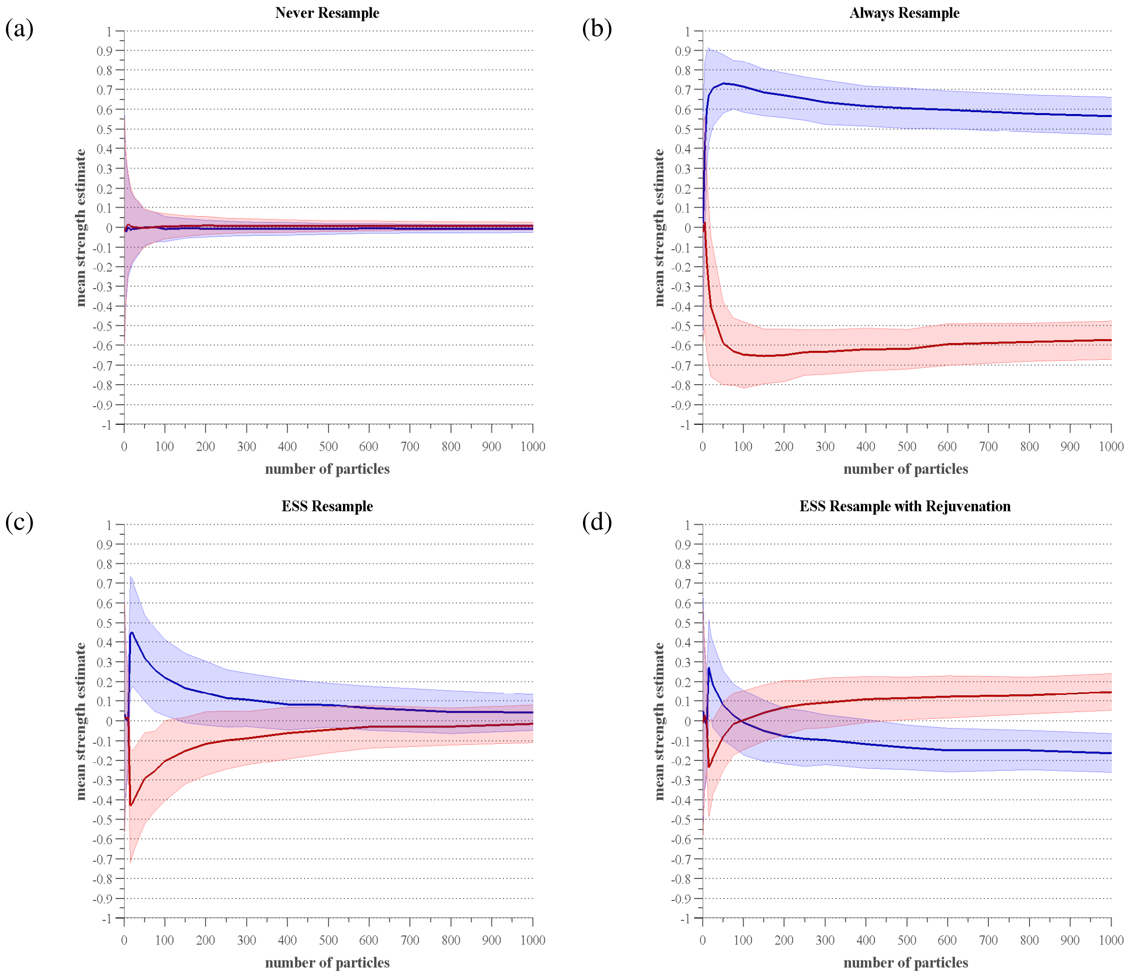

Figure 2: The performance of different resampling methods in our particle filter. The strength estimate produced for the generative-preventative (blue) and preventative-generative (red) versions of the contingencies is plotted against the number of particles used for each of four models: (a) never resample, (b) always resample, (c) resample only when the effective sample size (ESS) falls below a threshold, and (d) resample based on the ESS, and rejuvenate the particles (see text for details). The mean across 500 runs is shown with a heavy line, and the shaded fill indicates the standard deviation.

the algorithmic level of analysis that characterizes existing psychological process models. Using Monte Carlo methods to approximate Bayesian inference while also linking to models of psychological processes creates a new approach to the question of human inductive inference, resulting in what have been termed *rational process models* (Shi et al., 2008).

We have demonstrated how different resampling methods in a particle filter can produce different order effects in a causal learning task, potentially expanding the scope of the effects that can be explained using these models. Using a model with a bias for generative relationships and a sampling scheme that resamples only after the variance in particle weights becomes too high resulted in primacy effects similar to the results in Dennis and Ahn (2001) for small numbers of particles. Adding a rejuvenation step to this model after every 10 trials to match the experimental procedures of Collins and Shanks (2002) gives way to the observed recency effects in the literature.

These particle filter approximations to a Bayesian model of causal learning provide a more consistent explanation of the observed order effects in behavioral data. Our analysis is the first attempt at modeling order effects in causal learning using particle filters. We aim to explore other causal learning and contingency learning data with particle filter approximations in the future. This will include work to model not just how people estimate causal strength, but how they infer causal structure. We are currently conducting laboratory experiments to further explore how particle filter performance degrades with fewer particles in comparison to human performance degrading with higher cognitive load.

Acknowledgments. This work was supported by grant number FA-9550-10-1-0232 from the Air Force Office of Scientific Research.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

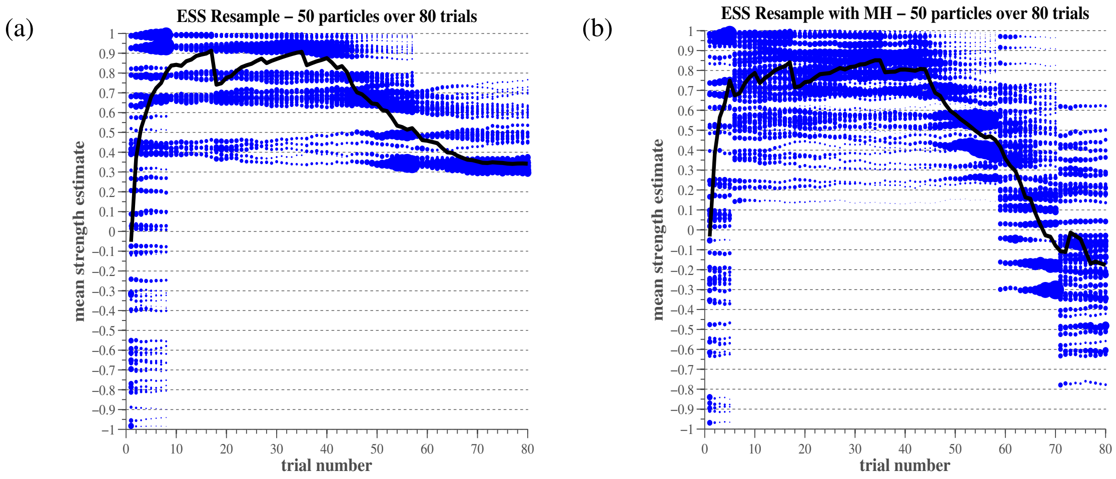

Figure 3: The performance of our particle filter over time for (a) Model 3, and (b) Model 4. The strength estimates produced by 50 particles for the generative-preventative data are plotted over trial number. The particles are represented as blue dots with the size of each dot a non-linear transformation of a particle's weight (for presentation purposes). The mean strength estimate over the set of particles for each trial event is given as a black curve.

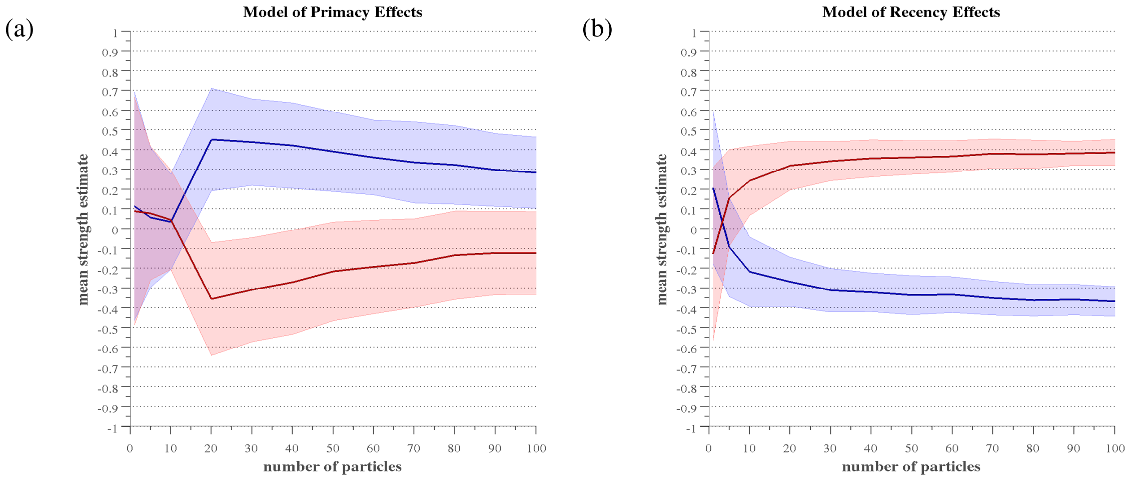

Figure 4: (a): The predictions of the modified ESS resampling particle filter on the contingency data from Dennis and Ahn (2001), and (b): the modified ESS resampling with rejuvenation particle filter on the task of Collins and Shanks (2002). The strength estimates produced for the generative-preventative (blue curves) and preventative-generative (red curves) versions of each task are plotted against the number of particles used, together with a shaded fill showing the standard deviation.

Cheng, P. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367-405.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, *89*(3), 539-551.

Collins, D., & Shanks, D. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, *30*(7), 1138.

Daw, N., & Courville, A. (2008). The pigeon as particle filter. *Advances in neural information processing systems*, *20*, 369–376.

Dennis, M., & Ahn, W. (2001). Primacy in causal strength judgments: The effect of initial evidence for generative versus inhibitory relationships. *Memory & Cognition*, *29*(1), 152.

Doucet, A., Freitas, N. de, & Gordan, N. (Eds.). (2001). *Sequential monte carlo methods in practice*. Springer.

Griffiths, T., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Science*, *14*, 357.

Griffiths, T., & Tenenbaum, J. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384.

Levy, R., Reali, F., & Griffiths, T. (2009). Modeling the effects of memory on human online sentence processing with particle filters. *Advances in neural information processing systems*, *21*, 937–944.

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

Oaksford, M., & Chater, N. (Eds.). (1998). *Rational models of cognition*. Oxford: Oxford University Press.

Sanborn, A., Griffiths, T., & Navarro, D. (2006). A more rational model of categorization. In *Proceedings of the 28th annual conference of the cognitive science society* (pp. 726–731).

Schustack, M., & Sternberg, R. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, *110*(1), 101.

Shi, L., Feldman, N., & Griffiths, T. (2008). Performing Bayesian inference with exemplar models. In *Proceedings of the 30th annual conference of the cognitive science society* (pp. 745–750).