The Causal Sampler: A Sampling Approach to Causal Representation, Reasoning and Learning

Zachary J. Davis (zach.davis@nyu.edu)
Bob Rehder (bob.rehder@nyu.edu)
Department of Psychology, New York University
6 Washington Place, New York, NY, USA 10003 USA

Abstract

Although the causal graphical model framework has achieved success accounting for numerous causal-based judgments, a key property of these models, the Markov condition, is consistently violated (Rehder, 2014; Rehder & Davis, 2016). A new process model—the *causal sampler*—accounts for these effects in a psychologically plausible manner by assuming that people construct their causal representations using the Metropolis-Hastings sampling algorithm constrained to only a small number of samples (e.g., < 20). Because it assumes that Markov violations are built into people's causal representations, the causal sampler accounts for the fact that those violations manifest themselves in multiple tasks (both causal reasoning and learning). This prediction was corroborated by a new experiment that directly measured people's causal representations.

Keywords: causal learning, causal reasoning, sampling

Introduction

The representation and use of causal knowledge is a central object of investigation in the cognitive sciences. Causal models have been found to affect cognition in a wide variety of inference problems, from reasoning and learning to decision-making and categorization (for a summary, see Rottman & Hastie, 2014; Waldmann & Hagmayer, 2013). One formal model of the representation of causal information — causal graphical models — has achieved success in modeling behavior across these tasks.

A foundational feature of causal graphical models is the Markov condition, which stipulates that the value of a node is independent of its non-descendants, conditional on its parents. This principle is crucial for statistical inference from causal graphical models (Pearl, 1988; Koller & Friedman, 2009), and has been argued to be necessary for a rigorous account of interventions (Hausman & Woodward, 1999).

Given the success of the causal graphical model formalism, one might expect to find the Markov condition satisfied in human behavior. In contrast, the causal inferences that people draw consistently violate the independence relationships implied by the Markov condition (Rehder, 2014; Rehder & Burnett, 2005; Rehder & Waldmann, 2016).

One explanation for Markov violations is that they represent a flaw in people's causal reasoning process. On this account, Markov violations would not necessarily manifest themselves on other causal-based tasks (e.g., causal learning). Rehder and Davis (2016) investigated this possibility by testing whether people honor the Markov condition during a causal hypothesis testing task. In fact, the hypothesis they favored reflected the same independence violations that characterize causal reasoning (details below).

Together, these findings pose a problem for current theories of causal cognition. We propose that the generality of these errors suggests that a reorientation is needed in our understanding of how people represent causal relationships. To this end, we propose a process model that conceives of causal cognition as based on simulation, rather than analytic calculation. The model outperforms traditional Bayes nets across tasks, and we test its predictions in a novel task.

Process Model

Building on recent work in cognitive science that investigates the role of sampling methods in accounting for judgments in a variety of domains (Hertwig & Pleskac, 2010; Lieder, Griffiths, & Goodman, 2012; Vul et al., 2014), we propose a model for resource-constrained inference using causal models. In particular, we propose that, when reasoning about causal systems, people attend to concrete cases and shift attention between those cases systematically. This process yields a joint distribution as a representation of the causal system, which can be used for inference in any task that can be modeled with causal graphical models.

Formalization

The proposed model is a variant of Metropolis-Hastings (MH) Markov Chain Monte Carlo, a computationally efficient rejection sampling method (Hastings, 1970). MH is defined by two components: a proposal distribution $\mathbb{Q}(q'|q)$ and a transition probability $a(q'|q)$, where q is the current state and q' is the proposal state in the random walk. Whereas MH models often deal with a continuous state space, the proposed model samples over the discrete states of a causal model. Figure 1B presents the eight states for the three variable graph shown in Figure 1A.

The sampling process uses the standard MH transition probability:

$$a(q'|q) = min\Big(1, \frac{\pi(q')}{\pi(q)}\Big)$$

where $\pi(q)$ is the joint probability of the graph being in state q given the graph's parameters (see the Appendix for an example of how $\pi(q)$ is calculated). The parameters reflect the particular beliefs of the participant (e.g. the causal strength between cause and effect).

We assume a proposal distribution $\mathbb{Q}(q'|q)$ that restricts reachable states q' to those that differ from the current state q by one binary variable. Each reachable state has an equal probability of being selected. Edges in Figure 1B denote reachable states for a node.

Note that this proposal distribution confers additional efficiency benefits. Because only one variable is changed, the ratio $\frac{\pi(q')}{\pi(q)}$ simplifies to

$$\frac{\pi(v_i', v_{-i})}{\pi(v_i, v_{-i})} = \frac{\pi(v_i'|v_{-i})\pi(v_{-i})}{\pi(v_i|v_{-i})\pi(v_{-i})} = \frac{\pi(v_i'|v_{-i})}{\pi(v_i|v_{-i})}$$

where v_i is the value of node i in q, and v_i' is the value in q'. This reduces the problem to calculating the relative conditional probabilities of two states, rather than representing the entire joint distribution. That calculating conditional probabilities only requires consideration of the node's *Markov blanket* further aids efficiency (Koller & Friedman, 2009).

The model thus far is simply an efficient MH model for estimating a causal graph's joint distribution. Importantly, however, we introduce a bias in the starting point for sampling: It always starts sampling from 'prototype' states, those in which nodes are either all 0 or all 1 (bottom left and top right corners of Figure 1B). This assumption is inspired by Johnson-Laird's influential Mental Models theory, in which the most easily represented state is the one where antecedent and consequent are both true (Johnson-Laird & Byrne, 2002). We propose that prototype states are the most easily represented states of a causal graph.

Regardless of our proposal distribution and biased initial samples, with many samples (e.g., 10^6), the causal sampler will converge to the normative distribution. However, we assume that people are resource-constrained and thus can only take a few samples (on the order of less than twenty). In this range, an MH model will overestimate the probability of states near the starting point (as it did not have time to fully explore the state space) and underestimate the remaining states. This effect is shown in Figure 2.

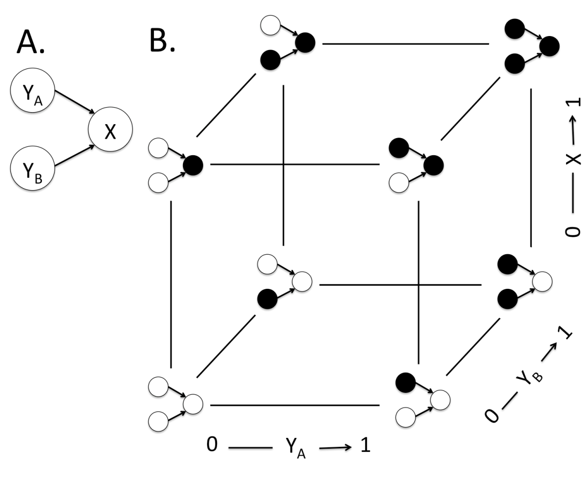

Figure 1: (A) Common effect network. (B) Possible concrete states of a common effect network. Filled in circles indicate a value of 1, empty circles indicate a value of 0.

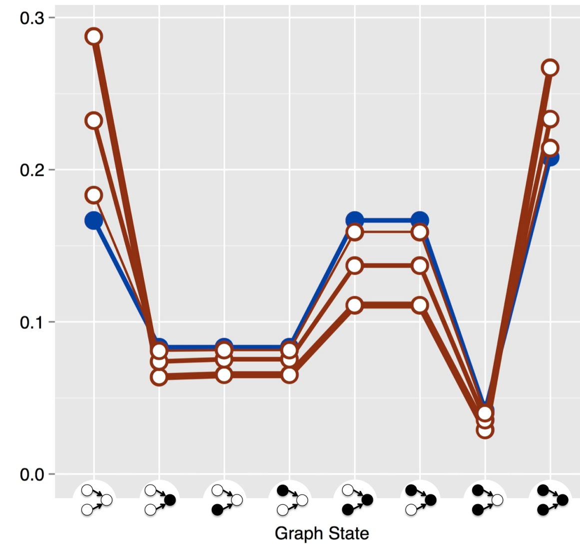

Figure 2: Joint distributions with causal strength = .5, causal prevalence = .5, strength of background causes = .33. The blue line (solid points) represents the joint distribution entailed by the normative model. Red lines (open points) represent the joint distributions simulated by the causal sampler, with thicker lines meaning fewer samples (thick = 4 samples, medium = 8, thin = 32).

An important prediction of the causal sampler model is that Markov violations are not resultant from a particular reasoning or learning process. Instead, these violations are built into the representations of causal graphs and so will propagate to any task that used the representation. To test this prediction, we compared our model to standard Bayes nets on existing data sets in causal learning and reasoning, as well as a new task introduced at the end of this paper.

Task 1: Causal Reasoning

The causal sampler model accounts for the independence violations found in human causal reasoning. For example, Rehder and Waldmann (2016) assessed the inferences people draw with the simple common effect graph in Figure 1A. Subjects were first instructed on two causal relationships that formed a common effect graph in the domains of economics, sociology, or meteorology (see the new experiment below for examples of these materials). The causal relationships were described as generative (a cause makes the effect more likely) and independent (each cause can bring about the effect on its own). Subjects were then asked to draw a number of causal inferences. For example, they were asked to estimate both $p(Y_A = 1|Y_B = 1)$ and $p(Y_A = 1|Y_B = 0)$ and also the same questions with the role of Y_A and Y_B reversed; thus, these inferences will be referred to as $p(Y_i = 1|Y_j = 1)$ and $p(Y_i = 1|Y_j = 0)$. The Markov condition stipulates that the two Ys should be conditionally independent, that is, that $p(Y_i = 1|Y_j = 1)$ should equal $p(Y_i = 1|Y_j = 0)$. The empirical results shown in the left hand side of Figure 3 (gray bars) reveal that subjects judged that $p(Y_i = 1|Y_j = 1) > p(Y_i = 1|Y_j = 0)$ instead. This violation of independence is also illustrated by the normative fit of the common effect graphical model in Figure 1A (blue solid line) to the ratings of Rehder and Waldmann's subjects (which included conditional probability queries other than those shown in Figure 3)[1]. As expected, the normative model is constrained to predict that $p(Y_i = 1|Y_j = 1) = p(Y_i = 1|Y_j = 0)$. This apparent expectation that the causes of a common effect graph are positively correlated has been observed in other studies (e.g. Rehder & Burnett, 2005; Rehder, 2014; Rottman & Hastie, 2016) and violations of the Markov condition have been observed with other graph topologies (see Hagmayer, 2016, and Rottman & Hastie, 2014, for reviews).

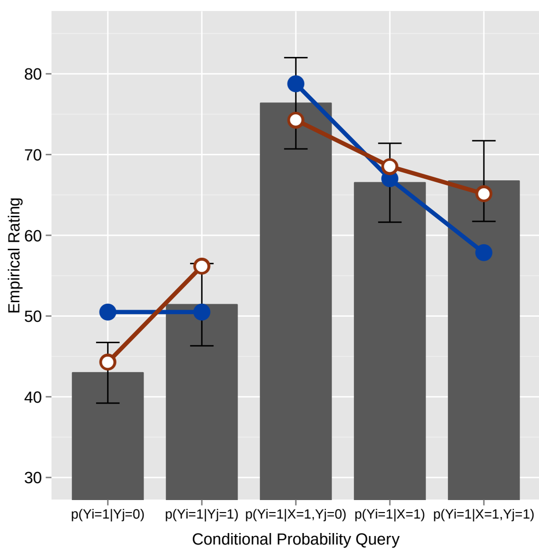

Figure 3: Data from Rehder & Waldmann (2014), Experiment 1. Sampler (red lines) and normative (blue lines, solid points) fits to conditional probability judgments. Error bars denote 95% confidence intervals.

Figure 3 also presents the best fit of the causal sampler to these data (red solid line) and shows that an average of 17.9 samples in fact reproduces subjects' belief that $p(Y_i = 1|Y_j = 1) > p(Y_i = 1|Y_j = 0)$[2]. It does so because the two prototype states are such that the causes are congruent $(Y_i = 1 \ \& \ Y_j = 1)$ or $(Y_i = 0 \ \& \ Y_j = 0)$. As was shown in Figure 2, the causal sampler overestimates these states, resulting in an inflated probability for states where the causes are congruent (e.g. $p(Y_i = 1|Y_j = 1)$), and underestimates states where the causes are incongruent (e.g. $p(Y_i = 1|Y_j = 0)$).

Note that the sampler also accounts for another reasoning error that subjects commit with the graph in Figure 1A. *Explaining away* is a signature property of common effect graphs. If X is observed to occur then the probability that Y_A is present of course increases. But if it is then further observed that the second cause Y_B is present then the probability that Y_A is present should *decrease*. (Conversely, if Y_B is observed to be absent then the probability of Y_A should increase.) In fact however, research finds that subjects often explain away too little or not at all (Morris & Larrick, 1995; Rehder, 2014; see Rottman & Hastie, 2014, for a review). The right three bars in Figure 3 illustrate the three conditional probability judgments relevant to explaining away: $p(Y_i = 1|X = 1, Y_j = 0)$, $p(Y_i = 1|X = 1)$, and $p(Y_i = 1|X = 1, Y_j = 1)$. The fits of the normative model to these data points reveal that explaining away with Rehder and Waldmanns subjects was indeed too weak (the slope of the blue line is steeper than the empirical ratings). In contrast, the fit of the sampler predicts this too-weak explaining away (the slope of the red line is shallower). Because it predicts both independence violations and weak explaining away, the sampler achieves a better fit to these data according to a measure (*AIC*) that corrects for its extra parameter (30.3 vs. 33.6).

Task 2: Causal Learning

The causal sampler also outperforms the normative model in a causal learning experiment. Rehder and Davis (2016) tested for the presence of independence violations in a hypothesis testing task by presenting subjects with a candidate theory that took the form of the graph in Figure 1A (again, in either the domain of economics, meteorology, or sociology). Subjects were then presented with hypothetical data and asked to rate the likelihood of observing the data if the theory was true. The correlation between Y_A and Y_B that obtained in the data was manipulated to be either negative, zero, or positive (all other aspects of the data, e.g., causal strengths, were held constant). The empirical results shown in Figure 4 (gray bars) revealed that subjects' ratings were largest when the inter-Y correlation was positive and smallest when it was negative.

The normative model's predictions for this task were derived by, for each of the three data sets, identifying the maximum likelihood parameters for the graph in Figure 1A to that data set. Using simple linear regression, the three maximum log-likelihoods were then scaled and translated onto the subjects' 0-100 ratings. The fitted predictions averaged over subjects (blue line in Figure 4) show the expected result that the data set with a zero inter-Y correlation is more likely than

[1] Fits were carried out per subject and identified parameters that minimized squared error. The parameters were w_Y (the marginal probabilities of Y_A and Y_B), w_{YX} (the strength, or causal power, of the links between the Ys and X), and w_X (the strength of alternative causes of X). Predicted conditional probabilities [0-1] were multiplied by a scaling parameter s to bring them into the range of subjects' ratings [0-100]. The best fitting parameters averaged over subjects were w_Y = .401, w_{YX} = .483, w_X = .178, and s = 158.6.

[2] Rather than explicit sampling, the causal samplers predictions for a given chain length has an analytic solution involving repeated multiplication of the matrix of transition probabilities between graph states defined by the Metropolis-Hastings rule. Fractional values of chain length involve computing the weighted average of the joint probability distributions that obtain when chain length is rounded up and down. The best fitting parameters averaged over subjects were w_Y = .440, w_{YX} = .469, w_X = .233, s = 137.7, and chain length = 17.9.

those with non-zero correlations, reflecting the independence between the causes stipulated by the normative model.[3]

The same process was followed for the causal sampler with the elaboration that we performed a grid search on the number of samples from 1 to 32. The fitted predictions (red line in Figure 4) reveal that the model, like the subjects, judged that the data set with the positive Y_A-Y_B correlation is most likely to be generated by the candidate theory (chain length averaged over subjects was 2.3). As in conditional probability judgments, it makes this prediction because biased sampling (starting at the prototypes) combined with a limited number of samples naturally generates the expectation that Y_A and Y_B will be positively correlated.

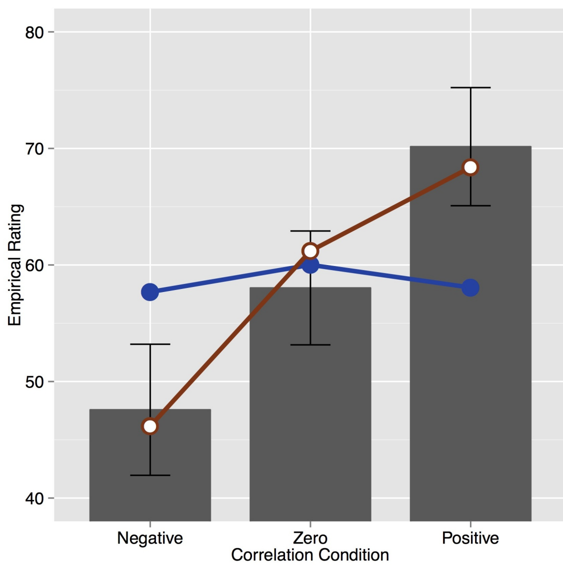

Figure 4: Rehder & Davis (2016). Sampler (red lines) and normative (blue lines, solid points) fits to data likelihood judgments. Error bars denote 95% confidence intervals.

Task 3: Expected Distributions

Recall that when the causal sampler's number of samples is limited, it warps a causal graph's joint distribution, overestimating prototype states and underestimating others (Figure 2). The following experiment tests this account using a novel methodology, one that directly asks participants to generate a distribution for a causal graph.

Method

Materials. Participants were presented with causal hypotheses in one of three domains: meteorology, sociology, or economics. Each domain had three variables (in economics: interest rates, trade deficits, and retirement savings; in meteorology: ozone levels, air pressure, and humidity; in sociology: urbanization, interest in religion, and socioeconomic mobility). Each variable could take on two possible values. One of these values was described as "Normal" and the other was either "High" or "Low". The values of the variables were mixed to prevent domain-specific beliefs from affecting the results (alternate values were either all "High", all "Low", or a mixture of "High" and "Low"). All hypotheses were of the form shown in Figure 1A, with two causes of one effect.

Procedure. Participants first studied screens of information that defined the variables, provided a mechanism describing how each cause could independently generate the effect, and a diagram of the causal relationships. They were then required to pass a multiple-choice test of this knowledge.

Next, participants were asked to generate a data set that they would expect to result from the causal graph. The causal relationship between smoking and lung cancer was used as an example. Subjects were shown the four cells formed by crossing smoker/non-smoker with lung cancer/no-lung cancer and how (in terms of how hypothetical people were allocated to the four cells) a greater proportion of smokers had lung cancer as compared to non-smokers. Subjects were asked to generate an analogous distribution in their assigned domain (economics, etc.). Specifically, they were given 50 pennies and asked to distribute them among the cells formed by crossing the three binary variables. They did so by placing the coins on a large sheet that contained the eight possible states (the position of the states on the sheet was randomized).

Design and Participants. The experiment consisted of a 3 (domain) by 4 (variable senses, e.g., all "High") between-subjects design. 60 New York University undergraduates received course credit for participation.

Results

Figure 5 presents how subjects allocated the 50 pennies to the eight states of the graph in Figure 1A (gray bars). Because these raw data are difficult to interpret, we computed measures that reflect the statistical relationships among the three binary variables implied by the pennies. In particular, we first normalized a subject's distribution and then computed the phi coefficients between a Y and an X ($\phi(Y_i,X)$; the pennies were aggregated so that the two Ys are interchangeable), between the Ys themselves ($\phi(Y_A,Y_B)$), and between the Ys conditioned on the presence of X ($\phi(Y_A,Y_B|X=1)$). These measures averaged over subjects are presented in Figure 6. First, the fact that $\phi(Y_i,X) >> 0$ indicates that subjects understood that the Ys were generative causes of X. Of greater theoretical importance is the fact that $\phi(Y_A,Y_B)$ was also significantly greater than 0, $t(59) = 3.62$, $p < .001$. This corroborates our claim that the violations of independence that obtain during causal reasoning (Figure 3) and hypothesis testing (Figure 4) are also manifested in peoples' causal representations (Figure 5).

The best fit of the normative model is shown superimposed

[3] Our lab has subsequently extended these finding to a more traditional hypothesis testing task in which subjects rate the posterior probability of the graph in Figure 1A relative to alternative hypotheses (those formed by removing one or both of the causal links).

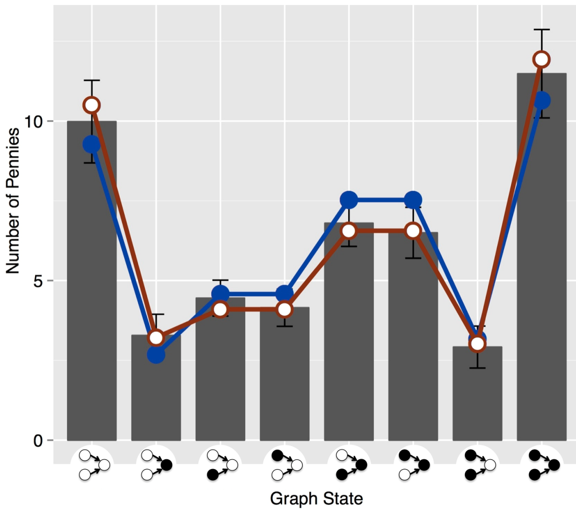

Figure 5: Causal sampler (red line) and normative (blue line, solid points) fits to participant-generated expected distribution judgments. Error bars denote 95% confidence intervals.

on the empirical data in Figure 5 (blue line)[4]. The figure indicates that the normative model tends to underpredict subjects' judgments for the two prototype states (111 and 000) and overpredict the remaining states. Phi coefficients computed for these fits (blue line in Figure 6) show the expected result that the normative model requires that $\phi(Y_A, Y_B) = 0$, at odds with subjects' distributions. Moreover, it sharply underpredicts $\phi(Y_A, Y_B | X = 1)$. Because of the explaining away phenomenon described above, the normative model requires that $\phi(Y_A, Y_B | X = 1)$ is negative (one cause is less likely when the other is present). Figure 6 shows that subjects' distributions implied a value of $\phi(Y_A, Y_B | X = 1)$ that is less negative (i.e., explaining away was again too weak).

The best fit of the causal sampler (red lines in Figs. 5 and 6) shows that it accounts for the fact that, relative to the normative model, the number of pennies is generally too large for the prototype states and too small for other states[5]. Of course, this pattern was expected on the basis of the theoretical predictions in Figure 2. Like the subjects, the causal sampler predicts that $\phi(Y_A, Y_B) > 0$ and that explaining away (as represented by $\phi(Y_A, Y_B | X = 1)$ is too weak relative to the normative model. As a result, it achieved a better fit to these data than the normative model (*AIC* of 3.2 vs. 10.8).

Discussion

Although causal graphical models have enjoyed success in explaining causal cognition, people consistently violate key predictions of these models. That independence violations manifest themselves in multiple tasks suggests that they arise from the causal representations that people construct. This conjecture was confirmed in an experiment using a new methodology that assessed, in a relatively direct way, people's causal representations. This result suggests that the fault lies not in how we reason or learn but how we represent.

This paper has proposed a process model that naturally constructs faulty causal representations. Importantly, it does so in a manner that is computationally efficient and psychologically plausible. The Metropolis-Hasting rule combined with the proposal distribution we advocate implies that at any one time reasoners only need to consider the relative likelihood of two graph states that differ by one variable, a computation that can be carried out very efficiently (because it involves only those nodes in the variable's Markov blanket; Koller & Friendman, 2009). Yet further efficiencies can be achieved for conditional probability queries (because sampling can be limited to those graph states that instantiate a query's antecedent). Note that this view suggests that humans *could* construct veridical causal representations—if only they had the cognitive resources to do so. The fault thus lies not in our causal representations per se but rather in the fact that causal judgments must be computed in finite time and with limited resources. Independence violations are thus an unavoidable consequence of the tradeoff between accuracy, speed, and effort.

The causal sampler perhaps gains some credence given the property it shares with the well-known Mental Model theory, namely, that reasoning is based on concrete states of the world (Goldvarg & Johnson-Laird, 2001; Johnson-Laird & Byrne, 2002). There are, however, some differences. Whereas the model theory never represents cause-present/effect-absent situations, the causal sampler, as a probabilistic model, merely asserts that such situations are un-

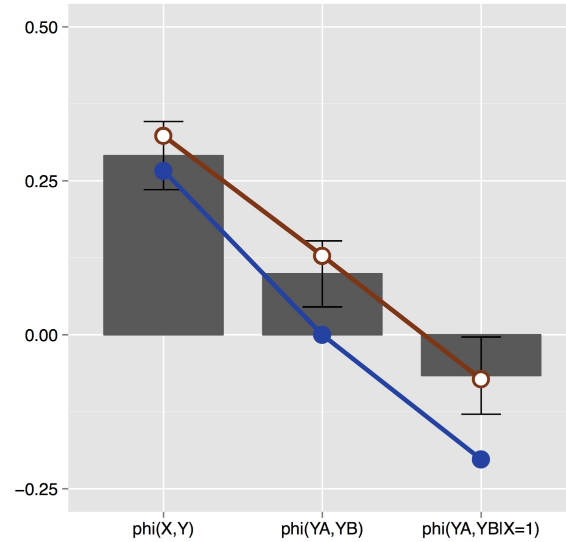

Figure 6: Causal sampler (red line) and normative (blue line, solid points) fits to participant-generated expected distribution judgments. Error bars denote 95% confidence intervals.

[4]The best fitting parameters (w_Y = .519, w_{YX} = 0.440, w_X = .243 averaged over subjects), were those that maximized the likelihood of the distribution of pennies.

[5]$w_Y = .534, w_{YX} = 0.410, w_X = .328$, chain length = 10.1.

likely (depending on the causal graph's parameters) and thus rarely sampled (cf. Khemlani, Barbey, & Johnson-Laird, 2014). There are also differences regarding which states reasoners initially consider (initial mental models are similar but not identical to the causal sampler's starting samples).

The causal sampler accounts for independence violations with other graph topologies. For example, suppose the direction of causality in Figure 1A is reversed, yielding a common cause graph. Independence is then captured by the *screening off* principle whereby the effects (Y_A and Y_B) are independent conditioned on the cause X. In fact, people judge that $p(Y_i = 1|X = 1, Y_j = 1) > p(Y_i = 1|X = 1, Y_j = 0)$ instead (Rehder, 2014; Rehder & Waldmann, 2016; Rehder & Burnett, 2005). The causal sampler predicts this result as well (because biased sampling induces a positive correlation between the Ys conditioned on X).

There are many possible directions for future research. For one, current models do not attempt to model the substantial variability in peoples causal inferences (Rehder, 2014; Rottman & Hastie, 2016). The stochastic nature of sampling may shed light on this important aspect of behavior. The causal sampler also makes predictions about reaction times. For example, it would predict that longer reaction times implies a less warped joint distribution (because more samples were taken).

Research in the causal graphical model tradition has rarely considered the cognitive processes involved in causal-based judgments. A limited sampling approach to building causal representations (a) is psychologically plausible, (b) accounts for the key discrepancy between graphical models and human judgments (Markov violations), and (c) explains why those discrepancies manifiest themselves in multiple causal-based tasks. Yet, it doesn't deny that people are sophisticated causal reasoners—they are, however, limited ones. As a process model, the causal sampler allows the causal graphical model framework to be extended to new phenomena, such as within- and between-subject variability and response times.

References

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science, 25*(4), 565-610.

Hausman, D. M., & Woodward, J. (1999). Independence, invariance and the causal Markov condition. *The British journal for the philosophy of science*, *50*(4), 521-583.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97-109.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115*

Johnson-Laird, P. N., & Byrne, R. M. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review, 109*(4), 646.

Khemlani, S. S., Barbey, A. K., & Johnson-Laird, P. N. (2014). Causal reasoning with mental models. *Frontiers in Human Neuroscience, 8*, 849.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Lieder, F., Griffiths, T. L., & Goodman, N. D. (2012). Burn-in, bias, and the rationality of anchoring. In P. PBartlett, et al. (Eds.), *Advances in neural information processing systems* (Vol. 25, pp. 2699-2707). Cambridge, MA: MIT Press.

Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review, 102*, 331-355.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology*, *72*, 54-107.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of object categories. *Cognitive Psychology, 50*, 264-314.

Rehder, B., & Davis, Z. (2016). Evaluating causal hypotheses: The curious case of correlated cues. In Papafragou, A., et al. (Eds.) (2016). *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Rehder, B., & Waldmann, M. R. (2016). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory and Cognition*, 1-16.

Rottman, B., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*.

Rottman, B. M., & Hastie, R. (2016). Do people reason rationally about causally related events? Markov violations, weak inferences, and failures of explaining away. *Cognitive Psychology, 87*, 88-134.

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*, 599-637.

Waldmann, M. R., & Hagmayer, Y. (2013). Causal reasoning. In D. Reisberg (Ed.), *Oxford Handbook of Cognitive Psychology* (pp. 733-752). New York: Oxford University Press.

Appendix

To calculate $\pi(q)$ (the probability of being in some state q), we simply use the normative calculation for each potential state. For example, when causal relations are generative, operate independently, and combine according to a noisy-or integration rule, $\pi(q)$ is defined as:

$$1-(1-b_j)\prod_{q_i \in Pa_k(q_j)}(1-m_{ij})^{ind(q_i)}$$

where b_j is the strength of causes exogenous to the model on the node, $Pa_k(q_j)$ denotes the parents of q_j in the causal model, m_{ij} denotes the causal strength between node j and parent i, and $ind(q_i)$ is an indicator function that yields 1 if feature q_i is present, 0 otherwise.