Uncertainty can explain apparent mistakes in causal reasoning

Nicolás Marchant[1] (nicolas.marchant@edu.uai.cl)**, Tadeg Quillien**[2] (tadeg.quillien@gmail.com) **& Sergio E. Chaigneau**[1] (sergio.chaigneau@uai.cl)
[1] Center of Cognitive and Social Neuroscience, Universidad Adolfo Ibáñez, Santiago, Chile
[2] School of Informatics, University of Edinburgh, UK

Abstract

Humans excel at causal reasoning, yet at the same time consistently fail to respect its basic axioms. They seemingly fail to recognize, for instance, that only the direct causes of an event can affect its probability (the Markov condition). How can one explain this paradox? Here we argue that standard normative analyses of causal reasoning mostly apply to the idealized case where the reasoner has perfect confidence in her knowledge of the underlying causal model. Given uncertainty about the correct representation of a causal system, it is not always rational for a reasoner to respect the Markov condition and other 'normative' principles. To test whether uncertainty can account for the apparent fallibility of human judgments, we formulate a simple computational model of a rational-but-uncertain causal reasoner. In a re-analysis of a recent causal reasoning study, the model fits the data significantly better than its standard normative counterpart.

Keywords: Causal reasoning; Markov violations, Inference judgments, Computational modeling

Introduction

The last decades have seen an explosion of research on causal cognition, guided by the advent of modeling tools such as Causal Bayes Nets (Pearl, 2000). Bayes nets are a formalism for normative causal reasoning, but researchers also hold that they might provide a good general hypothesis for how people represent and reason about causal systems (Glymour, 2003; Gopnik & Wellman, 2012; Hagmayer, 2016; Holyoak & Cheng, 2011; Rips, 2008; Rottman & Hastie, 2014; Sloman & Lagnado, 2015; Quillien & Lucas, 2023). Indeed, many studies suggest that causality is central to human cognition, and that people reason in a way that is well-approximated by algorithms for inference on causal Bayes nets (Waldmann, Holyoak, & Frantianne, 1995; Gopnik & Wellman, 2012; Griffiths & Tenenbaum, 2005; Rottman & Hastie, 2014; Marchant, Quillien, & Chaigneau, 2023). Against this background, it is surprising that one of the most replicable findings in the field is that people also consistently flout basic axioms of causal reasoning.

One of the most important characteristics of causal Bayes nets is the Markov condition. This condition stipulates that the state of any given variable in a causal model is independent of its non-descendants, conditional on the state of its direct parents (Pearl, 2000). For illustration, consider a chain causal model where X→Y→Z, and where we want to infer the state of variable Z. The Markov condition states that if the state of Y is known, then the state of X should be irrelevant for estimating the likelihood of Z (it is also said that Y "screens off" X). The Markov condition is a necessary assumption of any causally sufficient system (Pearl, 2000; Glymour, 2003). If humans are Bayesian reasoners, they should respect this condition. However, it has generally been reported that people violate the Markov condition when making inference judgments (Rehder & Burnett, 2005; Park & Sloman, 2013; Rottman & Hastie, 2016). This violation of a basic axiom of Bayes nets is generally considered an important shortcoming of the theory as a description of human reasoning (Rehder, 2014; Rehder & Waldmann, 2017; Sloman & Lagnado, 2015). This, and other apparently non-normative phenomena in causal-based reasoning, have led researchers to conclude that causal Bayes nets offer an incomplete account of human causal reasoning (Rottman & Hastie, 2016; Sloman & Lagnado, 2015; Rehder, 2018). In the current work, we highlight one way to reconcile people's judgments with the causal Bayes net framework. We argue that normative principles like the Markov condition hold for idealized reasoners that have perfect confidence in their model of the causal system – for a reasoner who is uncertain about the causal model, it might sometimes be rational to deviate from these principles.

The role of uncertainty in reasoning with causal models

In causal reasoning experiments, subjects are typically informed of a model characterized by causally related features, where each of the features assume discrete values (such as present/absent or high/low) and are sometimes characterized by base-rate and causal strength probabilities (i.e., p(effect|cause)). Subjects are then typically asked to infer the probability (e.g., using a slider scale ranging from 0 to 100) of a constituent feature occurring (or not; typically referred to as 1 for present and 0 for absent) given some known information about other constituent events. For example, researchers might first tell participants that z-radiation causes green spots, and specify the probability that someone gets green spots, in the presence and in the absence of z-radiation; then in a later phase, they ask participants (e.g.) the probability that someone has been exposed to z-radiation, given that they have green spots.

When researchers compare participants' judgments to the Bayes nets predictions, they implicitly assume that the participant is perfectly confident about the causal model that represents the system of interest. Formally, they assume that

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

the participant does not represent uncertainty about the structure or the parameters of the causal model. This is a strong assumption, ruling out for instance that the participant might hold both "74%" and "75%" to be plausible values for the probability with which the cause produces the effect.
Below we lay out some reasons why one might expect participants to maintain uncertainty about the causal model, and then explore the consequences of this assumption for normative principles of causal reasoning.

Sources of uncertainty

In causal reasoning experiments, researchers often describe the causal model verbally, as well as with a graphical model, but do not necessarily communicate explicitly the parameters of the model to the participant. In these cases, participants presumably have to infer a probability distribution over the causal model. In other cases, researchers also show participants several samples from the causal system (e.g. 30 people who were exposed / unexposed to z-radiation, and have / do not have green spots; see Rehder & Waldmann, 2017). In these cases, if participants learn the model by updating a prior distribution over possible models in the light of the evidence, we expect them to learn a posterior distribution over possible models. However, these latter procedures still allow some uncertainty (see, e.g., Griffiths & Tenenbaum, 2005; Lu, Yuille, Liljeholm, Cheng & Holyoak, 2008). In general, participants may also not completely discard the possibility that the experimenter might be deceitful, or mistaken about the correct causal model.

Hierarchical reasoning (Kemp, Perfors, & Tenenbaum, 2007) can also introduce uncertainty. Having wings causes birds to be able to fly, but this is not true of all species of birds: an ostrich cannot fly, regardless of whether it has wings. Participants might reason that the causal model given by the experimenter applies in some cases but not universally. For example, if the causal model states that low interest rates cause small trade deficits, participants might reason that this holds for countries in a particular region, and that countries in different regions might have different economic systems where no such relationship holds. Note that in various domains of cognition, people's priors appear to favor *sparsity* (Klayman & Ha, 1987; Oaksford & Chater, 1994; Hendrickson, Navarro, & Perfors, 2016; Navarro & Perfors, 2011). In particular, people's priors for causal reasoning favor a sparsity of causal relationships: by default, most variables are assumed to be causally unrelated (Lu et al., 2008). As such, even when an experimenter asserts that two variables are causally related, people might still assign a non-trivial probability to the contrary possibility.

Consequences of uncertainty

Here we provide some intuition for why uncertainty can change normative prescriptions in causal reasoning. We use the example of the Markov condition in a X → Y → Z causal chain. Formally, the Markov condition states that a variable is conditionally independent of its non-descendants, given its parents (Pearl, 2000). In the chain X → Y → Z, this means that Pr(Z|X, Y) = Pr(Z|Y): once we know the value of Y, knowing the value of X does not provide new information about the value of X.

Algebraically, the Markov condition follows straightforwardly from the factorization defined by the Bayesian network. The network topology X → Y → Z implies that we can write the joint probability distribution as:

$$\Pr(X,Y,Z) = \Pr(Z|Y)\Pr(Y|X)\Pr(X) \quad (1)$$

This factorization allows us to write Pr(Z|Y, X) as:

$$\Pr(Z|Y,X) = \frac{\Pr(Z,Y,X)}{\Pr(Y,X)} = \frac{\Pr(Z|Y)\Pr(Y|X)\,PrX)}{\Pr(Y|X)\Pr(X)} = \Pr(Z|Y) \quad (2)$$

However, if the reasoner is uncertain about the correct causal model, and entertains several possible hypotheses H_i, …H_n, then (even if every H assumes X → Y → Z) we must write Pr(Z|Y,X) as:

$$\frac{\sum_{H_i}\Pr(Z|Y,H_i)\Pr(Y|X,H_i)\Pr(X|H_i)\Pr(H_i)}{\sum_{H_i}\Pr(Y|X,H_i)\Pr(X|H_i)\Pr(H_i)} \quad (3)$$

The summation operations prevent us from canceling the identical terms in the denominators and numerators, and thus we cannot guarantee algebraically that Pr(Z|Y) = Pr(Z|Y, X), i.e., that the Markov condition holds.

In what follows we show more systematically that a model of a rational-but-uncertain agent can account for empirically observed violations of the Markov condition and other normative violations in causal reasoning.

Modeling uncertainty

Here we define a very simple formal implementation of our hypothesis, that we will compare to human data. We assume that, when given a causal model to reason with, people also consider alternative causal models in their computations (see also Meder, Mayrhofer, & Waldmann, 2014). Specifically, people consider two competing hypotheses about the causal model. According to hypothesis H, the causal model representing the system of interest is the model given by the experimenter. According to the alternative hypothesis H*, there is actually no causal relationship between the variables in the system. The model in H* is otherwise similar to H in terms of the variables it contains. We then assume that participants compute the joint distribution over variable states by marginalizing over hypotheses H and H*:

$$\begin{aligned}\Pr(X=x,Y=y,Z=z) &= \Pr(X=x,Y=y,Z=z|H)\Pr(H) \\ &+ \Pr(X=x,Y=y,Z=z|H*)\Pr(\mathrm{H}*)\end{aligned} \quad (4)$$

Where Pr(H) is the prior of the received causal model and Pr(H*) = 1-Pr(H) is the prior for the alternative causal model. This joint distribution can then be used to compute the conditional probabilities of interest, e.g., Pr(Z|Y).

Our model, viewed like this, is extremely simple, and we only assume that the relevant marginalizations are made on

the joint distribution of likelihoods that results from combining H and H* as shown in Eq. (4). Note that there are many other ways one could formally implement our hypothesis. For instance, one could model a reasoner who represents a probability distribution over many possible causal structures, and many possible model parameterizations. The simplicity of the current model is not only a matter of expository convenience: having a single free parameter Pr(H) limits excessive flexibility and the potential for overfitting.

In what follows, we will use Eq. (4) to provide accounts of several purported normative violations: Markov in causal chains, Markov in common cause structures, explaining away in common effect structures, and conservativeness, by modeling data made publicly available by Kolvoort and colleagues (Kolvoort, Fisher, van Rooij, Schulz & van Maanen, 2022).

Testing the model against empirical data

We test the predictions of our uncertainty-augmented model against data from a recent study. Kolvoort et al. (2022) asked participants to make 27 types of inferences about each of 3 causal structures with binary variables: a common-cause, a common-effect, and a chain structure (see Figure 1). For each structure, participants first were given verbal and graphical descriptions of the causal model, and then learned the associated probabilities by viewing 32 representative samples generated from the model. In the test phase, they made probability judgments, rating the probability of the presence of one variable given the state of other variables.

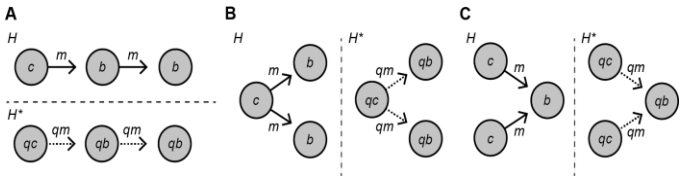

Figure 1: Causal structures and their parameters. Panel A shows a causal chain (i.e., X → Y → Z) model, Panel B a common cause (Y←X→Z) model, and Panel C a common effect (X→Z←Y) model. See Table 1 for the values of c, m and b used to generate model predictions. The *q* parameters for the alternative H* models are consistent with the sparsity assumption, with $qm = 0$, $qc = 0.5$, $qb = 0.5$, $p(\mathrm{H}) = 0.59$, and $p(\mathrm{H}^*) = 1 - p(\mathrm{H})$. See the main text for further details.

The causal models used by Kolvoort et al. are parameterized in the following way. If a variable has no parent in the graph, it has probability *c*. Otherwise the probability that variable E is present is given by a noisy-OR function:

$$\Pr(E|c_1 \dots c_n) = 1 - (1-b)\prod_{c_i}^{n}(1-m)^{c_i} \quad (5)$$

Where $c_1,\dots c_n$ denote the presence of the potential causes of *E* in the graph ($c_i = 1$ if present, 0 if absent), *m* denotes the strength of causal relationships, and *b* is the base-rate probability of *E* in the absence of any of its causes (see Figure 1). For instance, when there is a single cause present then $\Pr(E) = m+b-mb$.

The parameters used by Kolvoort et al. are shown in Table 1. We use these parameters in our modeling, effectively assuming that participants accurately incorporate the probabilities that they were taught for hypothesis H. As we mentioned above, we assume that people's belief about the causal model is a mixture of two different models (i.e., the given causal model H, and an alternative causal model H*). For modeling purposes, we assume that the H* model is a null model that represents the possibility that there is in fact no causal relationship between the three variables. In that regard, H* is parameterized with parameter values of $qc = .5$, $qm = 0$ and $qb = .5$; that is, all variables have base rate .5, and there is no causal relationship between variables. In other words, this model induces a joint probability distribution where each possible state of the system has probability 1/8. We set the base rate of variables in H* to .5 to reflect the fact that study stimuli in Kolvoort et al. (2022) were counterbalanced. For example, in one of the scenarios (economics), C_1=1 was operationalized as "high interest rates" for half of participants and "low interest rates" for the other half (while C_1=0 was always "normal interest rates").

Table 1: Model parameters used in Kolvoort et al. (2022).

	Common-cause, Chain	Common-effect
c	.5	.5
m	.67	.5
b	.25	0

Our model has a single free parameter Pr(H): the ground truth causal model H is assumed to be correct with probability Pr(H); otherwise H* is correct (with probability 1-Pr(H)). The joint probability distribution over possible variable states can then be computed by marginalizing over causal models H and H* by following Eq. (4). Conditional probabilities can then be derived from this joint distribution[1]:

$$\Pr(X = x|Y = y, Z = z) = \frac{\Pr(X = x, Y = y, Z = z)}{\Pr(Y = y, Z = z)} \quad (6)$$

[1] Note that it may be initially tempting to compute conditional probabilities by directly averaging the conditional probabilities defined by each causal model, i.e. compute $\Pr(X = x|Y = y, Z = z)$ as $\Pr(X = x|Y = y, Z = z|H)\Pr(H) + \Pr(X = x|Y = y, Z = z|H*)\Pr(H*)$. But this computation is not in fact conform to the laws of probability.

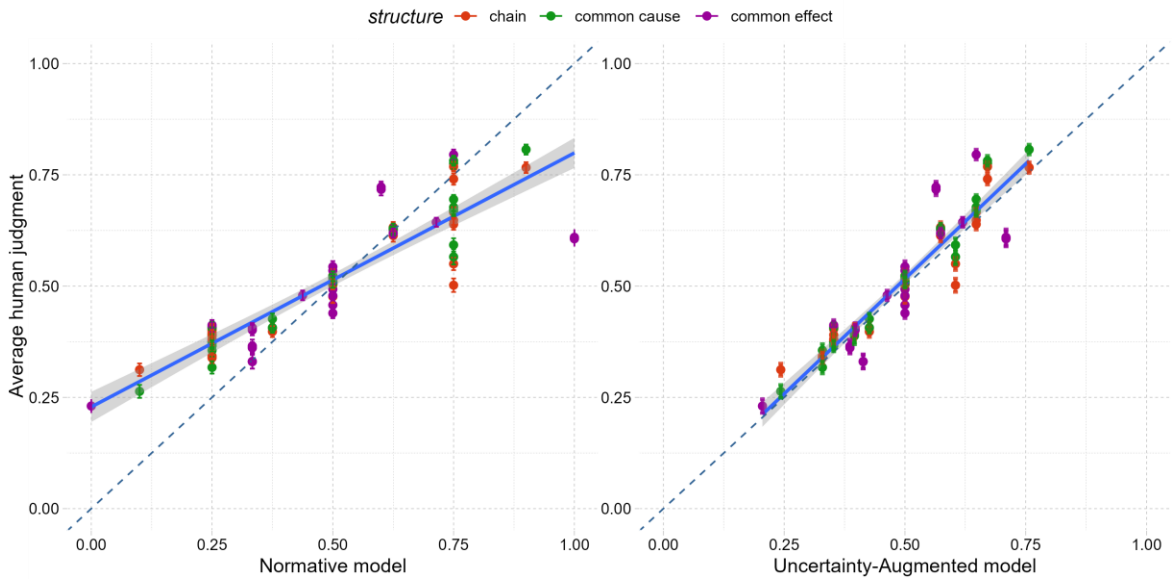

Figure 2: Average human probability judgments and model predictions. Each point represents one inference. Error bars represent the standard error of the mean. In all figures presented here, human data are from Kolvoort et al. (2022).

We fit the model at the group level, by finding the value of *Pr(H)* that minimizes the root mean squared error (RMSE) between model predictions and average human judgment (we find *Pr(H)* = .59). We also compute predictions for the normative model (which fully relies on the ground truth causal model H). R code to reproduce our analyses is available on the Open Science Framework.

Modeling results

The Uncertainty-Augmented model has a better fit (RMSE = .0521) to mean human judgments than the normative model (RMSE = .1147). Its predictions are also more highly correlated with mean human judgments, r(79) = .937, p < .001, than the normative model, r(79) = .900, p<.001; see Figure 2. This result also holds when accounting for the extra free parameter of the Uncertainty-Augmented model. We computed model fit using Leave-One-Out Cross-Validation, repeatedly training the model to predict human judgments on two causal structures (e.g., chain and common-cause), and then testing its fit on the third structure (e.g., common-effect). We obtain very similar results as when we fit the model directly (RMSE = .0528). The uncertainty-based model also has a significantly better fit to the human data than the normative model as assessed by their respective BICs[2]: Bayes Factor > 10^4.

These findings also hold at the individual level: 40 out of 43 participants were better fit by the Uncertainty-Augmented than the normative model (even when using the same value of *Pr(H)*, fitted at the group level, for all participants).

Conservatism On Figure 2, the line of best-fit for the normative model is shallower than the identity line. This shows conservatism: participants' judgments tend to be closer to 50% than is predicted by the normative model. By contrast, the line of best-fit for our model coincides almost perfectly with the identity line, indicating that it accurately accounts for conservatism in participants' responses. Conservatism has been shown in other studies of causal reasoning (Rottman & Hastie, 2016; see also Edwards, 1968). Here, conservatism falls out as a natural consequence of uncertainty about the causal model.

Markov violations in causal chains In a causal chain (X1 → Y → X2), Markov implies that the following conditional probabilities should all be equal: p(X2=1|Y=1, X1=1) = p(X2=1|Y=1) = p(X2=1|Y=1, X1=0). This pattern means that the X1 feature's effect on X2 is screened off by feature Y. However, the human pattern differs from the normative prediction in that the distal X1 cause does affect inferences about the state of the X2 effect: p(X2=1|Y=1, X1=1) > p(X2=1|Y=1) > p(X2=1|Y=1, X1=0).

Figure 3(A and B) shows that participants violated the Markov condition in the chain structure. Participants made different judgments (see the gray bars) for questions to which the normative model (red dots) gives the same answer. Our Uncertainty-Augmented model shows the same qualitative pattern of judgments as participants and as it has been reported in the literature (Park & Sloman, 2013; Rehder,

[2] BIC penalizes models with more free parameters. To compute log-likelihoods for the BIC, we assumed that responses for judgment i are generated from a normal distribution with mean μ_i and standard deviation σ, where μ_i is the model's prediction for judgment i. We fit σ to the data at the group level. We then computed the Bayes' Factor as $exp(-\left(BIC_{uncertainty\ model} - BIC_{normative\ model}\right)/2)$.

2014; Mayrhofer & Waldmann, 2015). The model is an especially good match to participants' judgments when the intermediate variable is absent (Y = 0). When the intermediate variable is present (Y = 1), the model still predicts a Markov violation, although of a smaller magnitude than participants'.

Markov violations in common-cause structures In a common cause model (X1←Y→X_2), Markov implies that the following conditional probabilities should all be equal: p(X1=1|Y=1, X2=1) = p(X1=1|Y=1) = p(X1=1|Y=1, X2=0). This pattern means that the state of feature X_i should depend only on the state of its direct cause. However, the human pattern differs from the normative prediction in that the second X2 effect does influence inferences about the state of the X1 effect independently from the state of the Y common effect: p(X1=1|Y=1, X2=1) > p(X1=1|Y=1) > p(X1=1|Y=1, X2=0).

Figure 3 (panel C) shows that participants violated the Markov condition in the common cause structure in the Kolvoort et al. (2022) data. Our model captures their qualitative pattern of judgments, which has also been found in the literature (Park & Sloman, 2013; Rehder, 2014; Mayrhofer & Waldmann, 2015; Rehder & Waldmann, 2017), although it does not perfectly capture the magnitude of the effect when the cause is present (i.e., Y = 1).

Insufficient Explaining Away in common effect structures In a common effect model X1→Y←X2, explaining away implies the following difference: p(X1=1|Y=1, X2=1) < p(X1=1|Y=1) < p(X1=1|Y=1, X2=0). This pattern means that knowing about the presence of a cause X2 should reduce the estimated probability for an alternative cause X1 being also present. However, though humans do discount the probability of the X1 cause if they know that the X2 cause was also present, they fail to discount as much as normatively predicted. Figure 3 (panel D) shows that participants did not conform to Explaining Away in the common-effect structure. In most studies, people tend to exhibit the Explaining Away effect, but to a lesser extent than is normatively prescribed (Rehder, 2014; Rottman & Hastie, 2016; Rehder & Waldmann, 2017).

In the current data, participants show an extreme version of the usual pattern: they not only fail to explain away as much as they should, but they also even show the reverse tendency, judging that the presence of a cause makes the other cause even more likely. Our model predicts the more commonly observed pattern, as it engages in explaining away but less so than the normative model, see Figure 3 (panel D).

Modeling discussion

Our simple model has a good overall fit to participants' judgments and is able to account qualitatively for four documented deviations from the normative model (Markov violations in chain and common-cause structures, conservatism and insufficient explaining away), with only one free parameter (i.e., Pr(H)). For simplicity, we assumed that participants accurately learned the ground truth causal

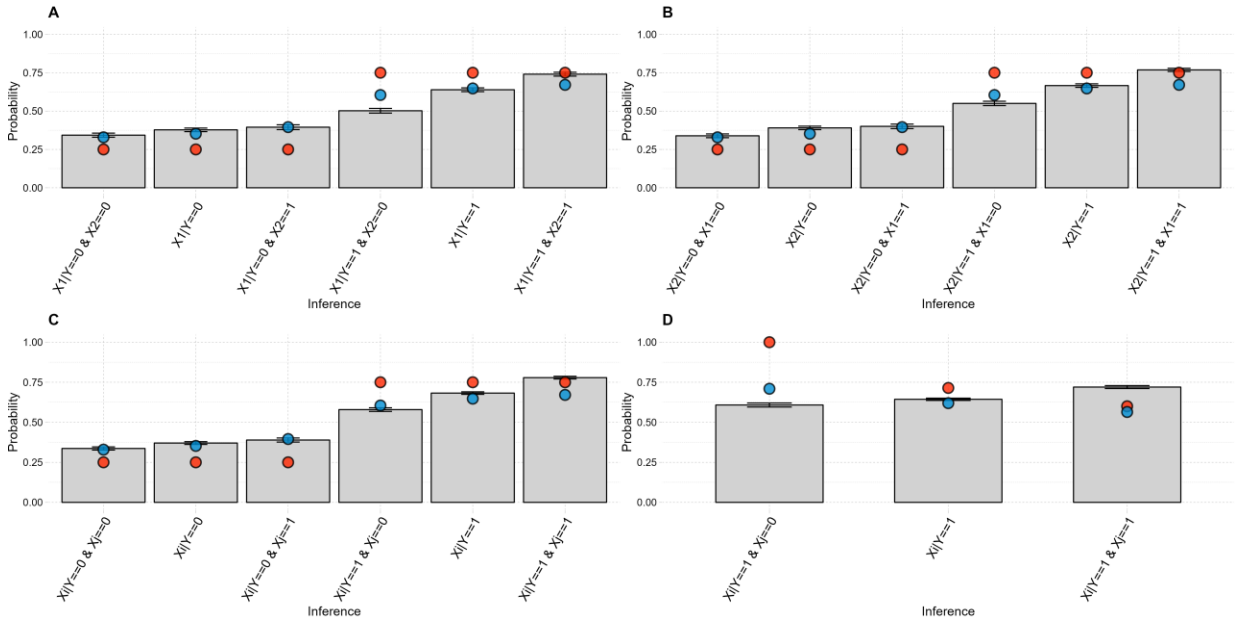

Figure 3: People's empirical ratings and model predictions on certain inferences. Average human judgments in Kolvoort et al. (2022) data (gray), along with normative (red dots) and uncertainty-augmented (blue dots) model predictions. In panels A and B, inferences relevant to assessing the Markov condition in the chain structure (X1 → Y → X2), when X1 is queried (A) or X2 is queried (B). Panel C shows relevant inferences for the Markov condition in common cause structure (X1←Y→X2). Panel D shows relevant inferences to assessing Explaining Away in common effect structure (X1→Y←X2). Error bars are standard errors of the mean.

model, but this might not have been perfectly the case. For instance, Markov violations were bigger when the intermediary variable Y was present than when it was absent, but these inferences should be symmetrical if participants inferred the ground truth causal model H correctly (since the marginal probability of each variable is 0.5). Exploratory analyses (not reported here) show that our model can reproduce this asymmetry in the size of the Markov violations by assuming participants learned a slightly different parameterization of the causal model.

The value of *Pr(H)* that gives the best account of our data is relatively low (*Pr(H)* = .59). Should we expect people to be that little confident in the causal model given by the experimenter? As outlined earlier, there are several potential sources of uncertainty about the applicability of the ground truth causal model. People might be uncertain about whether the causal model given by the experimenter holds in general; they could also be uncertain about whether the causal model applies in the particular case they are making a judgment about. To some extent, the best-fitting value of *Pr(H)* might also have been artificially pulled down by other sources of noise in the data (e.g. inattention or random responding).

General discussion

Our Uncertainty-Augmented model is able to predict several violations of Bayesian causal nets' prescriptions that are problematic for the literature. Whereas the standard normative model sees the reasoner as completely certain about the causal model describing the system of interest, our model incorporates uncertainty.

Our account predicts Markov violations to the extent that the values of the causal model parameters 'move together' across possible hypotheses that people have about the causal model. For example, in a causal chain X → Y → Z, people believe that either the X → Y and Y → Z relationships are both strong, or both relationships are weak. This assumption explains the Markov violation P(Z|Y,X) > p(Z|Y) because observing both Y and X provides evidence that the X→Y link is strong, which in turn provides evidence that the Y→Z link is also strong.

As such, our account predicts that the magnitude of Markov violations will track the extent to which people generalize evidence about the strength of a causal link to other causal links in the graph. We know of no direct evidence for this conjecture, but a series of studies by Park & Sloman (2013, 2014) offers indirect evidence. The authors manipulated whether the two causal links in a model operate via the same or different mechanisms. For example, in the "different mechanisms" condition, components A and B in a machine were connected by a blue line, while components B and C were connected by a red line; in the "same mechanism", both connections were of the same color. Markov violations were significantly higher in the "same mechanism" condition. Presumably, participants were more likely to generalize causal strength across relationships that rely on a similar mechanism.

We argue that the standard normative analysis neglects the important role of uncertainty in reasoning, and that therefore human causal reasoning might be more rational than it seems. Our model differs theoretically from many existing accounts of non-normative causal reasoning, which assume some degree of irrationality on the part of human reasoners (e.g. Rottman & Hastie, 2016; Rehder, 2018; Davis & Rehder, 2020). For example, systematic normative violations might in part be a byproduct of sampling-based approximation, in conjunction with limited cognitive resources (Davis & Rehder, 2020; Kolvoort, Temme & Van Maanen, 2023). Sampling-based models successfully predict an impressive number of features of human causal reasoning. On the other hand, some characteristics of Markov violations are not accounted for by these models. Under a sampling account, giving participants more time to think should allow them to take more samples, which should decrease the size of their Markov violations (Davis & Rehder, 2020). In contrast to this prediction, manipulations of time pressure do not affect the magnitude of Markov violations in causal reasoning tasks, although they affect overall accuracy (Kolvoort et al., 2022; Rehder, 2014). Also, sampling-based accounts do not directly account for the effect of mechanistic information (Park & Sloman, 2013; 2014; see discussion above).

Our analysis is closely related to theories that argue participants use a different representation of the causal model than the one explicitly provided by the experimenter (Buchanan, Tenenbaum & Sobel, 2010; Park & Sloman, 2013, 2014; Mayrhofer & Waldmann, 2015). However, these theories are typically designed to account for reasoning violations in one or two causal structures, while our account can explain normative violations in chains, common-cause and common-effect structures.

We note that it is likely that many different factors overall contribute to explaining why people fail to respect norms of good causal reasoning. Thus, we do not see our account as competing with, but rather complementing the other hypotheses mentioned above.

Our suggestion that uncertainty can play an important role in causal reasoning is consistent with previous research with a slightly different task. Meder, Mayrhofer and Waldmann (2014) found that when participants make conditional probability judgments on the basis of contingency data, these judgments do not depend only on the empirical conditional probabilities in the data: they are also independently modulated by the empirical correlation between the cause and effect variable. Participants' judgments were well-approximated by a Bayesian reasoner that is initially uncertain about whether C has a causal influence on E, and uses the empirical correlation between C and E to resolve this uncertainty.

Open science statement

R script for modeling the Uncertainty-Augmented model are open to download at the Open Science Foundation (OSF): https://osf.io/6xa7m/

References

Buchanan, D., Tenenbaum, J., & Sobel, D. (2010). Edge replacement and nonindependence in causation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (32).

Davis, Z. J., & Rehder, B. (2020). A process model of causal reasoning. Cognitive Science, 44(5), e12839.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17-52). New York: Wiley.

Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Sciences*, 7(1).

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, *51*(4), 334-384.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108. https://doi.org/10.1037/a0028044

Hagmayer, Y. (2016). Causal Bayes nets as psychological theories of causal reasoning: evidence from psychological research. *Synthese*, *193*(4), 1107–1126. https://doi:10.1007/s11229-015-0734-0

Hendrickson, A. T., Navarro, D. J., & Perfors, A. (2016). Sensitivity to hypothesis size during information search. *Decision*, *3*(1), 62–80. https://doi.org/10.1037/dec0000039

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual review of psychology*, *62*, 135-163. https://doi.org/10.1146/annurev.psych.121208.131634

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental science*, *10*(3), 307-321.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review*, *94*(2), 211-228.

Kolvoort, I., Fisher, E. L., van Rooij, R., Schulz, K., & van Maanen, L. (2022, August 3). Probabilistic causal reasoning under time pressure. *PsyArXiv*. https://doi.org/10.31234/osf.io/ej26r

Kolvoort, I., Temme, N., & van Maanen, L. (2023). The Bayesian Mutation Sampler explains distributions of causal judgments. *PsyArXiv*. https://psyarxiv.com/9kzb4/

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*(4), 955.

Marchant, N., Quillien, T., & Chaigneau, S. E. (2023). A context-dependent Bayesian account for causal-based categorization. *Cognitive Science*, *47*(1). https://doi.org/10.1111/cogs.13240

Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: Dispositional intuitions as a guide to causal structure. *Cognitive Science*, 39(1), 65–95. https://doi.org/10.1111/cogs.12132

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121*(3), 277–301. https://doi.org/10.1037/a0035944

Navarro, D. J., & Perfors, A. F. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychological review*, *118*(1), 120–134. https://doi.org/10.1037/a0021110

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608–631. https://doi.org/10.1037/0033-295X.101.4.608

Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the markov property in causal reasoning. *Cognitive Psychology*, *67*(4), 186–216. https://doi.org/10.1016/j.cogpsych.2013.09.002

Park, J., & Sloman, S. A. (2014). Causal explanation in the face of contradiction. *Memory and Cognition, 42*(5), 806-820. doi:10.3758/s13421-013-0389-3

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, UK: Cambridge University Press.

Quillien, T., & Lucas, C. G. (2023). Counterfactuals and the logic of causal selection. *Psychological Review*

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology*, *72*, 54–107. https://doi.org/10.1016/j.cogpsych.2014.02.002

Rehder, B. (2018). Beyond Markov: Accounting for independence violations in causal reasoning. *Cognitive Psychology*, *103*, 42–84. https://doi.org/10.1016/j.cogpsych.2018.01.003

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. Cognitive psychology, 50(3), 264–314. doi:10.1016/j.cogpsych.2004.09.002.

Rehder, B., & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory and Cognition*, *45*(2), 245–260. https://doi.org/10.3758/s13421-016-0662-3

Rips, L. J. (2008). Causal thinking. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundations* (pp. 597–631). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511814273.031

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological bulletin*, *140*(1), 109-139. https://doi.org/10.1037/a0031903

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. https://doi.org/10.1016/j.cogpsych.2016.05.002

Sloman, S. A., & Lagnado, D. (2015). Causality in Thought. Annual Review of Psychology, 66, 223-247. https://doi.org/10.1146/annurev-psych-010814-015135

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal Models and the Acquisition of Category Structure. Journal of Experimental Psychology: General, 124(2), 181–206. https://doi.org/10.1037/0096-3445.124.2.181.