Rationally uncertain: investigating deviations from Explaining Away and Screening Off in causal reasoning

Nicolás Marchant[1] (nicolas.marchant@edu.uai.cl), Guillermo Puebla[2] (pueblaramirezg@gmail.com), Tadeg Quillien[3] (tadeg.quillien@gmail.com), & Sergio E. Chaigneau[1] (sergio.chaigneau@uai.cl)

[1]Center of Cognitive and Social Neuroscience, Universidad Adolfo Ibáñez, Chile
[2]Instituto de Alta Investigación, Universidad de Tarapacá, Chile
[3]School of Informatics, University of Edinburgh, Scotland

Abstract

This work provides an alternative account for deviations in human causal reasoning from normative predictions based on Causal Bayesian Networks (CBNs). We highlight violations of the Markov condition (Screening Off) and insufficient Explaining Away. Different from other accounts, our model does not assume that people fail to honor normative predictions due to reliance on heuristics, hidden nodes and links or cognitive limitations. Instead, we propose that people are rationally uncertain about the received causal model they are asked to reason with. We fitted the model to published data from two experiments where people were asked to make probability estimates on inferences of interest within a causal model. We find that the model is able to i) reproduce deviations from normative predictions, and ii) predict changes in the magnitude of these deviations across contexts. We conclude that assuming that people, in order to be rational, will always fully believe in the information they receive about a causal model may be too strong an assumption.

Keywords: Causal reasoning; Markov condition; Explaining Away; Computational modeling

Introduction

People are generally good at using causal information to make inferences about the world (Sloman, 2005). They often rely on the causal structure of entities in the world to make diagnostic judgments (i.e., to diagnose a disease given a set of symptoms; Fernbach et al., 2010), to attribute responsibility and blame (Fenton et al., 2013; Lagnado et al., 2013; Quillien & Lucas, 2023), and to categorize different entities into common natural categories (Marchant et al., 2023a; Rehder, 2017a). A formalism that allows the mathematical specification of causal knowledge is the Causal Bayesian Network (CBN, Pearl, 2000). CBNs are prominent and widely used in recent research on causal reasoning because they allow a pictorial representation of events connected by nodes and arrows representing events and causal relations, respectively. In addition, CBNs formalize the causal relations through the joint distributions of the connected events, allowing mathematical inference of a queried event using probability calculus. For this reason, CBNs were posited as a normative model to account for human causal knowledge and reasoning (Glymour, 2003; Hagmayer, 2016; Rips, 2008).

The normative CBN formalism has been offered as an explanation for human causal reasoning, inference and categorization (Marchant et al., 2023a; Rehder, 2017a, 2017b; Rottman & Hastie, 2014). Problematically, however, though CBNs successfully capture people's general behavioral trends, people seem not to fully adhere to its basic axioms, as demonstrated in several studies (see Rottman & Hastie, 2016; Sloman & Lagnado, 2015). Here, we focus on the two most important deviations in causal inference judgments: violation of the Markov condition (also known as Screening Off) and insufficient Explaining Away.

The Markov condition holds that the state of any given event in the causal model is independent of its non-descendants, conditional on the state of its direct parents (Pearl, 2000). As an example, consider a common cause structure with three variables $X_1 \leftarrow Y \rightarrow X_2$ in which we want to infer the state of event X_2. The Markov condition holds that if the state of Y is known, then the state of X_1 should be irrelevant for estimating the likelihood of X_2. However, people often take the state of X_1 into account when estimating the likelihood of X_2, violating the Markov assumption and deviating from predicted normative responses (Park & Sloman, 2013; Rehder & Burnett, 2005; Rehder & Waldmann, 2017).

Second, people's empirical estimates often exhibit insufficient Explaining Away in common effect structures. The Explaining Away principle prescribes that the presence of one cause of an effect should reduce the likelihood that an alternative cause is also present (Rehder, 2014; Rehder & Waldmann, 2017; Rottman & Hastie, 2014). Suppose we have a common effect structure $X_1 \rightarrow Y \leftarrow X_2$, and we want to infer the state of X_1 under the assumption that the common effect Y is also present. The normative prediction is that the presence of X_2 should reduce the likelihood that X_1 is also present because it explains away the contribution of X_1. However, people often fail to properly discount the contribution of X_2, failing to respect the Explaining Away principle. There does not appear to be a unified and straightforward explanation of why people' estimates fail to honor the Markov condition and the Explaining Away principle. Indeed, several explanations have been offered over the course of the years:

1) Similarity judgments or heuristics: As noted in Rottman and Hastie (2016) and Rehder and Waldmann (2017), some people seem to judge the information provided by a causal model according to the combinations of presence and absence of its variables (i.e., similarity). The Beta-Q model (Rehder, 2018) assumes a rich-get-richer principle, which states that the strength of the causal inference is a function of the variables that are present minus the variables that are absent. This type of behavior is related to the "representativeness" heuris-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

tic (Tversky & Kahneman, 1973), assuming that the causal model with all events present is more "representative" of the prototypical abstraction of the causal model presented.

2) Hidden nodes and causal links: Another explanation for non-normative behavior in causal inference is that people may not base their judgments entirely on the given information, but instead posit additional variables (as nodes in CBN) and causal relations (Rottman & Hastie, 2016). Park and Sloman (2013) hypothesized that people may not respect the Markov condition when faced with a causal model that uses the same mechanism explanation (e.g., in a chain structure, all causal links operate using chemical reactions), but not when faced with a different mechanism (e.g., some links use a chemical and others a physical mechanism). They suggest that in the same-mechanism condition, people will represent a hidden general shared disabler as a form of hidden variable with causal links within the observable causes. This is consistent with the fact that people may bring some prior knowledge to the task itself (Mayrhofer & Waldmann, 2015).

3) Limited cognitive resources: Recently, Davis and Rehder (2020) developed the Mutation Sampler model, which assumes that a limited capacity cognitive system solves causal reasoning problems through a mental process akin to sampling. The Mutation Sampler assumes that people use a sampling procedure (through Markov Chain Monte Carlo approximations) from concrete states (e.g., the provided causal model) that are stored in memory. People start this sampling process from one of the causal model prototypes (i.e., all variables are present or all variables are absent) to generate a next state where only one variable is in a different state than the prototype. The Mutation Sampler predicts deviations from normative inferences because people are generally biased toward the initial prototype and because the number of samples is limited due to cognitive resource limitations (Kolvoort et al., 2023). Similarly, Wang and Sun (2020) suggested that deviations in causal inference occur because people generally do not revise or adjust their prior beliefs, and thus they showed minimal revisions in the causal network in dynamic situations (e.g., when there is a change in the causal network from the initial state to the end-state).

In this work we investigate an alternative theoretical and computational explanation for why people's judgments deviate from normative estimates when reasoning about causal events. We build on a model that takes into account the role of uncertainty in causal reasoning (Marchant et al., 2023b). In a nutshell, the Uncertainty-Augmented Model (UAM) assumes that people are generally uncertain about whether the received causal model is true (i.e., the causal model described by the experimenter). In this context, people also consider alternative models, which are then taken into account in their computations (see also Meder et al., 2014). For example, the experimenter might tell the participant that variable A causes variable B 75% of the time, but participants might still allow for the possibility that A's causal influence is stronger or weaker, or indeed that A may not have a causal influence on B at all. In contrast to proposals suggesting that people posit hidden nodes or causal links that enable or disable the variables in the given causal model (Park & Sloman, 2013), the UAM models uncertainty about the relationship between the variables that were explicitly mentioned by the experimenter.

In the current work, we apply our model to account for deviations from the Markov condition and from Explaining Away. In what follows, we fit the UAM to empirical data from Rehder and Waldmann (2017) and test it against its normative counterpart. The dataset in Rehder and Waldmann (2017) allows us to investigate two main questions. First, participants in these experiments exhibit robust normative violations: can the UAM reproduce these effects? Second, Rehder and Waldmann (2017) manipulated the information that was given to participants, and found that this manipulation had an impact on the magnitude of normative violations. This feature of the dataset allows us to ask the more stringent question of whether the model is able to capture variation in the magnitude of normative violations across contexts. In this way we are able to test the UAM more thoroughly than in previous work (Marchant et al., 2023b).

The uncertainty-augmented model

The UAM assumes that people are generally uncertain about the trustworthiness of a given causal model to reason with. In principle, there are many possible ways to model this uncertainty. For the sake of simplicity, we follow the simple implementation of Marchant et al. (2023b). We consider an experiment in which the experimenter gives information about a causal system to a reasoner: for example, variable X_1 causes variable Y, which in turn causes variable X_2. The reasoner considers two possible hypotheses about the causal system. Under the first hypothesis (which we call H), the experimenter is correct about the causal model; the second hypothesis (H') is a null model, according to which there is in fact no causal relationship between the variables: they are all statistically independent, with a base rate of 50% each (see also Meder et al. (2014) for related ideas).

Formally, people have a hypothesis H that represents the causal model as intended by the experimenter, but also have a hypothesis H' that represents an alternative non-causal model in which there are no causal relationships between its variables. Both hypotheses (H and H'), however, are similar in terms of the variables they contain.

To implement model predictions in the context of the studies from Rehder and Waldmann (2017), we assume that causal links are generative (i.e., a cause increases the likelihood of its effect) and independent. We operationalize generative causal links using a noisy-OR function (Cheng, 1997):

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

where the presence of a given variable E is determined by the presence of potential causes of E in the causal model

$c_1 \ldots c_n$ (where $c_i = 1$ if a parent cause is present, 0 otherwise), m denotes the causal strength and b is the base-rate probability of E in the absence of any of its causes. If a variable has no direct parents in the given causal model, it has probability c.

The novelty of the UAM is the incorporation of the H' alternative causal model which represents the probability that in fact there are no causal relations between the variables. In that regard, H' is parameterized with parameter values of $c = .5$, $b = .5$ and $m = 0$ (parameters c and b are set to .5 to reflect our assumption that they are as likely inside and outside of the alternative causal model if H' is true). In sum, the UAM assumes that people's representation is a mixture of two potential causal models, H (the received causal model) and H' (the alternative null causal model).

We assume that people compute conditional probabilities in a normative manner, given their uncertainty about the correct causal model. They do so by marginalizing over the two possible hypotheses about the causal model, H and H'. The marginalization can be carried out by applying the law of total probability for conditional probabilities:

$$\begin{aligned} Pr(X|Y,Z) &= \sum_{H_i} Pr(X|Y,Z,H_i)Pr(H_i|Y,Z) \\ &= Pr(X|Y,Z,H)Pr(H|Y,Z) + Pr(X|Y,Z,H')Pr(H'|Y,Z) \end{aligned} \tag{2}$$

This equation has an intuitive interpretation. The conditional probability $Pr(X|Y,Z)$ is a weighted average of the conditional probabilities under the different possible hypotheses about the causal model, and the $Pr(H_i|Y,Z)$ term specifies the weight given to hypothesis H_i. This weight depends on the value of the observations Y and Z, because observing Y and Z gives us some evidence about which causal model is most likely to be correct. For example, in a common cause structure $X_1 \leftarrow Y \rightarrow X_2$, observing both $Y = 1$ and $X_2 = 1$ gives some evidence in favour of hypothesis H (that variables in the network are causally related).

As such, even though each conditional probability term $Pr(X|Y,Z,H_i)$ in the equation is computed in a way that respects the Markov condition, the weighted average of these terms does not necessarily result in Markov-compliant inferences. Consider the inference $Pr(X_1|Y = 1,X_2 = 1)$, in a common cause structure $X_1 \leftarrow Y \rightarrow X_2$. Even though X_2 is 'screened-off' by Y, observing $X_2 = 1$ gives us some evidence in favor of hypothesis H, increasing the corresponding weight in the weighted average. In contrast observing $X_1 = 0$ would have given us evidence against H. As such, an uncertain but rational causal reasoner will infer a different value for $Pr(X_1|Y = 1,X_2 = 1)$ than for $Pr(X_1|Y = 1,X_2 = 0)$.

In the later sections, we fit the model to a rich empirical dataset of causal-based inferences collected by Rehder and Waldmann (2017). R code to reproduce our analyses is available at https://osf.io/.

Empirical testing on Rehder and Waldmann (2017) dataset

Rehder and Waldmann (2017) tested whether the amount of information given to participants about a causal model influences the magnitude of normative deviations in causal reasoning. In two experimental studies (a common effect structure in Exp. 1, and a common cause structure in Exp. 2) with a sample size of 144 undergraduate students in each, they implemented three experimental conditions (i.e., empirical only; description only and description + empirical, see below) across three domains (economics, sociology and meteorology) in which the amount of information regarding the causal model was manipulated. In the empirical-only condition, subjects were given a sample of data that provided some information about the joint probability distribution over the variables in the causal model, but did not receive information about the network structure[1]. In contrast, in the description-only condition subjects were only told to study the causal relations presented as sentences that described the structure of the causal network (i.e. which variables caused which) as well as a description of the causal mechanism. Finally, in the description + empirical conditions subjects were given both the description of the causal structure and a sample of data that contained information about the statistical correlations between variables.

After participants studied the causal model, they moved to the test phase, in which they were asked to make eight different types of inferences (see inferences a to h in Table 1). For each inference, participants had to respond using a slider labeled from 0 to 100% to indicate the probability that an unknown to-be-predicted variable was present in the provided causal model. The to-be-predicted variable was signaled as "???". There were arrows indicating causal relations between variables, except for the experience-only condition in which only the correlational information was provided (see Rehder & Waldmann, 2017). As it is usually common in these kinds of procedures, all variables within a causal model had binary values. For example, interest rates would be normal or low.

Because the causal domain (i.e., economics, sociology or meteorology) did not significantly affect the results, the authors collapsed across domains and compared the experimental conditions only (see Rehder, 2017a). In Exp. 1 (common-effect model) they found that the three different experimental conditions showed insufficient Explaining Away (inferences a, b and c), however, the deviation was greater for the description-only condition and lower for the empirical-only condition. In Exp. 2 (common cause), the greater deviation from the Markov assumption (inferences a, b and c) was obtained in the description-only condition, whereas the empirical-only condition showed a lower amount of devia-

[1] Specifically, participants observed 27 (in Exp.1) or 33 (in Exp.2) samples from the causal model (each sample is a triplet of realized variable values). In the common-effect structure (Exp.1), the causal model was parameterized with $c = .32$, $m = .83$, and $b = .12$. In the common-cause structure, it was parameterized with $c = .50$, $m = .67$, and $b = .20$.

Table 1: Inference of interest in Rehder and Waldmann (2017)

	Inference type
a	$Pr(X_1 = 1 \mid Y = 1, X_2 = 1)$
b	$Pr(X_1 = 1 \mid Y = 1)$
c	$Pr(X_1 = 1 \mid Y = 1, X_2 = 0)$
d	$Pr(X_1 = 1 \mid X_2 = 1)$
e	$Pr(X_1 = 1 \mid X_2 = 0)$
f	$Pr(Y = 1 \mid X_1 = 1, X_2 = 1)$
g	$Pr(Y = 1 \mid X_1 = 0, X_2 = 1)$
h	$Pr(Y = 1 \mid X_1 = 0, X_2 = 0)$

Note. Our nomenclature is symmetrical to both causal models, for common effect (Exp. 1) $X_1 \rightarrow Y \leftarrow X_2$ and for common cause (Exp. 2) $X_1 \leftarrow Y \rightarrow X_2$, so the inference types are equivalent to both experiments. Also, most of the inferences are symmetric (with the exception of H and H'). For instance, the inference (a) $Pr(X_1 = 1 \mid Y = 1, X_2 = 1)$ is symmetrical to its version of $Pr(X_2 = 1 \mid Y = 1, X_1 = 1)$.

tion. Also, for the common effect structure, violations of independence (inferences d and e in Exp. 1) of the causes were found. Independence refers to the assumption that one cause should be irrelevant when predicting the presence of the other cause. However, participants tend to violate the independence assumption in the common effect structure.

Overall, the authors suggest that the description-only and description+empirical conditions showed stronger deviations of normative predictions of CBNs because highlighting the causal structure intensifies the 'rich-get-richer' bias. In short, this bias assumes that people consider that one variable is more likely to be present because there are other variables within the causal model that are also present (Rehder, 2014). As we noted in the introduction, this explanation assumes that causal reasoning deviations occur because people use some sort of heuristic about the presence of variables.

We fitted our UAM to the Rehder and Waldmann (2017) dataset to offer a different explanation of deviations. Two versions of the UAM were fitted: A full-model with four free parameters (i.e., $Pr(H)$, c, m and b) and its normative counterpart. When fixing the prior parameter to $Pr(H) = 1$, the full UAM reduces to the normative CBN predictions. Model fittings were computed for each participant using the dfoptim [2] function in R (Varadhan, 2023) by applying the Nelder-Mead optimization algorithm with bounded parameter space (lower = 0, upper = 1). Then, we averaged the best-fitting parameters across participants for both experiments and for the three experimental conditions (see Table 2). We compared the two models by computing the Root Mean Squared Error (RMSE) between predictions and empirical data. Because our two models have a different number of free parameters (UAM = 4 and normative model = 3) we also calculated BIC[3] which penalizes by the number of free parameters.

For Exp. 1 (common effect) the UAM model achieves better fits for the three conditions (Empirical only: $RMSE = .0743$, $BIC = -10.58$; Description+Empirical: $RMSE = .0223$, $BIC = -29.85$; Description only: $RMSE = .0587$, $BIC = -14.35$) than the normative model (Empirical only: $RMSE = .1033$, $BIC = -7.38$; Description+Empirical: $RMSE = .0348$, $BIC = -24.81$; Description only: $RMSE = .0738$, $BIC = -12.77$). For Exp. 2 (common cause) the UAM again achieves better fits for all conditions (Empirical only: $RMSE = .0492$, $BIC = -17.17$; Description+Empirical: $RMSE = .0425$, $BIC = -19.50$; Description only: $RMSE = .0585$, $BIC = -14.41$) than the normative model (Empirical only: $RMSE = .0768$, $BIC = -12.12$; Description+Empirical: $RMSE = .0807$, $BIC = -11.34$; Description only: $RMSE = .0943$, $BIC = -8.84$).

Correlations between model's predictions and empirical responses across experimental conditions by each inference type, revealed that the UAM shows a better prediction of empirical data in Exp. 1 ($r = .96$, $p < .001$, 95% CI [.91 .98], see figure 1A) and also in Exp. 2 ($r = .97$, $p < .001$, 95% CI [.93 .99], see figure 1B). The normative model achieves a lower correlation coefficient in Exp. 1 ($r = .95$, $p < .001$, 95% CI [.88 .98]) and also in Exp. 2 ($r = .96$, $p < .001$, 95% CI [.91 .98]). As we will see next, what is critical of the UAM is that it entails the empirical deviations in causal reasoning while the normative model does not, suggesting that the incorporation of uncertainty may explain why people do not honor Markov assumptions and sufficient explaining away.

Accounting for the effects of the experimental manipulation

In Figure 2, we plot the empirical data and model predictions (for both models) broken down by experimental condition. Overall, the UAM is able to reproduce most of the empirical patterns found in the Rehder and Waldmann (2017) dataset. First, the UAM predicts the insufficient Explaining Away in the common effect structure (see inference types a, b and c; Figure 2, upper row). Second, the UAM is also able to track empirical responses for common cause structures: as shown in Figure 2 bottom row, it predicts Markov condition violations (inferences a, b, and c).

Interestingly, the UAM predictions can account for the effect of the experimental manipulation of Rehder and Waldmann (2017). As shown in Figure 2, overall model predictions (for both models) are lowest in the empirical-only con-

[2] dfoptim is a R package which computes derivative-free algorithms, which is convenient for non smooth functions.

[3] BIC were computed using the formula $BIC = -2 * LogLikelihood + k * log(n)$, where k is the number of free parameters of each model and n is the number of observations. Though we could not directly apply BIC in our case since we do not have a closed-form probabilistic equation describing our model where we can apply the maximum likelihood estimation method for our parameters, we approximate the BIC by treating the model results on all variables and participant data as a linear regression problem. For this, we computed the sum of squared residuals across all our predictions to estimate the LogLikelihood.

Table 2: Mean parameter values for UAM in Common Effect

	Common Effect				Common Cause			
Condition	$Pr(H)$	c	m	b	$Pr(H)$	c	m	b
Description only	.86(.22)	.57(.19)	.75(.19)	.22(.26)	.82(.18)	.56(.28)	.80(.21)	.24(.25)
Description + empirical	.91(.20)	.39(.17)	.54(.33)	.16(.20)	.84(.24)	.42(.25)	.64(.31)	.26(.20)
Empirical only	.88(.24)	.28(.16)	.50(.16)	.20(.28)	.78(.30)	.31(.24)	.47(.36)	.30(.20)

Note. In parenthesis the standard deviation.

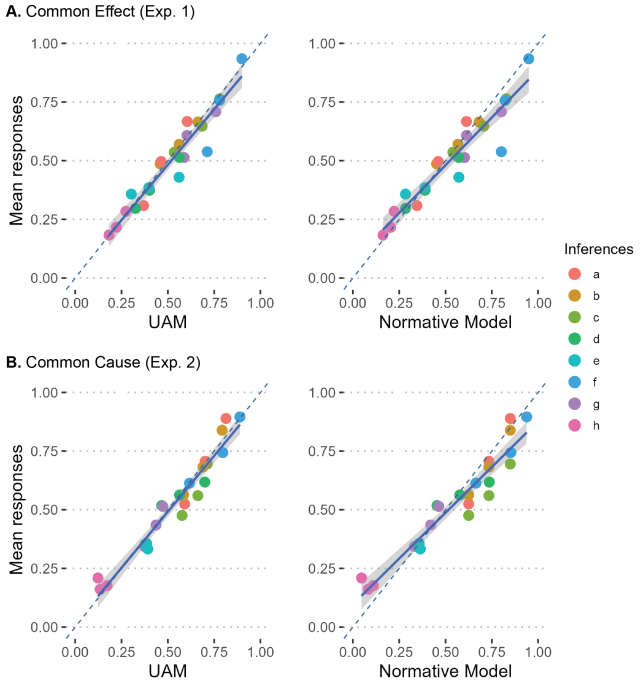

Figure 1: Correlations between model's predictions and empirical mean responses across inferences type for both experiments (panel A common effect (Exp. 1) and panel B common cause (Exp. 2)) in Rehder and Waldmann (2017).

dition, intermediate in the description+empirical condition, and highest in the description-only condition. This replicates the empirical results. More strikingly, the UAM (but, by definition, not the normative model where $Pr(H) = 1$) can reproduce the effects of the experimental manipulation on the magnitude of normative violations. In the common-effect condition, participants deviated from Explaining Away more in the description-only condition relative to the other two conditions, and the model reproduces this pattern (see inferences a, b, and c, Figure 2, upper row). In the common-cause condition, Markov violations are greater in the description-only and empirical+description conditions relative to the empirical-only condition, and this pattern also holds in the model predictions (see inferences a, b, and c, Figure 2. bottom row). These results suggest that the UAM is overall flexible enough to capture contextual manipulations of how information is presented in a causal inference domain.

In sum, our model is able to account for the key finding in Rehder and Waldmann (2017)'s study, namely that the researchers found "stronger deviations from the normative causal Bayes net model of causal reasoning in the conditions that described causal models compared to those that presented learning data." (Rehder & Waldmann, 2017, p.255). It is worth noting how our explanation differs from the explanation offered by the original authors. Rehder and Waldmann (2017) suggest that highlighting the causal structure of the system exacerbates a 'rich-gets-richer' bias in human causal reasoning. In contrast, our results suggest that participants might have inferred a different parameterization of the causal model in the different conditions. For example, the average best-fitting value of the causal strength parameter m in Exp.2 (common-cause) is $m = .80$ in the description-only, $m = .64$ in the empirical+description, and $m = .47$ in the empirical-only condition. This pattern might fall out of a human bias to assume that causal relations are near-deterministic (e.g., Lu et al., 2008; Schulz & Sommerville, 2006): this inductive bias presumably exerted less influence on participants' judgments when they were given learning data rather than a description of a causal model. Because the UAM predicts larger normative violations for higher values of the causal strength parameter (see simulations at the OSF: https://osf.io/), it can provide a natural explanation for the effect of interest.

Discussion

Deviations from the Markov condition and insufficient explaining away have been cornerstones of the argument that people may not be fully rational and fail to adhere to the basic probability calculus of Causal Bayesian Networks (CBNs; Rehder, 2018; Rottman & Hastie, 2016; Sloman & Lagnado, 2015). Normative predictions have been made through the implementation of CBNs. However, normative predictions of CBNs are incomplete because they are insufficient to explain common reasoning errors such as those discussed in this paper. Here, we adopt an approach that incorporates the pos-

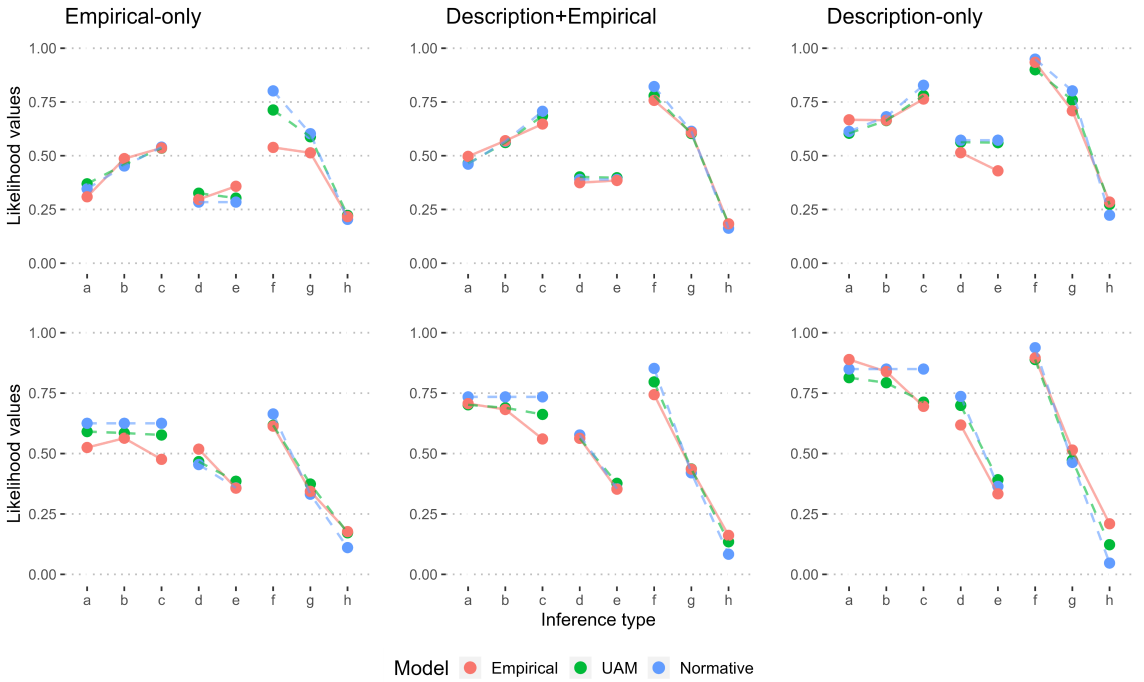

Figure 2: UAM and normative predictions compared to empirical responses in a common effect structure (upper panel; Exp. 1 from Rehder and Waldmann, 2017) and a common-cause structure (lower panel; Exp.2). Inference types are described in Table 1.

sibility that reasoners are uncertain about the received causal model (see Marchant et al., 2023b). Normative predictions of CBN assume that reasoners are completely certain that the received causal model is true (i.e., $Pr(H) = 1$), but as we showed in the modeling section, relaxing this assumption can go a long way in accounting for putative normative violations.

Rehder and Waldmann (2017) created different experimental conditions in which they manipulated the information participants were given about the received causal model. Their data reveal that people make systematic deviations from normative predictions, but also that the magnitude of these normative violations is influenced by the kind of information participants received about the causal model. Here, we showed that the Uncertainty-Augmented Model (UAM) is able to capture these two findings. As such the model can account not only for the existence of normative violations, but also variations in their magnitude across contexts.

Our explanation for the difference in the magnitude of normative violations across conditions has two key components. First, participants seem to have inferred a different parameterization of the causal model depending on the experimental condition[4]. Second, the UAM predicts that the magnitude of normative violations in causal reasoning depends on the causal model parameters inferred by the reasoner. To help the reader visualize how the model predictions depend on causal model parameters, we performed simulations by applying different value combinations of Pr(H), c and m. Those simulations can be found in the following OSF link: https://osf.io/. We hope these predictions inspire future empirical work.

Our explanation differs from those discussed in the literature (although we think it is possible that several different factors contribute to normative violations in causal reasoning; see Marchant et al. (2023b) for discussion). Instead of assuming a heuristic response (Rehder, 2018; Rottman & Hastie, 2016) or limited cognitive resources (Davis & Rehder, 2020; Kolvoort et al., 2023), the UAM assumes that people maintain some uncertainty about the causal representation of the relevant causal model. The parameterization of uncertainty as a null causal model different from the information received is somewhat similar to the use of contrast sets in other areas of reasoning. For example, Vance and Oaksford (2021) used contrast sets for a short learning period to explain empirical patterns in the implicit negation effect in conditional inference. They found that people revise their degree of beliefs from the information they receive as more data is provided to the reasoner. Taken together, our results suggest that assuming that rationality implies perfect belief in a received causal model may be an unnecessarily strong assumption.

[4]It is also possible that people have different levels of parameter uncertainty depending on the condition (e.g. they might consider that a wider range of values of c is possible in some conditions).

Acknowledgements

We thank Bob Rehder for providing us with his data from Rehder and Waldmann (2017) and for his valuable comments in the review process.

References

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological review*, *104*(2), 367.

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

Fenton, N., Neil, M., & Lagnado, D. A. (2013). A general structure for legal arguments about evidence using bayesian networks. *Cognitive science*, *37*(1), 61–102.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science*, *21*(3), 329–336.

Glymour, C. (2003). Learning, prediction and causal bayes nets. *Trends in cognitive sciences*, *7*(1), 43–48.

Hagmayer, Y. (2016). Causal bayes nets as psychological theories of causal reasoning: Evidence from psychological research. *Synthese*, *193*, 1107–1126.

Kolvoort, I. R., Temme, N., & van Maanen, L. (2023). The bayesian mutation sampler explains distributions of causal judgments. *Open Mind*, 1–32.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive science*, *37*(6), 1036–1073.

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. (2023a). A context-dependent bayesian account for causal-based categorization. *Cognitive Science*, *47*(1), e13240.

Marchant, N., Quillien, T., & Chaigneau, S. E. (2023b). Uncertainty can explain apparent mistakes in causal reasoning. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society*.

Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: Dispositional intuitions as a guide to causal structure. *Cognitive science*, *39*(1), 65–95.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121*(3), 277.

Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the markov property in causal reasoning. *Cognitive psychology*, *67*(4), 186–216.

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

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

Rehder, B. (2018). Beyond markov: Accounting for independence violations in causal reasoning. *Cognitive Psychology, 103*, 42–84.

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

Rehder, B. (2017a). Concepts as causal models: Categorization. *The Oxford handbook of causal reasoning*, 347–376.

Rehder, B. (2017b). Concepts as causal models: Induction. *The Oxford handbook of causal reasoning*, 377.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive psychology*, *50*(3), 264–314.

Rehder, B., & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & cognition*, *45*, 245–260.

Rips, L. J. (2008). Causal thinking. *Reasoning: Studies of human inference and its foundation*, 597–631.

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.

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological bulletin*, *140*(1), 109.

Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child development*, *77*(2), 427–442.

Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. Oxford University Press.

Sloman, S. A., & Lagnado, D. (2015). Causality in thought. *Annual review of psychology*, *66*, 223–247.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207–232.

Vance, J., & Oaksford, M. (2021). Explaining the implicit negations effect in conditional inference: Experience, probabilities, and contrast sets. *Journal of Experimental Psychology: General*, *150*(2), 354.

Wang, M., & Sun, J. (2020). A situation-modulated minimal change account for causal inferences about causal networks. *Quarterly Journal of Experimental Psychology*, *73*(12), 2403–2411.