Reasoning with Inconsistent Causal Beliefs

John V. McDonnell (john.mcdonnell@nyu.edu)[1]
Pedro Tsividis (tsividis@mit.edu)[2]
Bob Rehder (bob.rehder@nyu.edu)[1]

Abstract

Causal reasoning is a critical part of everyday cognition. We ask how people reason about causes when faced with inconsistent sources of knowledge. Causal models arise from multiple sources of information regarding their constituent parameters. Knowledge sources may be inconsistent both *within* parameters (when one source says a variable should appear often and another says it should appear rarely), and *between* parameters (when dependencies among parameters result in an internally inconsistent causal model). We provide a normative model for resolving both these sources of conflict. An experiment found that our model of belief integration predicted the qualitative pattern of adults causal inferences under uncertainty.

Keywords: Causal Learning; Causal Inference; Probabilistic Modeling

Introduction

From deciding on investment strategies to predicting others' reactions in social situations, we are constantly making probabilistic judgments about causal systems. However, because we receive information from multiple sources, we are often faced with contradictory beliefs. Consider the problem faced by an epidemiologist trying to understand the causes of chronic hypertension in a particular population. She reads a review paper suggesting that smoking tobacco causes hypertension in 50% of patients, and that all other potential causes of hypertension can be ruled out. The epidemiologist knows from survey data that 25% of the population of interest are smokers, and (independently) that 25% have hypertension. If she assumes her maximum likelihood estimates are true, she is left with an incoherent causal model: If smoking is the only cause of hypertension, and is effective half the time, then there should be half as many people with hypertension as there are smokers. Arriving at coherent causal beliefs will require adjustment. Perhaps hypertension isn't really as prevalent as she thought, or perhaps the smoking always causes hypertension. In this paper, we propose a normative probabilistic model for reasoning with these inconsistencies and explore the implications of that model in an experiment in which participants receive conflicting sources of evidence.

A Model

We assume that people's causal inferences are based on causal graphical models (CGMs), such as the one in Figure 1 in which $C1$ and $C2$ are believed to be independent causes of E. For simplicity, throughout the paper we will assume that all variables are binary (present or absent) and that all causal links are generative, bringing about their effects through a *noisy-or* functional form. We allow for the possibility of additional causes of E not shown in the graph by aggregating them into a single background cause that is always present (for a review of graphical models, see Koller & Friedman, 2009). A fully parameterized CGM is sufficient to answer virtually any question one might want to ask about the variables involved, including questions of conditional or joint probability, counterfactual reasoning, and predicting the effect of interventions (Pearl, 2000). However, complications arise when one recognizes that CGMs are constructed from many individual beliefs held by the reasoner. Since these beliefs may come from multiple sources that vary in their reliability, it is inevitable that they will sometimes contradict one another. We ask: How should one draw causal inferences in light of such inconsistencies?

To answer this question, we first note that inconsistencies can be either between or across parameters, where parameters represent one's beliefs about each constituent of the model. For example, a belief about the probability of a cause being present is one constituent; a belief about the strength of a causal relation is another. In the first section we advance a new proposal for representing uncertainty in CGMs and show how it solves the problem of within-parameter conflicts. We then tackle the more challenging problem of between-parameter conflicts.

Resolving conflicts within parameters

Consider the problem of representing the base rate of variable $C1$, represented by parameter c_1 in Figure 1. We suppose that beliefs about base rates may come from first-hand observations (observing the prevalence of $C1$), explicit, instruction (e.g., hearing that $C1$ is uncommon), and prior beliefs (e.g., a tendency to believe that events of this type arise rarely). Because probability is bounded to the range $[0, 1]$, the information from each of these sources can be encoded as a probability density function in that range. If knowledge is represented as a point estimate of the expected value of the variable combined with a confidence in that point estimate, this information can be encoded as a beta distribution. PDFs of beta distributions representing knowledge sources for $C1$ are shown at the top of Figure 1. The shape of a beta distribution is controlled by two parameters, α and β, constrained to be positive (we will refer to such parameters as knowledge parameters, or *k-parameters*, to emphasize that they represent participants' knowledge and to avoid confusion with the constituent parameters).

[1] New York University, Department of Psychology, 6 Washington Place, New York, NY 10003 USA

[2] Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, 77 Massachussetts Ave., Cambridge, MA 02139 USA

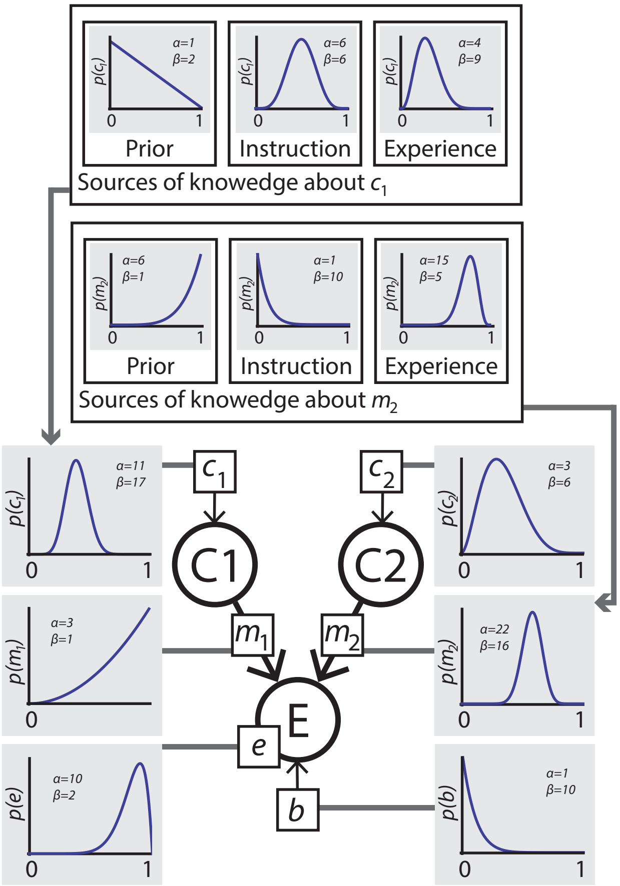

Figure 1: A simple common-effect causal graphical model. Here $C1$ and $C2$ are potential causes of E. Variables of boxes indicate the parameters of the model. Our conflict-resolution model assumes participants encode their beliefs about each of these variables as beta distributions, depicted alongside each of the variables. Above, knowledge sources for c_1 and m_2 are depicted, also represented as beta distributions.

The expected value of the odds is the ratio of the two k-parameters, while their sum can be interpreted as the confidence in that estimate.

Reconciling different beliefs encoded as beta distributions is simple. Bayes' rule gives the posterior belief as the renormalized product of the prior and likelihood distributions. In the case of beta distributions, this is derived by summing the parameters. This means that for the knowledge sources in Figure 1, if we denote the k-parameters α_{prior}, β_{prior}, $\alpha_{instruct}$, $\beta_{instruct}$, α_{exp}, and β_{exp}, the posterior is simply $\text{Beta}(\alpha_{prior}+\alpha_{instruct}+\alpha_{prior}, \beta_{prior}+\beta_{instruct}+\beta_{prior})$. This process is depicted in Figure 1 for c_1.

A similar treatment can be applied to the strengths of the causal relations in Figure 1, shown in the figure for the case of m_2. Following Cheng (1997), we assume that each link is represented as a *causal power*: the propensity of the cause, when present, to bring about the effect. Because they are probabilities, beliefs about causal powers can also be stored as beta distributions.

As depicted in Figure 1, the model consists of six parameters: the base rates of $C1$ and $C2$ (c_1 and c_2), the strengths of causal relationships $C1 \rightarrow E$ and $C2 \rightarrow E$ (m_1 and m_2), the strength of the background causes of E (b), and the rate at which E occurs (e). We suppose that belief in the value of each of them is represented as a posterior beta distribution.

Resolving conflicts between parameters

Computing the posterior beta distribution for each model parameter does not eliminate all potential inconsistencies, however. As illustrated in the introduction, when all the causes of an effect are fully described, the effect variable's rate of occurrence is no longer free to vary. Because the value of e is fixed, random draws from the individual beta posteriors will never return valid causal models (assuming infinite precision). This means we need a way to integrate information about effects into our beliefs over possible causal models without violating the constraints of the model.

Following Figure 1, let $\mathbf{V}$ represent the set of variables in the domain. For each $v \in \mathbf{V}$, belief in the probability of v is characterized by the PDF of a beta distribution, denoted π_v. These correspond to c_1, c_2, and e in the figure. Next, let $\mathbf{L}$ be the set of causal links in a model. For each $l \in \mathbf{L}$, the learner's belief in the causal power of l is characterized by a beta-distributed PDF denoted π_l (m_1 and m_2 in the figure). Finally, let $\mathbf{E} \subset \mathbf{V}$ be the effects. For each $e \in \mathbf{E}$, the belief in the background causes of e is characterized by a beta-distributed PDF denoted π_e (b in the figure).

We now define the posterior over valid causal models. Let $\mathbf{r}$, $\mathbf{m}$, and $\mathbf{b}$ be vectors describing the base rate of every variable in $\mathbf{V}$, the strength of every link in $\mathbf{L}$, and the strength of the background causes for every effect in $\mathbf{E}$, respectively. Under a noisy-or causal model, all effects are explained by the likelihood of their causes and the strength of those causes. This means the rate of occurrence for an effect e is constrained to be

$$r'_e = 1 - (1 - b_e) \prod_{(l \in \mathbf{L}) \wedge (l_e = e)} [1 - r_{l_c} m_l] \qquad (1)$$

where l_c and l_e are the cause and effect variables associated with causal link l. Enforcing this consistency, we can define the joint probability of a fully parameterized model as

$$P(\mathbf{r}, \mathbf{m}, \mathbf{b}) \propto \begin{cases} 0 & \text{if } \exists e \in \mathbf{E} \colon r_e \neq r'_e \\ \displaystyle\prod_{v \in \mathbf{V}} \pi_v(r_v) \prod_{l \in \mathbf{L}} \pi_l(m_l) \prod_{e \in \mathbf{E}} \pi_e(b_e) & \text{otherwise.} \end{cases} \qquad (2)$$

This is equivalent to saying that the posterior over joint model values is defined as the result of sampling from the beta distributions characterizing each of the variables in the model, discarding the inconsistent models. Our central hypothesis is that, when inconsistencies among be-

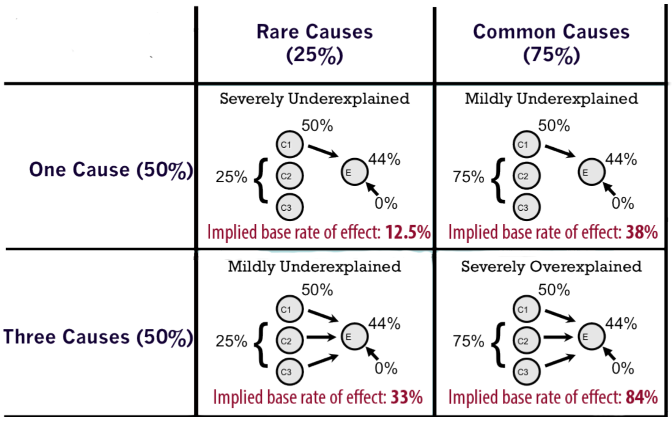

Figure 2: Design of the experiment.

liefs exist, people draw inferences as if they are reasoning with the *maximum a posteriori* causal model.

An Experiment

One important prediction of our model is that when faced with an invalid model, reasoners will make tradeoffs among the parameters to find a valid causal model (Equation 2 above). We tested this hypothesis by querying the adjustments individuals make to their instructed causal model given a variety of inconsistent beliefs.

Each participant learned about four binary variables in one of three domains: economics, meteorology, or sociology. Participants learned about four relevant variables and the beliefs of experts in their domain. In the domain of economics, for example, the four variables were interest rates (moderate/high), trade deficits (moderate/large), retirement savings (moderate/high), and job mobility (moderate/high). Depending on the condition, one or three of those variables (denoted C_1, C_2, and C_3) were been described as generative causes of a fourth (E).

A 2×2 design (depicted in Figure 2) explored the effect of varying the number of causal links (one vs. three) and the base rates of those causes ("rare" vs. "common"). In all conditions, the effect was "somewhat common," (occurring 44% of the time), and the causes brought about the effect half the time.

Manipulation of these two parameters should of course result in changes to participants' estimates of the values of those parameters. However, because they also imply between-parameter conflicts, they should also result in compensating changes to other parameters in order to yield a consistent model. For example, consider the condition in which only one rare cause is instructed. Here the participants were told about a cause happening 4 times out of 16 and bringing about its effect 50% of the time. As indicated in the figure, these facts imply a base rate for the effect of .125, which conflicts with its instructed base rate of .44. Thus, the effect is *underdetermined* by the causes in the model. There are many ways in which reasoners could compensate: They could adjust the likelihood of the effect downward, increase the likelihood of the cause, increase the causal strength of the cause, or increase the likelihood of background causes. Participants must choose a combination of these adjustments to reason with a valid causal model.

Conversely, in the case where participants were instructed about three causes, each common, the implied base rate of the effect is .84, that is, the effect is now *over-determined*, and the reverse adjustments are needed to help form a valid model. Note that because we did not explicitly control the confidence subjects should hold in their beliefs about individual parameters, we do not make predictions regarding which variables will be adjusted, but rather only that some subset will be adjusted to attain a consistent model.

To measure the adjustments made by participants, we followed this instruction with a series of questions designed to assess their beliefs about the parameters of the causal model.

Method

Participants A total of 234 subjects were tested, consisting of 114 New York University undergraduates who received course credit and 120 online subjects who received a small cash incentive. Subjects were randomly assigned to the 1-link/rare, 1-link/common, 3-link/rare, and 3-link/common, conditions. Participants whose numerically-coded responses over the course of the experiment had a standard deviation of less than 2 were excluded, leaving 54, 52, 57, and 47 participants in each of the conditions, respectively.

Materials Three knowledge domains were tested: Economics, meteorology, or sociology. In each domain, the same four variables were used, so the same variables always played the role of C_1, C_2, C_3, and E. Subjects in the 3-link conditions learned three causal links: $C_1 \rightarrow E$, $C_2 \rightarrow E$, and $C_3 \rightarrow E$. Those in the 1-link conditions learned only $C_1 \rightarrow E$.

Base rate information was displayed as an instruction screen displaying a pictorial representation of the rates at which each of the variables was observed. Base rates of the causes were described as occurring 25% of the time ("rare"), or 75% ("common"). E was always depicted as occurring 44% of the time ("somewhat common"). The base rates were illustrated using a pictorial graph showing them the values of each variable for 16 random systems from a "survey". For example, $\frac{7}{16}$ of systems were always shown to have the effect.

Causal information was conveyed in writing. For each causal relationship, participants were told that the cause brought about the effect 50% of the time. They were also given a short description of the mechanism underlying the relation. For example, if told that large trade deficits cause high job mobility, they were also told, "The flood of cheap imports means that many domestic manufacturing jobs are lost and workers must find new employment." Participants were also told that experts believed there were no other causes of the effect.

Procedure After being introduced to the domain, participants were presented with screens presenting experts' beliefs about the base rates of the variables and their causal relationships. After the instructions, online subjects were quizzed on their memory for the instructions. This repeated in a loop until they were able to correctly answer all questions.

Participants were asked four types of test questions presented in blocks. Block order was randomized. For all question types, participants responded with an estimated probability on a scale of 1–11 by choosing one of 11 radio buttons, with the two extremes labeled "very unlikely" (on the

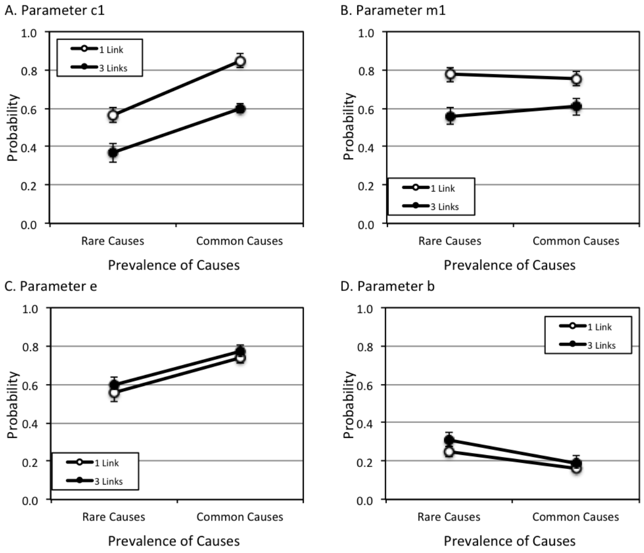

Figure 3: Causal model parameters derived from the empirical results from the experiment compared across all conditions (without modeling learning or prior belief). (A) The base rates of the causes (parameter c_1). (B) The strength of the causal links (parameter m). (C) The base rate of the effect (parameter e). (D) The strength of alternative causes (parameter b). Error bars are standard errors.

left) to "very likely" (on the right). The first questions were joint probability questions. In these questions, subjects were given the states of all four variables in their scenario and asked to rate the likelihood of those variable states being observed. All 16 possible questions formed by varying the state of the three binary variables were asked. A second question type consisted of conditional probability questions. For these questions, subjects were given the state of three of the four variables and asked to rate the likelihood of the fourth. All 8 possible conditional probability judgments formed asking for the probability of E as a function of the presence or absence of the three causes were asked. In addition, eight questions asked for the probability of one of the causes with the effect either present or absent (with the other causes always stipulated to be absent). Finally, the third and fourth question types directly queried the base rates of the four variables and the strengths of the three potential causal links. To avoid the possibility that subjects would forget the initial information about the causal model, a "theory reminder" was presented on the right side of the screen, accompanying each question.

Finally, subjects were given the chance to learn more about their causal system by observing a sample of 32 instances from that domain. The sample was drawn from a model in which the causes had a base rate of .50, two causal links ($C_1 \rightarrow E$ and $C_2 \rightarrow E$) of strength .50, with no alternative causes of E. Subjects were asked to "consider how these data might change your beliefs about the causal relationships in this system and the overall likelihoods of the variables involved." Participants were then asked to re-answer all the previous questions. Then they cycled through once more, again observing a sample of 32 instances, and again re-answering the questions. These responses were used in model fitting, but because learning effects appeared to be small the results of these test phases will not be reported here.

Results

To characterize subjects' judgments of joint and conditional probability, we fit those judgments to a causal model with three causes and one effect, yielding eight parameters: c_1, c_2, and c_3 (the likelihoods of the causes), m_1, m_2, and m_3 (the strengths of the putative causal links), b (the background cause of e) and e, the likelihood of the effect (see the Appendix for details).

Figure 3 summarizes the effects of our two manipulations on the causal model parameters. For purposes of comparison, we only present those parameters that were involved in a causal relationship in all conditions (c_1, m_1, e, and b). A 2×2 ANOVA with the causes' base rate and the number of causal links as factors was performed for each panel. A main effect of the base rate manipulation on estimates of base rates of C_1 (plotted in Figure 3A) confirmed the success of that manipulation ($F(1, 206) = 35.45$, $MSE = .034$, $p < .001$). Importantly, the manipulation also resulted in an increase in the prevalence of the effect (parameter e in Figure 3C, $F(1, 206) = 19.00$, $MSE = .074$, $p < .001$) and a decrease in the strength of the background causes (parameter b in Figure 3D ($F(1, 206) = 10.23$, $MSE = .058$, $p < .01$). That is, to accommodate the more prevalent causes, participants compensated by increasing the base rate of the effect and decreasing the effectiveness of alternative causes, as predicted by our model. There was no effect of the manipulation on causal strengths (parameter m_1 in Figure 3B).

The manipulation of the number of causal links also had two important effects. First, it reduced the base rate of C_1 (parameter c_1 in Figure 3A, $F(1, 206) = 35.45$, $MSE = .073$, $p < .001$). Second, it reduced the strength of the $C_1 \rightarrow E$ causal relationship (parameter m_1 in Figure 3B, $F(1, 206) = 20.67$, $MSE = .088$, $p < .001$). Apparently, to accommodate two additional causal links, participants compensated by decreasing the effectiveness of the $C_1 \rightarrow E$ link, reducing both c_1 and m_1 as predicted by our model. Changing the number of causal links did not have a significant effect on either the prevalence of the effect (parameter e in Figure 3C, $F < 1$) or the strength of alternative causes (Figure 3D, $F(1, 206) = 2.16$, $MSE = .058$, $p = .19$). None of the 2-way interactions approached significance (all Fs < 1).

Discussion

The results of this experiment supported the claim that when people are given inconsistent information, they draw inferences as if they're reasoning with the most likely causal model. Increasing the base rates resulted in participants believing that the effect is more likely and alternative causes are weaker. Increasing the number of causes led participants to adjust their beliefs about the causes, weakening both their efficacy and their base rates. We now assess whether our theoretical model can provide not only a good qualitative account of these data, but an acceptable quantitative one as well.

Theoretical Modeling

Recall that although our theoretical model of uncertainty and belief integration specifies that reasoners will adjust

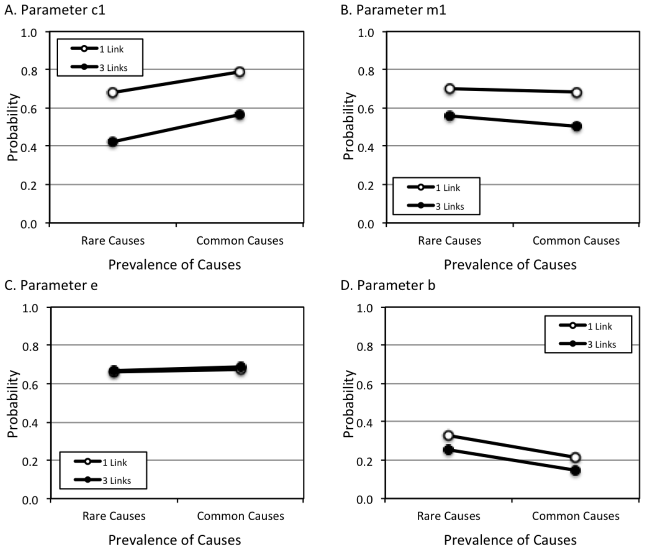

Figure 4: Causal parameters derived from the fit of the theoretical model. (A) The base rates of the causes (parameter c_1). (B) The strength of the causal links (parameter m). (C) The base rate of the effect (parameter e). (D) The strength of alternative causes (parameter b).

parameters in order to reason with a consistent causal model, it does not specify which parameters will be adjusted in the absence of any information about the confidence with which beliefs about those parameters are held. To assess our model's potential for providing a quantitative account of the experiment, we fit it to those results treating confidence in each constituent parameter as a free parameter (a fifth control condition, not reported here, was also fit in which subjects were instructed on the same model that generated the learning data; see Procedure section). For any constituent parameter of the model k (a variable or causal link), confidence is represented as a beta distribution with k-parameters α_k and β_k. We recast those k-parameters into the pair $v_k = \frac{\alpha_k}{\alpha_k+\beta_k}$, representing the expected value of the parameter, and $t_k = \alpha_k + \beta_k$, representing the overall confidence with which the expected value is held. We assume that reasoners may have different levels of confidence in the different types of parameters, represented as four free k-parameters: t_c (the base rates of the cause), t_m (the strength of the causal links), t_b (the strength of the alternative causes), and t_e (the base rate of the effect). Because we assume that subjects do not perfectly encode the initial numerical information provided about each parameter, those are free parameters as well. Let k-parameters v_m, v_b, and v_e represent the instructed values of m (causal link strengths, described to subjects as .5), b (the strength of the alternative causes, described as 0), and e (the base rate of the effect, described as .44), respectively. v_{m0} was the initial strength of the links on which subjects were not instructed. k-parameters v_{c-r}, v_{c-m} v_{c-c} represent the initial base rates of the causes in the rare, moderate, and common conditions, respectively (described as .25, .5, and .75).

This model was fit to the group level causal model parameters fit in the experiment. Eight parameters (c_1, c_2, c_3, m_1, m_2, m_3, b, and e) were estimated per phase per condition. This included the learning phases as well as a fifth control condition not reported here. This involved fitting $8 \times 3 \times 5 = 120$ data points with 10 parameters. The parameters that minimized squared error were $t_c = 106$, $t_b = 299$, $t_e = 3252$, $v_{c-r} = .162$, $v_{c-m} = .351$, $v_{c-c} = .485$, $v_m = .454$, $v_{m0} = .069$, $v_b = .122$, and $v_e = .7$. The correlation between observed and predicted values was .964. The predictions are depicted in Figure 4, which is analogous to Figure 3. Figure 4 reveals that the model is able to capture the effects of the causal strength manipulation, namely the base rate of C_1 (parameter c_1, Figure 4A) and the strength of the $C_1 \rightarrow E$ link (parameter m_1, Figure 4B) both decrease as the number of causal relations increases (compare with Figure 3A and B, respectively). It is also able to reproduce the effects of the base rate manipulation on the strength of the alternative causes (parameter b, Figure 4D), namely, that alternative causal strength decreases as the base rates of the causes increase (compare with Figure 3D). Less successfully, it predicts an increase in the base rate of the effect with larger base rates of the causes (parameter e, Figure 4C), although the magnitude of that change is much smaller than the one exhibited by subjects (compare with Figure 3C). Note the insensitivity of parameter e to changes in the other causal model parameters is a manifestation of the large confidence the model places on its initial value ($t_e = 3252$, vs. all other ts < 300).

General Discussion

In ecologically valid settings, causal reasoning often takes places with multiple knowledge sources that are potentially inconsistent with one another. To specify how causal inferences should be drawn in such situations, we developed an account of how uncertainty about causal models might be represented and then showed how to derive the most likely causal model that is sensitive to each knowledge source yet resolves inconsistencies between them. Our central hypothesis was that people would draw causal inferences as if they were reasoning with the most likely consistent model.

The qualitative predictions of this model were confirmed in an experiment manipulating two instructed parameters: the base rates of the causes and the number of causal links. Making causes more prevalent resulted in alternative causes becoming weaker and the effect becoming more prevalent. Making causal relations more numerous resulted in the causes becoming rarer and other causal links becoming weaker. We know of no other model that is capable of predicting these sorts of effects.

Our model also yielded moderately good quantitative fits to the data. One result it was unable to reproduce was participants' tendency to adjust the likelihood of the effect (e) when the causal base rates were adjusted but not when the number of causes was adjusted. Moreover, we acknowledge that these fits used a large number of parameters, necessitated by the fact that confidence in each instructed model parameter was not specified experimentally and so needed to be free parameters. We are conducting follow-up studies manipulating instructed confidence in the information provided to participants.

Although our representation of uncertainty was sufficient to account for our empirical results, its assumption that the distributions of the causal model parameters are independent is unrealistic in some situations. For example, Lu et al. (2008) have modeled the traditional causal learning experiment as one in which the prior distribution is a two-dimensional density function on the strength of the to-be-learned causal link and the strength of alternative causes. Multivariate representations of uncertainty like this may be common. In addition, one might imagine that reasoners not only have experiential knowledge about the base rates of variables (Figure 1), but also about configurations of variables. Finally, in addition to changing parameters to attain consistency, reasoners might also change the function relating an effect to its causes (e.g., by assuming that the causes combine interactively rather than independently) or even the structure of the model itself (e.g., deleting a causal link, as proposed by Griffiths & Tenenbaum, 2005).

One facet of the data not discussed above was the distinction between our explicit queries (e.g., directly querying the causal efficacy of C_1) and our implicit ones, such as judgments of conditional and joint probability. Perhaps unsurprisingly, explicit queries more closely resembled the likelihood information participants were given, unadjusted for consistency, suggesting that such questions do not invoke inconsistency resolution processes we have specified here. We also observed (but did not report here) that participants' causal models changed little as a result of observing data. Whether this result reflected a kind of anchoring effect (initial judgments influenced later ones) or participants large confidence in the initial domain theories on which they were initially instructed remain questions for future research.

Acknowledgments

This work was supported by the Air Force Office of Scientific Research, Grant No. FA9550-09-NL404.

References

Cheng, P. (1997). From covariation to causation: a causal power theory. *Psych. Rev. 104*, 367.

Griffiths, T. L. & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cog. Psych. 51*, 334–384.

Koller, D. & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques.* Cambridge, MA: The MIT Press.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psych. Rev. 115*, 955–984.

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

Appendix

Participants' causal models were simultaneously fit to participants' judgments of conditional and joint probability. Fitting assumed that participants formed a common-effects model with C_1, C_2, and C_3 as potential causes of E. The joint is defined as the probability that the four variables will take any particular combination of values. From the axioms of probability, we derive

$$p(E, C_1, C_2, C_3) = p(E|C_1, C_2, C_3)p(C_1, C_2, C_3). \tag{3}$$

Because C_1, C_2, and C_3 are assumed to be independent,

$$p(E, C_1, C_2, C_3) = p(E|C_1, C_2, C_3)p(C_1)p(C_2)P(C_3). \tag{4}$$

Assuming that the causes bring about their effects according to a noisy-or rule, the probability that E is present given the status of the causes is given by

$$p(E = 1|C_1, C_2, C_3) = 1 - (1 - b) \prod_{i \in \{1,2,3\}} (1 - m_i C_i), \tag{5}$$

where presence or absence is coded as 1 or 0, respectively.

Equations 4 and 5 are sufficient to specify the probability of any combination of the variables as a function of the parameters c_1, c_2, c_3, m_1, m_2, m_3, and b.

Separate c, m, and b parameters were estimated for each participant for each test phase. To transform responses on the 1–11 scale into probabilities, we applied a nonlinear (power) transformation. This necessitated fitting a power parameter, γ. Each subject's rankings were predicted as follows:

$$rating_{\text{cond}}(r_{b,i}) = 10 p_k(r_i; c_b, m_b, B_b)^{\gamma_{\text{cond}}} + 1 \tag{6}$$

$$rating_{\text{joint}}(o_{b,i}) = 10 p_k(r_i; c_b, m_b, B_b)^{\gamma_{\text{joint}}} + 1 \tag{7}$$

where $r_{b,i}$ and $o_{b,i}$ are the subject's conditional and joint judgments, respectively, on trial i in phase b. γs were fit within participant and question type, and constrained to the range $[0, 5]$. This resulted in $7 \times 3 = 21$ causal model parameters and the two γ parameters (23 in total) used to fit $32 \times 3 = 96$ responses. Parameters were fit using gradient descent.

The model fits included a control condition in which participants were instructed with a causal model that conformed to the data they observed. The results are not elaborated in this paper due to space constraints.

This model achieved a respectable correlation between its predictions and the empirical data points (.950; .783 averaged over subjects).