

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Reasoning with Conjunctive Causes

Permalink

<https://escholarship.org/uc/item/6c8397h1>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN

1069-7977

Author

Rehder, Bob

Publication Date

2011

Peer reviewed

Reasoning with Conjunctive Causes

Bob Rehder (bob.rehder@nyu.edu)

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

Abstract

Conjunctive causes are causes that all need to be present for an effect to occur. They contrast with *independent causes* that by themselves can each bring about an effect. We extend existing “causal power” representations of independent causes to include a representation of conjunctive causes. We then demonstrate how independent vs. conjunctive representations imply sharply different patterns of reasoning (e.g., *explaining away effects* for independent causes as compared to *exoneration effects* for conjunctive causes). An experiment testing how people reason with independent and conjunctive causes found that their inferences generally matched the model’s prediction, albeit with some important exceptions.

Rather than operating in a vacuum, causes frequently interact with other factors to produce their effects. For example, the conjunction of two or more variables is often necessary for an outcome to occur. A spark may only produce fire if there is fuel to ignite, a virus may only cause disease if one’s immune system is suppressed, the motive to commit murder may result in death only if the means to carry out the crime are available. Sometimes, conjunctive causes take the form of *enablers*. For example, the presence of oxygen enables fire given spark and fuel. In contrast, *disablers* interact with existing causes by preventing normal outcomes. Although the eight ball’s path to the side pocket may appear inevitable, it may be interrupted by an earthquake, a falling ceiling tile, or a spilled beer.

The last 20 years has seen a growing interest in the role of causal knowledge in numerous areas of cognition. Many studies have investigated how causal relations are learned from observed correlations (Cheng, 1997; Gopnik et al., 2004; Griffiths & Tenenbaum, 2005; 2009; Lu et al., 2008; Sobel et al., 2004; Waldmann et al., 1995). Others have tested the impact of causal knowledge on various forms of reasoning, including inference (Rehder & Burnett, 2005; Kemp & Tenenbaum, 2009), interventions (Sloman & Lagnado, 2005; Waldmann & Hagmeyer, 2005), analogy (Holyoak et al., 2010; Lee & Holyoak, 2008), generalization (Rehder, 2006; 2009), and classification (Rehder & Hastie, 2001; Rehder 2003a; b; Rehder & Kim, 2006; 2009; 2010). But although some studies have investigated the *learning* of interactive causes (e.g., Novick & Cheng, 2004), their role in *reasoning* has received little attention. This article tests how people reason with one sort of interactive cause—conjunctions.

How *should* one reason with conjunctive causes? One popular framework for modelling learning and reasoning with causal knowledge is *Bayesian networks* or *causal graphical models* (hereafter, *CGMs*). In CGMs, variables

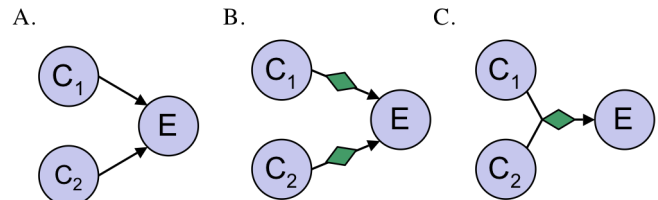


Figure 1.

are represented as nodes and causal relations as directed edges. For example, Figure 1A presents a CGM in which variables C₁ and C₂ are causes of variable E. CGMs are popular in part because they specify the *causal Markov condition* that stipulates patterns of conditional independence between variables and which has important implications for how one learns and reasons with causal knowledge.

By itself, however, a CGM says nothing about the functional relationship between an effect and its causes. For example, Figure 1A does not specify whether C₁ and C₂ are independent or interactive causes of E. Two possibilities are represented in Figures 1B and 1C. In these figures, we assume that C₁, C₂, and E are binary variables that are either present or absent. Diamonds represent independent *generative causal mechanisms*, processes that work to produce the effect when their causes are present. Figure 1B represents the fact that C₁ and C₂ are independent causes of E—that is, that E might be brought about by C₁ or C₂. Figure 1C represents that C₁ and C₂ are conjunctive causes of E—E is brought about only when C₁ and C₂ are both present. As mentioned, there are other ways that E might depend on an interaction between C₁ and C₂ (e.g., C₂ might *disable* the causal link between E and C₁), but here we focus on the contrast between independent and conjunctive causes. Below we specify the (independent or conjunctive) functions that relate an effect and its causes and derive the different patterns of inferences implied by those functions.

Other frameworks do not readily distinguish between alternative interpretations of Figure 1A. For example, reasoners may treat it as an associative network, in which case they may infer one variable given others without regard to the direction of causality. Or, they may treat it as a dependency network that is sensitive to causal direction (E depends on C₁ and C₂) but not their functional relationship. Accordingly, our goal is to first establish that reasoners indeed distinguish between independent and conjunctive causes and then determine whether they do so in the manner predicted by our proposed representation of conjunctions.

Reasoning with Conjunctive Causes

To test how people reason with conjunctive causes, subjects

were instructed on a novel category with six features. For example, subjects who learned Romanian Rogos (a type of automobile) were told that Rogos have a number of typical or characteristic features (e.g., butane-laden fuel, a loose fuel filter gasket, hot engine temperature, etc). In addition, subjects were instructed on the interfeature causal relations shown in Figure 2. Two features (IC_1 and IC_2 in Figure 2) were described as independent causes of IE whereas CC_1 and CC_2 were described as conjunctive causes of CE. Subjects were then presented with an inference test in which they predicted one Rogo feature given the state of others.

To derive predictions for this experiment, we first specify the joint probability distribution for each of the two CGMs represented by the subnetworks in Figure 2 and then use those distributions to derive expected inferences.

Specifying the Joint Distributions

Independent cause network. We first specify the joint distribution for the independent cause network, $p_k(IC_1, IC_2, IE)$, that is, the probability that IC_1 , IC_2 , and IE will take any particular combination of values in category k . From the axioms of probability theory we have,

$$p_k(IC_1, IC_2, IE) = p_k(IE | IC_1, IC_2) p_k(IC_1, IC_2) \quad (1)$$

Because IC_1 and IC_2 have no common causes in the independent cause network (and because the *causal sufficiency* constraint on CGMs rules out them having a hidden common cause, Spirites et al. 1993) they are assumed to be independent. Eq. 1 thus becomes,

$$p_k(IC_1, IC_2, IE) = p_k(IE | IC_1, IC_2) p_k(IC_1) p_k(IC_2) \quad (2)$$

$p_k(IE | IC_1, IC_2)$ can be written as a function of parameters that characterize the generative causal mechanisms that relate IE to its causes. Specifically, $m_{IC_1, IE}$ and $m_{IC_2, IE}$ are the probabilities that those mechanisms will produce IE when IC_1 and IC_2 are present, respectively. In terms introduced by Cheng (1997), these probabilities refer to the “power” of the causes. In addition, to allow for the possibility that IE has additional causes not shown in Figure 2, b_{IE} is the probability that IE will be brought about by some other cause. With these definitions, the probability that IE is present is given by the familiar “fuzzy-or” equation,

$$p_k(IE = 1 | IC_1, IC_2) = 1 - (1 - b_{IE}) \prod_{i=1,2} (1 - m_{IC_i, IE})^{ind(IC_i)} \quad (3)$$

where $ind(IC_i)$ returns 1 when IC_i is present and 0 otherwise.

Equations 2 and 3 are sufficient to specify the probability of any combination of IC_1 , IC_2 , and IE. These expressions

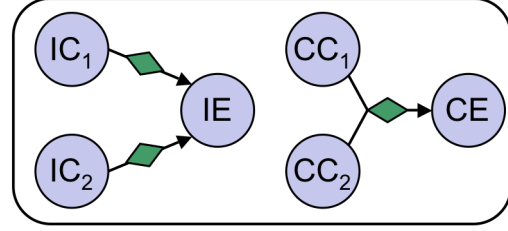


Figure 2.

are shown in the left hand side of Table 1. For example, the probability that IC_1 and IE are present and IC_2 absent is

$$\begin{aligned} p_k(IC_1 = 1, IC_2 = 0, IE = 1) \\ &= p_k(IE = 1 | IC_1 = 1, IC_2 = 0) p_k(IC_1 = 1) p_k(IC_2 = 0) \\ &= [1 - (1 - b_{IE})(1 - m_{IC_1, IE})] c_{IC_1} (1 - c_{IC_2}) \end{aligned}$$

where c_{IC_1} and c_{IC_2} are the probabilities that IC_1 and IC_2 , will appear in members of category k , respectively.

Conjunctive cause network. The joint distribution for the conjunctive cause network, $p_k(CC_1, CC_2, CE)$, can be written in a manner analogous to Eq. 2,

$$p_k(CC_1, CC_2, CE) = p_k(CE | CC_1, CC_2) p_k(CC_1) p_k(CC_2) \quad (4)$$

The conjunctive cause network in Figure 2 differs from independent causes in having one generative causal mechanism. We extend the notion of “causal power” to conjunctive causes by assuming that, when CC_1 and CC_2 are both present, that mechanism will bring about CE with probability $m_{IC_1, IC_2, IE}$. Thus we have,

$$p_k(CE = 1 | CC_1, CC_2) = 1 - (1 - b_{CE}) (1 - m_{IC_1, IC_2, IE})^{ind(CC_1, CC_2)} \quad (5)$$

where b_{CE} is the probability that CE will be brought about by some other cause, and $ind(CC_1, CC_2)$ returns 1 when CC_1 and CC_2 are both present and 0 otherwise. Equations 4 and 5 are sufficient to specify the probability of any combination of CC_1 , CC_2 , and CE, as shown in the right hand side of Table 1. c_{CC_1} and c_{CC_2} are the probabilities that CC_1 and CC_2 will appear in members of category k , respectively

Theoretical Predictions

Given the joint distributions in Table 1, it is straightforward to compute the conditional probability of any feature given the state of any other features in the same subnetwork. To demonstrate the qualitative pattern of these inferences, we instantiate the joint distributions by assigning the causal model parameters with values that are hypothetical but also reasonable in light of conditions established in the upcom-

Table 1

Independent Causes				Conjunctive Causes			
IC_1	IC_2	IE	$p_k(IC_1, IC_2, IE)$	CC_1	CC_2	CE	$p_k(CC_1, CC_2, CE)$
				$c_{IC_1} = c_{IC_2} = .67;$			
				$m_{IC_1, IE} = m_{IC_2, IE} = .75;$			
				$b_{IE} = .20$			
				$c_{CC_1} = c_{CC_2} = .67;$			
				$m_{CC_1, CC_2, IE} = .75;$			
				$b_{CE} = .20$			
1	1	1	$[1 - (1 - m_{IC_1, IE})(1 - m_{IC_2, IE})(1 - b_{IE})] c_{IC_1} c_{IC_2}$	1	1	1	$[1 - (1 - m_{CC_1, CC_2, IE})(1 - b_{CE})] c_{CC_1} c_{CC_2}$
1	1	0	$(1 - m_{IC_1, IE})(1 - m_{IC_2, IE})(1 - b_{IE}) c_{IC_1} c_{IC_2}$	1	1	0	$(1 - m_{CC_1, CC_2, IE})(1 - b_{CE}) c_{CC_1} c_{CC_2}$
1	0	1	$[1 - (1 - m_{IC_1, IE})(1 - b_{IE})] c_{IC_1} (1 - c_{IC_2})$	1	0	1	$b_{CE} c_{CC_1} (1 - c_{CC_2})$
0	1	1	$[1 - (1 - m_{IC_2, IE})(1 - b_{IE})] (1 - c_{IC_1}) c_{IC_2}$	0	1	1	$b_{CE} (1 - c_{CC_1}) c_{CC_2}$
0	0	1	$b_{IE} (1 - c_{IC_1}) (1 - c_{IC_2})$	0	0	1	$b_{CE} (1 - c_{CC_1}) (1 - c_{CC_2})$
0	1	0	$(1 - m_{IC_2, IE})(1 - b_{IE})(1 - c_{IC_1}) c_{IC_2}$	0	1	0	$(1 - b_{CE})(1 - c_{CC_1}) c_{CC_2}$
1	0	0	$(1 - m_{IC_1, IE})(1 - b_{IE}) c_{IC_1} (1 - c_{IC_2})$	1	0	0	$(1 - b_{CE}) c_{CC_1} (1 - c_{CC_2})$
0	0	0	$(1 - b_{IE})(1 - c_{IC_1})(1 - c_{IC_2})$	0	0	0	$(1 - b_{CE})(1 - c_{CC_1})(1 - c_{CC_2})$

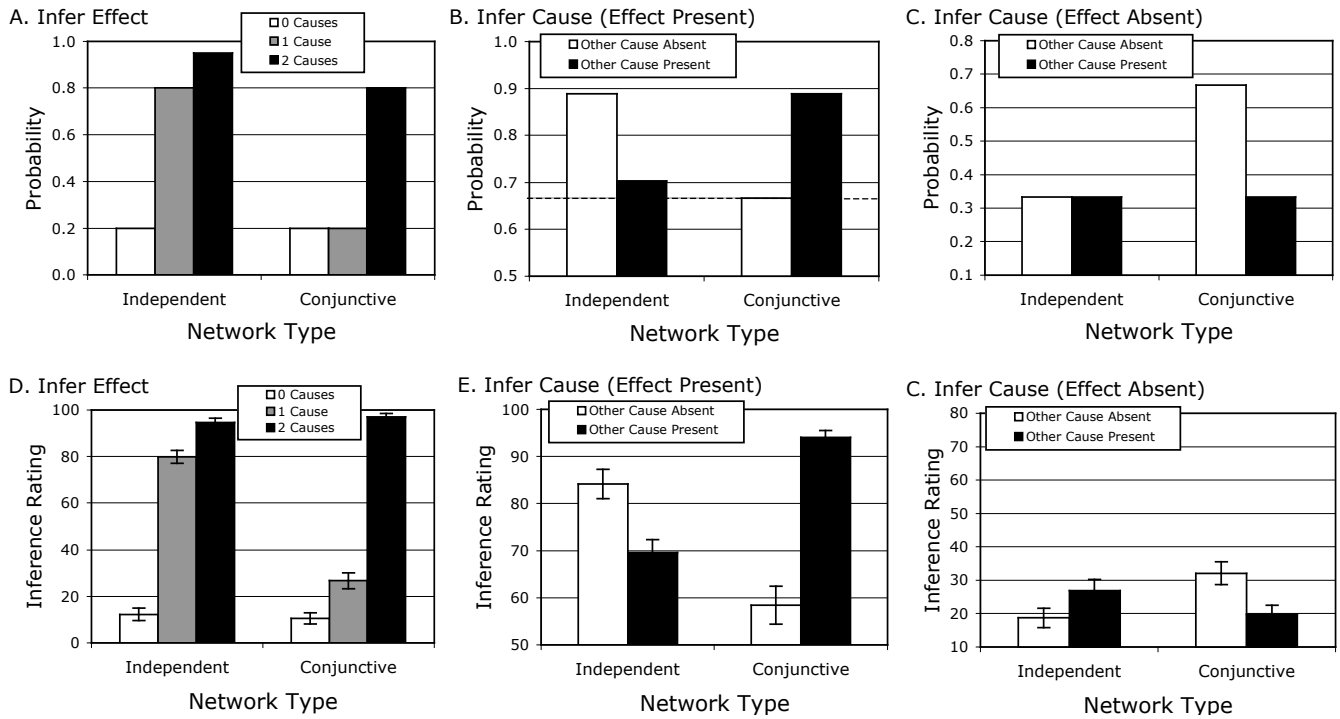


Figure 3. Panels A, B, and C present the predicted probability that a feature will be present as a function of the presence or absence of other features for the independent and conjunctive cause networks. (A) The probability of the effect as a function of the number of causes present. (B) The probability of a cause as a function of whether the other cause is present, assuming the effect is present. (C) The probability of a cause as a function of whether the other cause is present, assuming the effect is absent. The corresponding empirical results are presented in panels D, E, and F. Error bars are standard errors.

ing experiment. Because they are described as typical category features, each cause is assumed to be moderately prevalent among category members (the $cs = .67$), each causal mechanism is moderately strong (the $ms = .75$), and the alternative causes of the effect features are weak (the $bs = .20$). For example, given these parameter values, the probability of IC_2 conditioned on the presence of IE and the absence of IC_1 is,

$$\begin{aligned}
 p_k(IC_2 = 1 | IC_1 = 0, IE = 1) \\
 &= p_k(IC_2 = 1, IC_1 = 0, IE = 1) / p_k(IC_1 = 0, IE = 1) \\
 &= .178 / (.178 + .022) = .890
 \end{aligned}$$

Predictions for the independent and conjunctive causal networks for three distinct types of inference problems are shown in Figures 3A, 3B, and 3C. First, Figure 3A presents the probability of the effect as a function of the number of causes that are present for both the independent and conjunctive cause networks. For independent causes, the probability of the effect of course increases monotonically with the number of causes. (The probability of the effect is .20 even when both causes are absent because of the potential of additional causes, represented by $b_{IE} = .20$.) In contrast, for conjunctive causes, the probability of the effect increases from its baseline of .20 only when *both* causes are present.

Figure 3B presents inferences to a cause when the effect is present as a function of the state of the other cause. For an independent cause network, the probability of a cause is higher when the other cause is absent versus present. This is

the well-known *explaining away* phenomenon in which the presence of one cause of an effect makes other causes less likely. For example, the discovery of the murder weapon in a suspect's possession lowers the probable guilt of other suspects. Morris and Larrick (1995) have shown how explaining away is expected under a wide range of conditions.

The conjunctive cause network, in contrast, shows the opposite pattern, namely, the probability of a cause is *lower* when the other cause is absent versus present. For example, murder requires not only the motive but also the means, so discovering that a murder suspect didn't possess the means to carry out the crime (e.g., proximity to the victim) decreases his likely guilt. We refer to this as the *exoneration effect*. To our knowledge, this effect has not been noted by previous investigators.

Finally, Figure 3C presents the probability of a cause when the effect is absent as a function of the state of the other cause. On one hand, although independent causes are negatively correlated when the effect is present (the explaining away effect), they are uncorrelated when the effect is absent (and thus the probability of a cause is unaffected by state of the other cause). In contrast, the probability of a conjunctive cause is lower when the other cause is present. This represents another form of exoneration: Your brother, who promised to attend your Thanksgiving dinner this year but failed to arrive, is exonerated from responsibility (e.g., of returning an insincere RSVP) when you learn that his flight was canceled due to a snowstorm.

Table 2

Features	Causal Relationships
High amounts of carbon monoxide in the exhaust	High amounts of carbon monoxide in the exhaust causes a long-lived generator. The carbon monoxide increases the pressure of the exhaust that enters the turbocharger, resulting in the turbocharger drawing less electricity from the generator, extending its life. [Independent]
Damaged fan belt	A damaged fan belt causes a long-lived generator. When the damaged fan belt slips, the generator turns at lower RPMs, which means that it lasts longer. [Independent]
Long-lived generator	
Butane-laden fuel	Butane-laden fuel and loose fuel filter gaskets together cause a hot engine temperature. Loose fuel filter gaskets allow a small amount of fuel to seep into the engine bearings. This normally has no effect. However, if there is butane in the fuel, it undergoes a chemical reaction that creates heat as a byproduct. Thus, when a car has both butane-laden fuel and a loose filter gasket, the engine runs at a hot temperature. [Conjunctive]
Loose-fuel filter gaskets	
Hot engine temperature	

Overview of Experiment

The following experiment assesses whether people's inference judgments are consistent with the predictions just presented. As this is the first test of how people reason with conjunctive causes, our initial goal was to test whether subjects manifest the qualitative phenomenon that distinguish them from independent causes. Accordingly, subjects were not provided with values corresponding to the causal model parameters, that is, exact information about the probability of each cause (the *c* parameters), the strength of the causal links (the *m* parameters), or the possibility of alternative causes (the *b* parameters). Instead, we assess whether subjects exhibit, for example, explaining away for independent causes and exoneration for conjunctive causes.

Method

Materials. Six novel categories were tested: two biological kinds (Kehoe Ants, Lake Victoria Shrimp), two nonliving natural kinds (Myastars [a type of star], Meteoric Sodium Carbonate), and two artifacts (Romanian Rogos, Neptune Personal Computers). Each category had six binary feature dimensions. One value on each dimension was described as typical of the category. For example, participants who learned Romanian Rogos were told that "Most Rogos have a hot engine temperature whereas some have a normal engine temperature," "Most Myastars have high density whereas some have a low density," and so on.

Subjects were also provided with causal knowledge corresponding to the structures in Figure 2. Each independent causal relationship was described as one typical feature causing another, accompanied with one or two sentences describing the mechanism responsible for the causal relationship. Each conjunctive causal relationship was described as two features together causing a third. Table 2 presents an example of independent and conjunctive causes for Rogos.

The assignment of the six typical category features to the causal roles in Figure 2 (IC₁, IC₂, IE, CC₁, CC₂, and CE) was balanced over subjects such that for each category one triple of features played the role of IC₁, IC₂, and IE and the other played the role of CC₁, CC₂, and CE for half the subjects and this assignment was reversed for the other half. The features and causal relationships for all six categories are available from the authors.

Procedure. Participants first studied several computer screens of information about the category. Three initial screens presented the category's cover story and which fea-

tures occurred in "most" versus "some" category members. The fourth screen described the three causal relationships and the causal mechanisms. A fifth screen presented a diagram like that in Figure 2 (with the names of the category's actual features).

When ready, participants took a multiple-choice test that tested them on this knowledge. While taking the test, participants were free to return to the information screens; however, doing so obligated them to retake the test. The only way to proceed was to take the test all the way through without errors and without asking for help.

Subjects were then presented with classification and inference tests, counterbalanced for order. (The results of the classification test are not the topic of this article and are not discussed further.) During the inference test, participants were presented with a total of 24 inference problems, 12 for each subnetwork. They were asked to (a) predict the effect given all possible states of the causes (4 problems), predict each cause given all possible states of the effect and the other cause (8 problems). For example, participants who learned Rogos were asked to suppose that a Rogo had been found that had butane-laden fuel and a loose fuel filter gasket and to judge how likely it was that it also had a hot engine temperature. Responses were entered by positioning a slider on a scale where the left end was labeled "Sure that it doesn't" and the right end was labeled "Sure that it does" The position of the slider was scaled into the range 0-100. The order of presentation of the 24 test items was randomized for each participant. So that judgments did not depend on subjects' ability to remember the causal relations, they were provided with a printed diagram similar to the one in Figure 2. Subjects were asked to make use of those causal relations in answering the inference questions.

Participants. 48 New York University undergraduates received course credit for participating in this experiment. There were three between-subject factors: the two assignments of physical features to their causal roles, the two task presentation orders, and which of the six categories was learned. Participants were randomly assigned to these 2 x 2 x 3 = 12 between-participant cells subject to the constraint that an equal number appeared in each cell.

Results

Initial analyses revealed no effects of which category subjects learned, the assignment of features to causal roles, or feature presentation order, and so the results are presented collapsed over these factors.

Feature inference ratings are presented in Figures 3D, 3E, and 3F. These ratings generally reflected the predictions shown in Figures 3A-C. Unsurprisingly, for both independent and conjunctive causes, subjects judged that the presence of the effect was rated to be very likely (ratings > 90) when two causes were present and very unlikely (< 15) when they were absent. In addition however, Figure 3D shows that subjects were sensitive to the different functional relationships relating the effects to their causes. They were much more likely to predict the effect when one cause was present when the causes were independent (rating of 79.9) as compared to conjunctive (26.8).

Statistical analysis supported these conclusions. A 3 x 2 ANOVA of the data in Figure 3D revealed a main effect of the number of causes, $F(2, 94) = 470.36$, $MSE = 364$, $p < .0001$, a main effect of network type, $F(1, 47) = 65.91$, $MSE = 335$, $p < .0001$, and an interaction, $F(2, 94) = 82.62$, $MSE = 278$, $p < .0001$, reflecting how the networks differed in how inference ratings increased with the number of causes. In particular, when only one cause was present, ratings were much higher for the independent vs. conjunctive cause networks, $t(47) = 9.88$, $p < .0001$.

Second, Figure 3E shows inference ratings when subjects predicted a cause when the effect is present. When causes were independent, subjects exhibited explaining away: The cause was rated higher when the other cause was absent (84.2) versus present (69.6). In contrast, this pattern was reversed for conjunctive causes (58.4 vs. 94.0), that is, subjects exhibited exoneration. A 2 x 2 ANOVA of the data in Figure 3E revealed an effect of the state of the other cause, $F(1, 47) = 9.68$, $MSE = 549$, $p < .01$, no effect of network type, $F < 1$, but an interaction, $F(1, 47) = 83.62$, $MSE = 362$, $p < .0001$. Ratings were higher when the other cause was absent, $t(47) = 3.03$, $p < .01$, when causes were independent (explaining away) whereas they were lower when they were conjunctive, $t(47) = 9.27$, $p < .0001$ (exoneration).

Finally, Figure 3F shows that when reasoning about conjunctive causes subjects also exhibited exoneration when predicting a cause in the absence of an effect: The cause was rated higher when the other cause was absent (32.1 vs. 19.8 when present). The independent cause network, in contrast, did not show this pattern. Analyses of the data in Figure 3F revealed no main effects, both F s < 1, but an interaction, $F(1, 47) = 15.43$, $MSE = 326$, $p < .001$. In particular, ratings were higher when the other conjunctive cause was absent, $t(47) = 2.80$, $p < .01$ (exoneration).

These analyses confirm that subjects exhibited many of the key phenomena distinguish reasoning with independent vs. conjunctive causes. Nevertheless, Figures 3D-F also reveal a couple of ways in which the observed ratings differ from the predicted ones. First, the model predicted equal probabilities for certain inferences that in fact were rated differently by subjects. For example, for conjunctive causes, an effect should be equally probable regardless of whether zero or one cause is present (Figure 3A). In contrast, subjects judged that the effect was more probable in the presence of one cause (26.8) versus none (10.5), $t(47) = 5.25$, $p < .0001$. And, for independent causes, a cause should be independent of the state of the other cause when the effect is

absent (Figure 3C) but subjects judged it more likely when the other cause was present vs. absent (26.9 vs. 18.8), $t(47) = 2.30$, $p < .05$. These results are consistent with a *typicality effect* in which features are judged to be more probable when other typical features are present, even when those other features are (according to our model) independent of the feature being inferred. Rehder and Burnett (2005) found typicality effects for a large variety of causal networks. We discuss this result at greater length below.

Second, there are also signs that subjects were ignoring feature base rates. For example, for conjunctive causes, the probability of a cause when the effect and the other cause are both absent should be its base rate (in Figure 3C, .67). However, for this inference problem subjects produced a rating of only 32.1 on a 100-point scale. Although the inference ratings should not be directly interpreted as probabilities (because subjects were not explicitly told how the scale maps onto probabilities) a rating that is below the midpoint of the scale strongly suggests that subjects did not view the probability of the cause as corresponding to its base rate (which, because it was described as a typical category feature, should be > .50). That is, as in so many other studies, our subject appears to be exhibiting base rate neglect.

General Discussion

The first question asked in this research is whether human causal reasoners are sensitive to the different functional relationships that can tie an effect to its causes. The answer is that they are. Inferences differed sharply depending on whether causes were independent or conjunctive.

A second question was whether those inferences would exhibit the qualitative patterns predicted by a causal power representation of causal knowledge. The answer is that they (mostly) did. For example, when causes are independent, subjects should (and did) exhibit explaining away, that is, judged that a feature was less likely to be the cause of an effect when another cause was present. This result is unsurprising in light of the numerous demonstrations of explaining away in the social psychology (e.g., Morris & Larrick, 1995) and cognitive (e.g., Rehder & Burnett, 2005) literatures. In contrast, when causes are conjunctive, subjects should (and did) exhibit exoneration effects. To our knowledge, this article is the first to demonstrate both that exoneration effects are entailed by a causal power representation of conjunctions and that human causal reasoners in fact exhibit that effect.

Other frameworks for representing causal knowledge are unable to readily explain these results. For example, simple spreading activation networks are unable to account for the present result because such networks are insensitive to both the distinction between independent and conjunctive causes. Particularly troublesome for such networks are cases in which the presence of one variable *decreases* the probability of another. For example, in the context of learning, the phenomenon of explaining away is at the heart of CGM explanations of *backward blocking*, a phenomenon notorious for the difficulties it poses for associative learning theories (e.g., Sobel et al., 2004; although see Van Hamme & Waserman, 1994, for updated versions of associative models

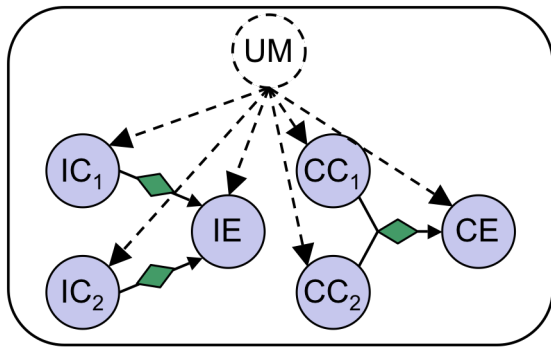


Figure 4.

that attempt to account such phenomena). Our demonstration of exoneration with conjunctive causes when the effect is absent (Figure 3F) presents an analogous challenge for associative accounts: the probability of a cause is lower when its conjunct is present.

As mentioned, the goal of this initial experiment was to assess whether subjects exhibited the qualitative pattern of inferences predicted by our model. In future work we intend to present more stringent test of the model by providing subjects with information corresponding to the parameters of the causal model, such as the strength of the causal relations (the m parameters) and alternative causes (the b parameters). Morris & Larrick (1995) systematically evaluated how subjects' inferences varied with these factors for independent but not conjunctive causes.

Nevertheless, even taking into account that we did not provide causal strength information, we found that subjects' ratings diverged from the predictions in one qualitative way: They exhibited a typicality effect in which inferences were stronger whenever more features were present, regardless of whether those features were independent of the feature being predicted. Rehder and Burnett (2005) also found pervasive typicality effects and proposed that people assume that categories possess underlying causal mechanisms that relate observable features (UM in Figure 4). The underlying mechanism provides an alternative inferential path such that the presence of one feature makes the operation of the category's normal mechanism more certain, which in turn increases the likelihood of other typical features (see Mayrhofer et al., 2010 for an alternative formalization). Our future work will also focus on whether such proposals provide a full account of reasoning with conjunctive causes. Yet another question is whether a typicality effect appears for inferences not involving features of categories.

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. *Psychological Review*, *104*, 367-405.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., & Kushnir, T. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 3-23.

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

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*, 56.

Holyoak, K. J., Lee, J. S., & Lu, H. (2010). Analogical and category-based inferences: A theoretical integration with Bayesian causal models. *JEP:General*, *139*, 702-727.

Lee, H. S., & Holyoak, K. J. (2008). The role of causal models in analogical inference. *JEP:LMC*, *34*, 1111-1122.

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

Mayrhofer, R., Hagmayer, Y., & Waldmann, M. R. (2010). Agents and Causes: A Bayesian Error Attribution Model of Causal Reasoning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, Austin, TX: Cognitive Science Society (pp. 925-930).

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

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review*, *111*, 455-485.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science*, *27*, 709-748.

Rehder, B. (2003b). A causal-model theory of conceptual representation and categorization. *JEP:LMC*, *29*, 1141-1159.

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

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *JEP:General*, *130*, 323-360.

Rehder, B. & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *JEP:LMC*, *32*, 659-683.

Rehder, B. & Kim, S. (2009). Classification as diagnostic reasoning. *Memory & Cognition*, *37*, 715-729.

Rehder, B. & Kim, S. (2010). Causal status and coherence in causal-based categorization. *JEP:LMC*, *36*, 1171-1206.

Shafto, P., Kemp, C., Bonawitz, E. B., Coley, J. D., & Tenenbaum, J. B. (2008). Inductive reasoning about causally transmitted properties. *Cognition*, *109*, 175-192.

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*, 189-228.

Sloman, S. A., & Lagnado, D. A. (2005). Do we "do"? *Cognitive Science*, *29*, 5-39.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*, 303-333.

Van Hamme, L., & Wasserman, E.A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*, 127-151.

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *JEP:General*, *124*, 181-206.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *JEP:LMC*, *31*, 216-227.