Explaining "Explaining Away"*

Michael P. Wellman
University of Michigan
Department of EECS
Ann Arbor, MI 48109
wellman@engin.umich.edu

Max Henrion†
Rockwell International Science Center
444 High St, #400
Palo Alto, CA 94301
henrion@sumex-aim.stanford.edu

March 16, 1994

Abstract

Explaining away is a common pattern of reasoning in which the confirmation of one cause of an observed or believed event reduces the need to invoke alternative causes. The opposite of explaining away also can occur, in which the confirmation of one cause *increases* belief in another. We provide a general qualitative probabilistic analysis of intercausal reasoning, and identify the property of the interaction among the causes, *product synergy*, that determines which form of reasoning is appropriate. Product synergy extends the qualitative probabilistic network (QPN) formalism to support qualitative intercausal inference about the directions of change in probabilistic belief. The intercausal relation also justifies Occam's razor, facilitating pruning in search for likely diagnoses.

Appeared as a correspondence in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(3):287-292, 1993.

[0]Portions of this paper originally appeared in *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning* [16].

†Supported by the National Science Foundation under grant IRI-8807061 to Carnegie Mellon and by the Rockwell International Science Center.

1 Explaining Away

Keeping track of the dependency or causal structure among events is critical in uncertain reasoning. One fundamental reason is the inherent asymmetry between *predictive* (or causal) reasoning, from cause to effect, and *diagnostic* (or evidential) reasoning, from effect to cause. Pearl [9] clearly illustrates this asymmetry with the "sprinkler" example, depicted in Figure 1. Either A, "it rained last night,"[1] or B, "the sprinkler was on last night," could cause C, "the grass is wet." C could in turn cause E, "the grass is cold and shiny," as well as F, "my shoes are wet."

Observation of one effect, E, cold and shiny grass, is evidence for C, wet grass, and hence predicts the other effect, F, wet shoes. Confirmation of one cause, A, rain, also leads to the expectation of C, wet grass. But it does *not* provide any evidence for the alternate cause B, sprinkling. Suppose prior observation of wet grass had led to defeasible acceptance of sprinkling. In a default reasoning scheme, confirmation of rain should lead to a *retraction* of the hypothesis that the sprinkler had been on. In a probabilistic reasoning scheme, it should lead to a *reduced probability* of the sprinkler hypothesis, even though the possibility of simultaneous sprinkling and rain is allowed.

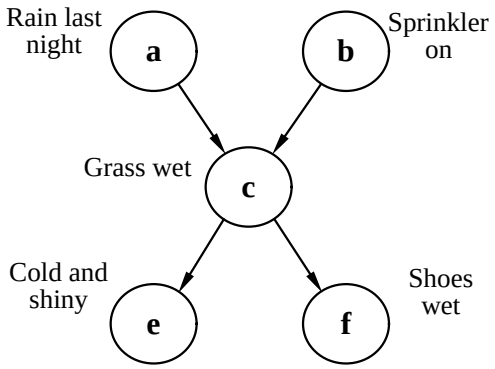

Figure 1: Causal diagram for the "sprinkler" example [9].

This common and intuitively compelling pattern of reasoning is called *explaining away*, because one cause explains the observed effect and so reduces the need to invoke other causes. This qualitative pattern of reasoning is entirely compatible with Bayesian inference when probabilistic influences reflect causal relationships [3, 9]. It is also the essence of Occam's razor: slice away hypotheses that are unnecessary to account for the evidence. Indeed, Paek [8] applies minimization of causal justifications to realize the explaining away pattern in a circumscriptive logic.

[1] By convention, uppercase letters denote propositional literals, while lowercase letters denote variables. Thus variable a, "rain last night," can take on the value A or its negation $\bar{A}$.

Pearl [9] uses the revealed asymmetry of inference with respect to causal direction to argue for incorporating causal relations in default reasoning schemes. Although inference rules implementing explaining away have been well studied [2, 9], precise and general conditions under which this pattern is valid have not appeared in the literature. Pearl provides these conditions for the special case of linear/Gaussian models [10, page 351]. Geffner provides a probabilistic justification of explaining away in terms of ϵ-semantics [1]. Both of these demonstrations are illustrative, but do not capture the full range of situations in which such inference is appropriate.

Explaining away is an example of *intercausal inference* [3]—that is, reasoning between two causes with a common effect—in contrast with pure causal or pure evidential reasoning. Although explaining away is often intuitively compelling, there are cases in which it appears inappropriate. Consider the following example, illustrated by the causal model of Figure 2. You notice this newspaper headline about a well-known politician: "Senator Jones Killed in Car Accident." You idly wonder whether she might have been drunk. The headline gives no indication of whether she was at fault in the accident, or even whether she was a driver or passenger. You had no previous information about her driving or drinking habits, but you know that alcohol is a major cause of fatal car accidents. Reading on, you find out that Jones was indeed the driver and that no other vehicle was involved in the accident. How does this new information affect your belief that she had been drinking? Without knowledge of any accident, the fact that the Senator was driving might *reduce* the suspicion that she had been drinking. Given the accident, however, the fact that she was the driver would *increase* the suspicion. Note that this pattern of plausible reasoning is the *opposite* of explaining away: Knowledge of a common effect renders a positive dependence between the causes, even though the causes were independent or even negatively dependent a priori.

The goal of this paper is to provide a general analysis of intercausal reasoning that accounts for both of the illustrated patterns of reasoning, and that makes precise the conditions differentiating them. Our choice of a probabilistic approach reflects the uncertainty central to causal explanation tasks, and is supported by the observation that explaining away is a natural consequence of some generic structures commonly employed in probabilistic modeling.

Although the probabilistic formulation refers to quantitative degrees of belief, it does not necessarily require precise numerical probabilities for application. Indeed, our analysis is qualitative, concerning the direction—but not the magnitude—of probabilistic dependencies. Our premise is that the critical distinctions correspond to intuitive categories of interaction among causes, and that further precision would be impractical or less convenient and, for many purposes, unnecessary. This position is supported by the observation that common vocabulary includes numerous qualitative concepts of causal interaction. For example, we often say that causal factors act inde-

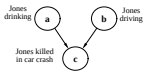

Figure 2: The drinking-and-driving example. Explaining away fails in this case because the two causes are positively related given their common effect.

pendently or synergistically, that one cause (a "gating condition") enables or inhibits another, or that a set of available inputs are complementary or substitutable with one another. Rain and sprinkling independently cause wet grass; drinking amplifies the causal relation between driving and car accidents.

We formalize these concepts using the qualitative probabilistic network (QPN) representation [15], an abstraction of Bayesian networks. The analysis of intercausal reasoning extends this formalism by introducing new qualitative characterizations of causal interactions, complementary with the existing QPN *synergy* relations.

In the remainder of this paper, we present a formal analysis of qualitative intercausal relations. After reviewing the notion of qualitative probabilistic influence in Section 2, in Section 3 we analyze intercausal reasoning with uncertain causal influences, and identify the conditions for explaining away to occur. We generalize these conditions in Section 4 to handle prior intercausal relationships and partial evidence on the effect. Finally, in Section 5 we present a view of Occam's razor suggested by intercausal relations.

2 Qualitative Probabilistic Networks

Our analysis of intercausal inference under uncertainty is based on the QPN formalism for qualitative probabilistic reasoning [15]. In a qualitative probabilistic network, variables are represented as nodes in a graph, with directed edges defining prob-

abilistic relationships. As in Bayesian networks [10] and other graphical schemes, connectedness in the graph represents the dependency structure of the underlying probability distribution [11]. However, rather than specify the distribution precisely with numeric probability tables, QPNs merely constrain the conditional probabilities using qualitative influences.

Associated with each edge is a sign, $\delta \in \{+, -, 0, ?\}$, denoting the direction of *qualitative influence* between nodes. Figure 3 depicts an example QPN representing beliefs about the health of a friend. Event A, that our friend has a cold, increases[2] the probability of C, that he is sneezing. Event B, that he has an allergic reaction, also increases this probability. On the other hand, event F, that he recently took an antihistamine, reduces the probability of sneezing. Event D, that our friend is allergic to cats, increases the probability of an allergic reaction, as does E, that a cat is present. (Whereas, for ease of exposition, the variables in our examples are binary, the definition of qualitative influence that follows, like most other definitions and theorems, applies equally to multivalent discrete and continuous variables.)

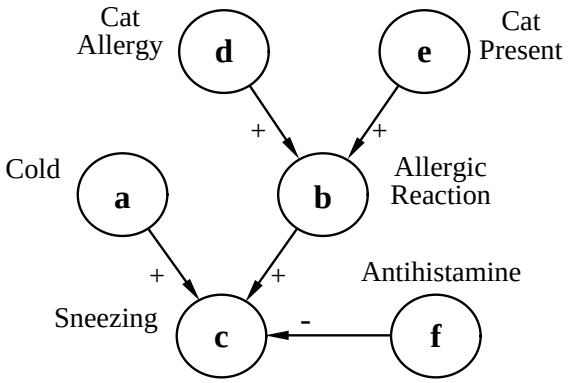

Figure 3: An example QPN representing beliefs about a friend's health. Arrows labeled "+" and "−" denote positive and negative causal influences, respectively.

For the general definition of qualitative influences, consider a QPN with a directed edge from a to c, and optionally some other variables, collectively denoted x, with links to c. In Figure 3, for example, x would comprise b and f. This structure dictates that the probability distribution for c can be specified conditionally on a and x.

Definition 1 (qualitative influence) *We say that* a positively influences c *in a QPN* G*, written* $S^+(a, c, G)$*, if and only if (iff) for all values* $a_1 > a_2$*,* c_0*, and all assignments* x *to other predecessors of* c *in* G*,*

$$\Pr(c \geq c_0 | a_1 x) \geq \Pr(c \geq c_0 | a_2 x).$$

[2] We use terms such as *increase* and *decrease* in the nonstrict sense, unless explicitly stated.

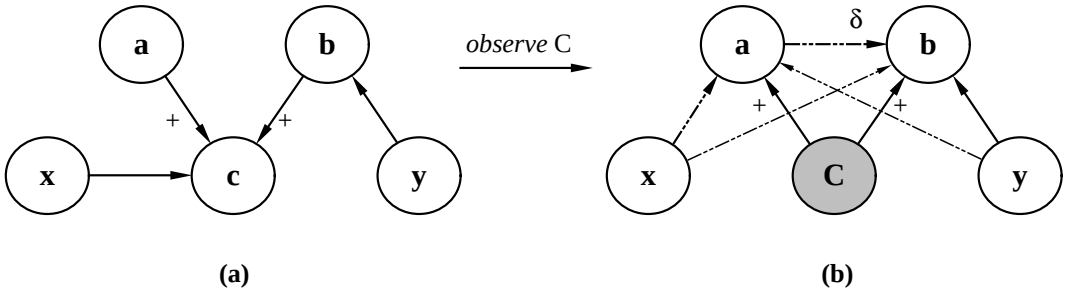

Figure 4: The schematic QPN transformation for intercausal inference. The qualitative influence, δ, of a on b upon observation of C indicates whether explaining away occurs.

An equivalent condition is that the probability density function (or mass function in the discrete case) for a given c and x, $f_a(\cdot|cx)$, obeys the *monotone likelihood-ratio property*:

$$\frac{f_a(a_1|c_1x)}{f_a(a_1|c_2x)} \geq \frac{f_a(a_2|c_1x)}{f_a(a_2|c_2x)}, \tag{1}$$

for all $a_1 > a_2$, $c_1 > c_2$, and x. This property ensures that increasing a increases the expected value of c.[3] We can replace $\geq$ in (1) by $\leq$ or $=$, yielding the conditions for negative influence, S^-, or zero influence, S^0, respectively. S^0 means a and c are independent for all values of x. Note that the condition S^+ requires the inequality (1) to hold for all assignments x, and similarly for S^- and S^0. Therefore, it is quite possible that none of the three conditions hold. The condition $S^?$ indicates that the qualitative influence is ambiguous or that it is not known which, if any, of the relations holds.

3 Probabilistic Intercausal Relations

Suppose we observe our friend sneezing, C, which raises the probability of his having a cold, A, and the probability of his having an allergic reaction, B. If we know that he is allergic to cats, D, then learning that a cat is present, E, lends confirmation of the allergic reaction B. This explains away the sneezing, and so reduces the probability of the cold, A.

This process of intercausal inference can be cast as transformation of a causal graph or QPN, illustrated in general form in Figure 4. Again, a and b are causes of c.

[3] In writing these ratios here and elsewhere, we assume that all conditional-probability terms are well defined and nonzero. These assumptions could be relaxed at the expense of explicatory complexity. For further discussion of these probabilistic inequalities, see [7, 15].

For generality, we allow that there may be other causes of c (collectively represented by x), and that a and b in turn may have causal antecedents (b's are collectively labeled y; a's do not figure in the example). Figure 4a depicts this initial situation. Note that because their only connecting path is via direct links to c, a and b are marginally independent, although they are conditionally dependent given c.

The basic explaining-away scenario starts with an observation of the effect variable to be explained, c. Suppose that c is propositional, and that the observed value is C. To represent observation in a probabilistic network, we instantiate the observed node and modify the dependency structure in the graph so that the nodes of interest become conditional on the observation. The evidence instantiation is tantamount to reversing the links from a and b to c [12], as shown in Figure 4b. The signs on the reversed links remain positive, indicating that observing C increases the probability of higher values of a and b. In addition, the reversals introduce a new *intercausal* link between a and b, accounting for the fact that the variables become dependent on observing C. The explaining-away pattern is characterized exactly by the negativity of this intercausal influence. For propositional a and b, the relation $S^-(a,b)$ in the graph of Figure 4b would mean that

$$\Pr(B|ACxy) \leq \Pr(B|Cxy) \leq \Pr(B|\bar{A}Cxy);$$

hence, belief in A decreases belief in B.

Even if we knew that the signs on the original links from a to c and b to c were positive, without further constraint the sign on this new intercausal link would be ambiguous [15]. The question is this: What condition on the causal combination of a and b would enable us to derive a negative intercausal influence on observing C?

Theorem 1 (explaining away) *Let a and b be predecessors of c in a QPN G, and let x denote an assignment to c's other predecessors, if any. Let $obs(c_0, G)$ denote the QPN obtained from G on observation of c_0. Suppose $S^0(a,b,G)$. Then $S^-(a,b,obs(c_0,G))$ iff for all $a_1 > a_2$, $b_1 > b_2$, and x,*

$$\frac{f_c(c_0|a_1b_1x)}{f_c(c_0|a_1b_2x)} \leq \frac{f_c(c_0|a_2b_1x)}{f_c(c_0|a_2b_2x)}. \tag{2}$$

This follows directly from Bayes's rule, reversing the dependence of c on b.[4]

Because it plays such a pivotal role in explaining away, we introduce terminology and notation for the intercausal relation (2).

[4]Complete proofs of this and other results are provided in the appendix. A propositional version of Theorem 1 appears in [6].

Definition 2 (product synergy) *Let a and b be predecessors of c in G, and let x denote an assignment to c's other predecessors, if any. Variables a and b exhibit* negative product synergy *with respect to a particular value c_0 of c in G, written $X^-(\{a,b\}, c_0, G)$, if, for all $a_1 > a_2$, $b_1 > b_2$, and x,*

$$f_c(c_0|a_1b_1x)f_c(c_0|a_2b_2x) \leq f_c(c_0|a_1b_2x)f_c(c_0|a_2b_1x). \tag{3}$$

Note that (3) is just the product form of (2). Thus, negative product synergy requires that the proportional increase in the probability of c_0 on raising b is smaller for higher values of a. Hence, the causal contribution of a given variable is greatest when that variable is the only active (high-valued) cause. It is this type of interaction that underlies explaining away.

We define *positive product synergy*, X^+, and *zero product synergy*, X^0, by substituting $\geq$ and $=$, respectively, for $\leq$ in (2). Theorem 1 is also valid with either "+" or "0" substituted for "−" in both the intercausal influence S^- and corresponding product synergy X^-. As for qualitative influences, the negative, zero, and positive product synergies are not exhaustive. The condition $X^?$ indicates that the product synergy is ambiguous or that it is not known which, if any, of the relations hold.

We illustrate the main result by reconsidering the two examples of explaining away. In Figure 1, there is a negative intercausal relation between rain, A, and the sprinkler, B, given their common effect, wet grass, C. Figure 3 displays a corresponding negative intercausal relation between the cold and allergic reaction given their common effect, sneezing. According to Theorem 1, this kind of relationship is appropriate if and only if we believe that negative product synergy holds in each of these cases. That is, our beliefs about the causal effects must satisfy:

$$\frac{\Pr(C|AB)}{\Pr(C|A\bar{B})} \leq \frac{\Pr(C|\bar{A}B)}{\Pr(C|\bar{A}\bar{B})}. \tag{4}$$

In words, the proportional increase in probability of C, wet grass, due to learning B, sprinkling, is smaller given A, rain, than given $\bar{A}$, no rain. Or, the proportional increase in probability of sneezing due to learning that our friend has an allergy is less given a cold than given no cold. Both of these conditions seem eminently plausible—given one cause is present, the incremental effect of the second cause is less than it would be if the first were absent.

If negative product synergy does not seem immediately compelling, one can also derive it as a generalization of the *leaky noisy*-OR [4, 10], a plausible model for either situation. The noisy-OR dictates that each of the two causes may be sufficient alone to cause the effect, and that the causal mechanisms are independent. The *leakiness* allows that, even if neither A nor B occurs, C may occur for another unspecified reason (a *leak*, L). It is easy to show that the leaky noisy-OR relation implies negative

product synergy with respect to the presence of the effect, and so leads to explaining away [16]. This result generalizes straightforwardly to cases with more than two causal variables. In contrast, *noisy*-NOR models—where causes lead to the *negation* of the effect—exhibit *zero* product synergy.

Now let us reconsider examples for which explaining away does not seem to apply. The drinking and driving Senator from Figure 2 is one such instance. The case from Figure 3 of the two causes of an allergic reaction is another. Given that an allergic reaction, B, is observed, knowledge that our friend is allergic to cats, D, would tend to increase the probability that a cat is present, E, and vice versa. There is a positive intercausal relationship between D and E, given B. According to the positive version of Theorem 1, this relationship holds iff positive product synergy applies—that is, iff

$$\frac{\Pr(B|DE)}{\Pr(B|D\bar{E})} \geq \frac{\Pr(B|\bar{D}E)}{\Pr(B|\bar{D}\bar{E})}. \tag{5}$$

For our example, this condition says that the proportional increase in probability of an allergic reaction, due to the cat being present, is greater given that our friend is allergic to cats, then it would be if he were not. This is evident, given that the cat would have only indirect effects, if any, if he were not allergic to cats. Therefore, the right-hand side of (5) would be at or near unity, whereas the left-hand side would be significantly larger.

4 Extensions

4.1 Dependent Causes

The premise of Theorem 1 requires that causes a and b be marginally independent. We can generalize the result, as long as any prior dependence between the causes is in the same direction as the intercausal effect of observing their common finding:

Theorem 2 $S^{\delta}(a, b, G) \wedge X^{\delta}(\{a, b\}, c_0, G) \Rightarrow S^{\delta}(a, b, obs(c_0, G))$.

For example, suppose we know our neighbor habitually listens to weather reports and turns off the sprinkler when rain is forecast. This negative prior relation between the two causes is in the same direction as the intercausal relation, and hence the tendency of the sprinkler to explain away the rain hypothesis is only strengthened.

On the other hand, suppose we believe in Murphy, the perverse raingod who likes to make it rain soon after a sprinkler has been used. This induces a positive prior dependence between the causes, rain and sprinkler. In this case the intercausal relationship after observing wet grass becomes ambiguous and cannot be determined by purely qualitative analysis.

4.2 Indirect Evidence

Theorem 1 also presumes that the effect variable, c, is observed directly. Can we generalize the main result to situations where we have only indirect evidence for c?

Suppose we observe the value of variable e, an effect of c. For example, in the sprinkler model, we might observe E, cold and shiny grass. To determine the intercausal implications of this observation, we investigate the interaction relation of a and b on e when c is factored out. This situation is depicted in Figure 5.

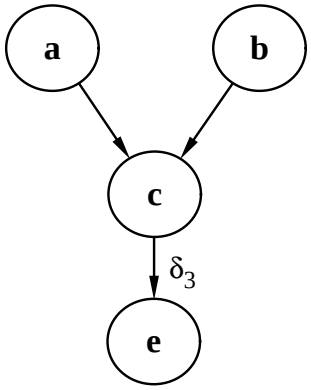

Figure 5: Two causes a and b, with partial evidence, e, for their common effect, c.

To propagate intercausal reasoning through indirect evidence, we appeal to another synergy concept, previously introduced for QPNs [15]:

Definition 3 (additive synergy) *Let a and b be predecessors of c in G, and let x denote an assignment to c's other predecessors, if any. Variables a and b exhibit* negative additive synergy *with respect to variable c in G, written $Y^-(\{a,b\},c,G)$, if, for all $a_1 > a_2$, $b_1 > b_2$, x, and c_0,*

$$\Pr(c \geq c_0|a_1b_1x) + \Pr(c \geq c_0|a_2b_2x) \leq \Pr(c \geq c_0|a_1b_2x) + \Pr(c \geq c_0|a_2b_1x).$$

Positive additive synergy, Y^+, and *zero additive synergy*, Y^0, are defined similarly, substituting $\geq$ and $=$, respectively, for $\leq$. An important difference between additive and product synergy is that the former is defined with respect to the *variable* c, rather than to a particular value c_0—that is, the additive synergy condition holds for all values of c. The disparity is due to the distinct roles of these relations in qualitative probabilistic inference. Note, however, that when c is a propositional variable, $Y^\delta(\{a,b\},c)$ is identical to $X^\delta(\{a,b\},C)$, except in substituting addition for multiplication in (3) (or differences for quotients in (2)). Although neither subsumes the other in general, when both of the individual influences of each cause on the effect

have unambiguous signs (+ or −), then there are entailment relationships between them. See [16] for a detailed exposition of these relationships.

The following result establishes (for the propositional case) that evidence positively related to the effect maintains intercausal relations given some particular patterns of product and additive synergy.

Theorem 3 *Let $red(c, G)$ denote the QPN obtained from G by reducing (averaging out) variable c. Suppose $X^{\delta_1}(\{a, b\}, C, G)$, $Y^{\delta_2}(\{a, b\}, c, G)$, $S^{\delta_3}(c, e, G)$, $S^0(a, e, G)$, and $S^0(b, e, G)$. Then $X^{\delta_1}(\{a, b\}, E, red(c, G))$ if either of the following:*

1. *$\delta_1 = \delta_2$ and $\delta_3 = +$.*
2. *$\delta_1 = -\delta_2$ and $\delta_3 = -$.*

Under certain circumstances, we can generalize Theorem 3 to the case of non-propositional c. In essence, product synergy extends from c_0 to e_0 as long as e_0 supports c_0 but does not distinguish among $c \neq c_0$.[5] For propositional c, it matters only whether the observed value e_0 was more likely given C than $\bar{C}$.

5 Occam's Razor and Intercausal Reasoning

Suppose that there are several causal hypotheses—each of which could explain an observed effect by itself—related to the finding according to a negative product synergy relationship. Given the negative intercausal relations between each pair of hypotheses given the finding, invoking one hypothesis reduces belief in the others. This process is analogous to the action of Occam's razor in slicing away hypotheses that are multiplied beyond necessity.

On the other hand, if two or more causes interact with a *positive* product synergy, their joint occurrence may be a more likely explanation of the finding than would be either alone. The synergistic effects of drinking and driving, and of cat allergies and cats are two examples. We might be tempted to invoke "Occam's glue" in such cases, as the multiple hypotheses adhere together to form a coherent scenario. But perhaps it is more appealing to regard the conjunctive relation as suggesting their combination as a single compound hypothesis. Seen in this light, they are not being multiplied beyond necessity, and so not actually contravening the principle of parsimony.

Note that when there are multiple evidence variables, positive intercausal relationships and complementary hypotheses can arise even when all synergy relations are negative. Consider the QPN in Figure 6a, where three diseases—represented by

[5] To establish this, we divide c into values for which X^{δ} holds (C) and those for which it does not ($\bar{C}$), then apply the previous theorem. In the process, we must be careful that the division does not invalidate the conditional independence of a and b from e given c.

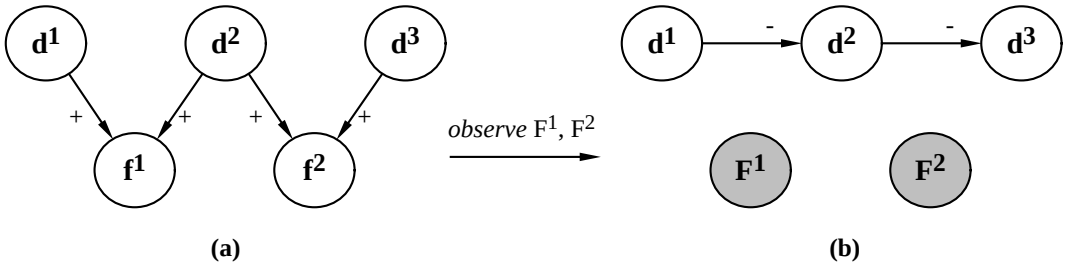

Figure 6: Multiple findings and complementary hypotheses. (a) A QPN with three diseases that can cause two findings. (b) Observing findings F^1 and F^2. Explaining away produces two negative intercausal influences, which can be chained to reveal a positive relation between d^1 and d^3.

propositional variables d^1, d^2, and d^3—can variously account for two findings, f^1 and f^2. Suppose that all influences are positive, and that the pairwise interactions satisfy negative product synergy. According to Theorem 1, given both findings F^1 and F^2, we obtain the two negative intercausal influences $S^-(d^1, d^2)$ and $S^-(d^2, d^3)$, depicted in Figure 6b. Chaining these, we can conclude $S^+(d^1, d^3)$ on removal of d^2, indicating that events D^1 and D^3 are complementary. This conclusion fits the intuitive observation that the findings can be explained either by the single disease D^2 or by the combination D^1 and D^3. If D^1 and D^3 are common diseases and D^2 is relatively rare, it is quite possible that the combination is more probable than the single disease. Thus, the intercausal analysis dictates how causal events should be clustered in compound hypotheses. Events complementary in the causal explanation become related by positive influences, without explicit set-covering computations.

Qualitative intercausal reasoning has also proven useful in the design of algorithms for quantitative probabilistic diagnosis. Because exact inference is intractable for large multiply-connected networks, there has been considerable interest in approximation algorithms. One such approach for diagnosis is to use heuristic search to find the most probable hypotheses that can explain the observed findings. In one large medical diagnosis application, called QMR-BN (Quick Medical Reference–Belief Network) [13], there are almost 600 diseases, and hence 2^{600} potential diagnoses. However, in most cases only a tiny fraction of these diagnoses have substantial probability. Search-based algorithms, such as TopN [5], concentrate on the most probable hypotheses. Given the relative probabilities of the candidate diagnoses, TopN computes bounds on their absolute probabilities. The bounds may be successively narrowed as the search continues.

The key to the design of efficient search-based algorithms is an *admissibility heuristic* that allows them to prune subtrees that can provably lead only to hypotheses whose probability is less than some threshold. The TopN algorithm starts out by ex-

amining single-disease hypotheses, extending them incrementally. Intercausal analysis can identify which additional diseases are complementary, and therefore can possibly lead to more probable hypotheses. It also reveals which diseases are competitive, and therefore can lead only to less probable hypotheses. Thus, intercausal analysis provides a suitable basis for an admissibility heuristic. Because QMR-BN uniformly assumes noisy-OR relations among diseases and findings, the diseases are always competitive. Initial results for this network, using this pruning criterion, show rapid convergence to narrow probability bounds in most cases [5]. The analysis described in this paper generalizes this approach to handle networks not only with noisy-OR relations, as in QMR-BN, but with any interactions satisfying negative product synergy.

6 Conclusions

Intercausal relations play a central role in the combination of diagnostic and predictive reasoning. The qualitatively significant property of interacting hypotheses is whether they compete with or complement one another in explaining the observed findings. In the former case, one cause explains away the other given the observation. In addition, we have shown that explaining away is not the only pattern of intercausal reasoning. To account for this distinction, we have derived a general probabilistic criterion, *negative product synergy*, that precisely justifies explaining away.

The main appeal of qualitative probabilistic relations is that they require minimal precision, yet capture some of the most significant behaviors. But qualitative probabilistic inference may be useful even for numerical systems, as a means of explanation to human users in a way that might correspond more directly to intuitive categories [6].

We also believe that it may be computationally advantageous to maintain these qualitative distinctions even when numeric information is available. As described in Section 5, intercausal relations qualitatively restrict the reasonable patterns in which to cluster events in compound hypotheses. These constraints can be exploited in diagnosis to prune the space of composite hypotheses at a high level, based on qualitative admissibility.

A Proofs

Theorem 1 *Let a and b be predecessors of c in a QPN G. Let $obs(c_0, G)$ denote the QPN obtained from G on observation of $c = c_0$. Suppose $S^0(a, b, G)$. Then $S^-(a, b, obs(c_0, G))$ iff $X^-(\{a, b\}, c_0, G)$.*

PROOF Let y denote the predecessors of b in G, and x the predecessors of c other

than a and b, if any. The distribution for a given b, x, and y on observation of c_0 is, by Bayes's rule,

$$f_a(a|bc_0xy) = \frac{f_c(c_0|abxy)f_a(a|bxy)}{f_c(c_0|bxy)}. \tag{6}$$

By conditional independence, we can drop the y condition from the f_c terms, and the x condition from the f_a term on the right-hand side. The qualitative influence of a on b is positive iff (6) obeys the monotone likelihood ratio property (1), and negative iff the inequality of (1) is reversed. Substituting (6) in the likelihood ratio for a_i given a pair of values for b, $b_1 > b_2$, we obtain

$$\frac{f_c(c_0|a_ib_1x)f_a(a_i|b_1y)f_c(c_0|b_2x)}{f_c(c_0|a_ib_2x)f_a(a_i|b_2y)f_c(c_0|b_1x)}. \tag{7}$$

Since $f_c(c_0|b_jx)$ does not depend on a_i, the ratio (7) is increasing or decreasing in a_i in direct correspondence with

$$\frac{f_c(c_0|a_ib_1x)f_a(a_i|b_1y)}{f_c(c_0|a_ib_2x)f_a(a_i|b_2y)}. \tag{8}$$

By the conditional independence of a and b given y (the S^0 condition), $f_a(a_i|b_1y) = f_a(a_i|b_2y)$, so these terms may be canceled from the expression, leaving

$$\frac{f_c(c_0|a_ib_1x)}{f_c(c_0|a_ib_2x)}.$$

The direction of change of this expression with respect to a_i is exactly the product synergy condition. □

Theorem 2 $S^\delta(a,b,G) \wedge X^\delta(\{a,b\},c_0,G) \Rightarrow S^\delta(a,b,obs(c_0,G))$.

PROOF Proceed as for Theorem 1, up to the reference to unconditional independence. The ratio (8) can be factored into two parts,

$$\left(\frac{f_c(c_0|a_ib_1x)}{f_c(c_0|a_ib_2x)}\right)\left(\frac{f_a(a_i|b_1y)}{f_a(a_i|b_2y)}\right).$$

The first part increases according to the sign of product synergy, the second contingent on the direct influence of a on b prior to observation of c_0. When the two agree, the direction of the entire expression is determined, establishing the qualitative influence of a on b posterior to the observation. □

Theorem 3 *Let $red(c,G)$ denote the QPN obtained from G by reducing (averaging out) variable c. Suppose $X^{\delta_1}(\{a,b\},C,G)$, $Y^{\delta_2}(\{a,b\},c,G)$, $S^{\delta_3}(c,e,G)$, $S^0(a,e,G)$, and $S^0(b,e,G)$. Then $X^{\delta_1}(\{a,b\},E,red(c,G))$ if either of the following:*

1. $\delta_1 = \delta_2$ *and* $\delta_3 = +$.

2. $\delta_1 = -\delta_2$ *and* $\delta_3 = -$.

PROOF Let $H_{i,j} = \Pr(E|a_i b_j x)$ and $G_{i,j} = \Pr(C|a_i b_j x)$. Since e is conditionally independent of a and b given c,

$$H_{i,j} = \Pr(E|C)G_{i,j} + \Pr(E|\bar{C})(1 - G_{i,j}).$$

Expanding terms and simplifying, the product of two H expressions is

$$H_{i,j}H_{k,l} = G_{i,j}G_{k,l}\Delta^2 + (G_{i,j} + G_{k,l})\Pr(E|\bar{C})\Delta,$$

where $\Delta = \left[\Pr(E|C) - \Pr(E|\bar{C})\right]$, which is positive or negative according to δ_3. Since Δ^2 is always positive, the comparison of a pair of H products is the same as for the corresponding G products if the comparison of second additive terms also agrees. When $\Delta > 0$, the sign of this second comparison is determined by the additive synergy relation, and when $\Delta < 0$, by its negation. □

Acknowledgments

We thank Marek Druzdzel for discussions on several aspects of this work, and Lyn Dupre for technical editing. Ramesh Patil and the anonymous reviewers provided useful suggestions on the presentation of this material.

References

[1] Hector Geffner. On the logic of defaults. In *Proceedings of the National Conference on Artificial Intelligence*, pages 449–454, St. Paul, MN, 1988. AAAI.

[2] Hector Geffner. Causal theories for nonmonotonic reasoning. In *Proceedings of the National Conference on Artificial Intelligence*, pages 524–530, Boston, MA, 1990. AAAI.

[3] Max Henrion. Uncertainty in artificial intelligence: Is probability epistemologically and heuristically adequate? In J. Mumpower et al., editors, *Expert Judgment and Expert Systems*, volume 35 of *NATO ISI Series F*, pages 105–130. Springer-Verlag, Berlin, 1987.

[4] Max Henrion. Some practical issues in constructing belief networks. In Laveen N. Kanal, Tod S. Levitt, and John F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*. North-Holland, Amsterdam, 1989.

[5] Max Henrion. Search-based methods to bound diagnostic probabilities in very large belief nets. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pages 142–150, Los Angeles, CA, 1991.

[6] Max Henrion and Marek J. Druzdzel. Qualitative propagation and scenario-based explanation of probabilistic reasoning. In Piero P. Bonissone, Max Henrion, Laveen N. Kanal, et al., editors, *Uncertainty in Artificial Intelligence 6*. North-Holland, Amsterdam, 1991.

[7] Paul R. Milgrom. Good news and bad news: Representation theorems and applications. *Bell Journal of Economics*, 12:380–391, 1981.

[8] Eunok Paek. A circumscriptive theory for causal and evidential support. In *Proceedings of the National Conference on Artificial Intelligence*, pages 545–549. AAAI, 1990.

[9] Judea Pearl. Embracing causality in default reasoning. *Artificial Intelligence*, 35:259–271, 1988.

[10] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

[11] Judea Pearl, Dan Geiger, and Thomas Verma. Conditional independence and its representations. *Kybernetika*, 25:33–44, 1989.

[12] Ross D. Shachter. Evidence absorption and propagation through evidence reversals. In *Proceedings of the Workshop on Uncertainty in Artificial Intelligence*, pages 303–310, Windsor, ON, 1989.

[13] M. Shwe, B. Middleton, D. E. Heckerman, et al. Probabilistic diagnosis using a reformulation of the Internist-1/QMR knowledge base: I. The probabilistic model and inference algorithms. *Methods of Information in Medicine*, 30:241–255, 1991.

[14] Michael P. Wellman. *Formulation of Tradeoffs in Planning Under Uncertainty*. Pitman, London, 1990.

[15] Michael P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44:257–303, 1990.

[16] Michael P. Wellman and Max Henrion. Qualitative intercausal relations, or Explaining "explaining away". In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 535–546, 1991.