The Causal Interpretation of Bayesian Networks

Chapter · September 2008

DOI: 10.1007/978-3-540-85066-3_4

Kevin Korb
Monash University (Australia)

Ann E. Nicholson
Monash University (Australia)

The Causal Interpretation of Bayesian Networks

Kevin B. Korb and Ann E. Nicholson

Clayton School of Information Technology
Monash University
Clayton, Victoria 3800, Australia
{kevin.korb,ann.nicholson}@infotech.monash.edu.au

Summary. The common interpretation of Bayesian networks is that they are vehicles for representing probability distributions, in a graphical form supportive of human understanding and with computational mechanisms supportive of probabilistic reasoning (updating). But the interpretation of Bayesian networks assumed by causal discovery algorithms is causal: the links in the graphs specifically represent direct causal connections between variables. However, there is some tension between these two interpretations. The philosophy of probabilistic causation posits a particular connection between the two, namely that causal relations of certain kinds give rise to probabilistic relations of certain kinds. Causal discovery algorithms take advantage of this kind of connection by ruling out some Bayesian networks given observational data not supported by the posited probability-causality relation. But the discovered (remaining) Bayesian networks are then specifically causal, and not simply arbitrary representations of probability.

There are multiple contentious issues underlying any causal interpretation of Bayesian networks. We will address the following questions:

- Since Bayesian net construction rules allow the construction of multiple distinct networks to represent the very same probability distribution, how can we come to prefer any specific one as “the” causal network?
- Since Bayesian nets within a Verma-Pearl pattern are strongly indistinguishable, how can causal discovery ever come to select exactly one network as “the” causal network?
- Causal discovery assumes faithfulness (that d-connections in the model are accompanied by probabilistic dependency in the system modeled). However, some physical systems cannot be modeled faithfully under a causal interpretation. How can causal discovery cope with that?

Here we introduce a causal interpretation of Bayesian networks by way of answering these questions and then apply this interpretation to answering further questions about causal power, explanation and responsibility.

Keywords: causal discovery, faithfulness, Bayesian networks, probabilistic causality, intervention, causal power, causal responsibility.

1 Introduction

In the last decade Bayesian networks have risen to prominence as the preferred technology for probabilistic reasoning in artificial intelligence, with a proliferation of techniques for fast and approximate updating and also for their automated discovery from data (causal discovery, or “data mining” of Bayesian networks). Philosophers of science have also begun to adopt the technology for reasoning about causality and methodology (e.g., [2]; [23]). Both the causal discovery and the philosophical analysis depend upon the propriety of a causal interpretation of the Bayesian nets in use. However, the standard semantics for Bayesian networks are purely probabilistic (see, e.g., [39]), and various of their properties — such as the statistical indistinguishability of distinct Bayesian networks [53, 7] — seem to undermine any causal interpretation. As a result, skeptics of causal interpretation are growing along with the technology itself.

2 Bayesian Networks

We begin with a perfectly orthodox (we hope) introduction to the concepts and notation used in Bayesian networks (for a more detailed introduction see [28, Chap 2]); readers familiar with these are advised to skip to the next section. A **Bayesian network** is a directed acyclic graph (dag), M, over a set of variables with associated conditional probabilities θ which together represent a probability distribution over the joint states of the variables. The fully parameterized model will be designated $M(\theta)$. For a simple (unparameterized) example, see Figure 1. *Rain* and *Sprinkler* are the **root** nodes (equivalently, the **exogenous** variables), which report whether there is rain overnight and whether the automatic sprinkler system comes on. The **endogenous** (non-root) variables are *Lawn*, which describes whether or not the lawn is wet, and *Newspaper* and *Carpet*, which respectively describe the resultant sogginess and muddiness when the dog retrieves the newspaper. Note that we shall vary between talk of variables and their values and talk of **event types** (e.g., rain) and their corresponding **token events** (e.g., last night’s rain) without much ado.

Probabilistic reasoning is computationally intractable (specifically, NP-hard; [10]). The substantial advantage Bayesian networks offer for probabilistic reasoning is that, if the probability distribution can be represented with a sparse network (i.e., with few arcs), the computations become practicable. And, most networks of actual interest to us are sparse. Given a sparse network, the probablistic implications of observations of a subset of the variables can be readily computed using any of a large number of available algorithms [28]. In order for the computational savings afforded by low arc density to be realized, the lack of an arc between two variables must be reflected in a probabilistic independence in the system being modeled. Thus, in a simple

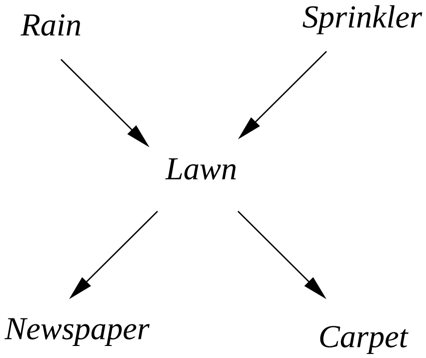

Fig. 1. A simple Bayesian network.

two-variable model with nodes X and Y, a missing arc implies that X and Y are probabilistically independent. If they are not, then the Bayesian network simply fails to be an appropriate model. Thus, in our example, *Rain* and *Sprinkler* must be independent of each other; if the sprinkler system is turned off in rainy weather, then Figure 1 is simply the wrong model (which could be made right by then adding *Rain* $\rightarrow$ *Sprinkler*).

X and Y being probabilistically independent just means that $P(X = x_i|Y = y_j) = P(X = x_i)$ for any two states x_i and y_j. **Conditional independence** generalizes this to cases where observations of a third variable may induce an independence between the first two variables, which may otherwise be dependent. Philosophers, following [43], have tended to call this relationship **screening off**. For example, *Rain* and *Newspaper* are presumably dependent in Figure 1; however, if we hold fixed the state of the lawn — say, we already know it is wet — then they are no longer probabilistically related: $P(\mathit{Newspaper}|\mathit{Lawn}, \mathit{Rain}) = P(\mathit{Newspaper}|\mathit{Lawn})$, which we will commonly abbreviate as $\mathit{Newspaper} \perp\!\!\!\perp \mathit{Rain}|\mathit{Lawn}$. Given these and like facts for other variables, Figure 1 is said to have the **Markov property**: that is, all of the conditional independencies implied by the Bayesian network are true of the actual system (or, equivalently, it is said to be an **independence map (I-map)** of the system). We shall often assume that our Bayesian networks are Markov, since, as we have indicated, when they are not, this is simply because we have chosen an incorrect model for our problem.

In the opposite condition, where all apparent dependencies in the network are realized in the system, the network is said to be **faithful** to the system (or, the network is called a **dependence-map (D-map)** of the system). A network which both satisfies the Markov property and is faithful is said to be a **perfect map** of the system.[1] There is no general requirement for a Bayesian

[1] The concept of "faithfulness" comes from the "faithfulness condition" of Spirtes, et al. [47]; they also, somewhat confusingly, talk about graphs and distributions being "faithful to each other", when we prefer to talk about perfect maps.

network to be faithful in order to be considered an adequate probabilistic model. In particular, arcs can always be added which do nothing — they can be parameterized so that no additional probabilistic influence between variables is implied. Of course, there is a computational cost to doing so, but there is no misrepresentation of the probability distribution. What we are normally interested in, however, are I-maps that are **minimal**: i.e., I-maps such that if any arc is deleted, the model is no longer an I-map for the system of interest. A minimal I-map need not necessarily also be a perfect map, although they typically are; in particular, there are some systems which have multiple distinct minimal I-maps.

We shall often be interested in probabilistic dependencies carried by particular paths across a network. A **path** is a sequence of nodes which can be visited by traversing arcs in the model (disregarding their directions) in which no node is visited twice. A **directed path** proceeds entirely in the direction of the arcs traversed. A fundamental graph-theoretic concept is that of two nodes X and Y being **d-separated** by a set of nodes $\mathbf{Z}$, which we will abbreviate $X \perp Y | \mathbf{Z}$. Formally,

Definition 1 (D-separation)
X and Y are **d-separated** *given $\mathbf{Z}$ (for any subset $\mathbf{Z}$ of variables not including X or Y) if and only if each distinct path Φ between them is cut by one of the graph-theoretic conditions:*

1. *Φ contains a chain $X_1 \longrightarrow X_2 \longrightarrow X_3$ and $X_2 \in \mathbf{Z}$.*
2. *Φ contains a common causal structure $X_1 \longleftarrow X_2 \longrightarrow X_3$ and $X_2 \in \mathbf{Z}$.*
3. *Φ contains a common effect structure $X_1 \longrightarrow X_2 \longleftarrow X_3$ (i.e., an* **uncovered collision** *with X_1 and X_3 not directly connected) and neither X_2 nor any descendant of X_2 is in $\mathbf{Z}$.*

This is readily generalized to sets of variables $\mathbf{X}$ and $\mathbf{Y}$.

The idea of d-separation is simply that dependencies can be cut by observing intermediate variables (1) or common causes (2), on the one hand, and induced by observing common effects (or their descendants), on the other (3). As for the latter, if we assume as above that the automated sprinkler and rain are independent in Figure 1, they will not remain so if we presuppose knowledge of the state of the lawn. For example, if the lawn is wet, something must explain that, so if we learn that there was no rain overnight, we must increase our belief that the sprinkler system came on.

A more formal version of the **Markov property** is then: a model has the Markov property relative to a system if the system has a conditional independence corresponding to every d-separation in the model. I.e.,

$$\forall X, Y, \mathbf{Z} \;\; (X \perp Y | \mathbf{Z} \Rightarrow X \perp\!\!\!\perp Y | \mathbf{Z})$$

The opposite condition to d-separation is **d-connection**, when some path between X and Y is *not* blocked by $\mathbf{Z}$, which we will write $X \not\perp Y | \mathbf{Z}$. Faithfulness of a graph is equivalent to

$$\forall X, Y, \mathbf{Z} \ (X \not\perp Y|\mathbf{Z} \Rightarrow X \not\perp\!\!\!\perp Y|\mathbf{Z})$$

A concept related to d-connection, but not the same, is that of an active path. We use active paths to consider the probabilistic impact of observations of some variables upon others: a path between X and Y is an **active path** in case an observation of X can induce a change in the probability distribution of Y. The concepts of d-connected paths and active paths are not equivalent because a d-connected path may be inactive due only to its parameterization.

3 Are Bayesian Networks Causal Models?

Such are Bayesian networks and some of their associated concepts. The illustrative network of Figure 1 is itself clearly a causal model: its arcs identify direct causal relationships between event types corresponding to its nodes — rain *causes* lawns to get wet, etc. But it is clear that there is no necessity in this. We could represent the very same probabilistic facts with a very different network, one which does not respect the causal story, reversing some of the arcs. Indeed, Chickering [7] introduced a transformation rule which allows us to reverse any number of the arcs in a Bayesian network:

Rule 1 (Chickering's Transformation Rule) *The transformation of M to M', where $M = M'$ except that, for some variables C and E, $C \longrightarrow E \in M$ and $C \longleftarrow E \in M'$ (and excepting any arc introduced below), will allow the probability distribution induced by M to be represented via M' so long as:*

> *if any uncovered collision is introduced or eliminated, then a covering arc is added (e.g., if $A \longrightarrow C \longrightarrow E \in M$ then A and E must be directly connected in M').*

Given any probability distribution, and any causal Bayesian network representing it, we can use Chickering's rule to find other, *anti-causal*, Bayesian networks. Clearly, this rule only *introduces* arcs and never eliminates any. Thus, when starting with a causal model and applying a sequence of Chickering transformations to find additional models capable of representing the original probability distribution, we can only start from a simpler model and reach (monotonically) ever more complex models. For example, we can apply this rule to our *Sprinkler* model in order to find that the alternative of Figure 2 can represent the probabilities just as well. Except that, this model clearly does not represent the probabilities *just as well* — it is far more, needlessly, complex. Since, in general, parameter complexity is exponential in the number of arcs, this complexification is highly undesirable computationally. And the new complexity introduced by Chickering transformations fail to represent the very same causal structure, and under a causal interpretation they will imply falsehoods about the consequences of interventions on its variables. The model arrived at in Figure 2, for example, has a wet carpet watering the lawn.

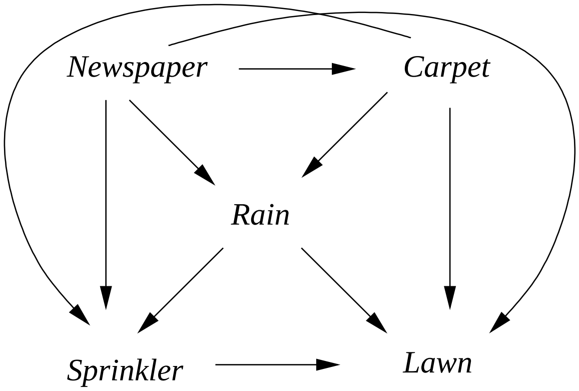

Fig. 2. A less simple Bayesian network.

The models in a sequence of Chickering transformations progress monotonically from simpler to more complex; the set of probability distributions representable by those models (by re-parameterizing them) form a sequence of monotonically increasing supersets, with the final model capable of representing all of (and usually more than) the distributions representable by its predecessors.

Intuitively, we might come to a preference for causal models then on simplicity grounds: causal models are the simplest of Bayesian networks capable of representing the probabilistic facts (such nets we will call **admissible models**), whereas the transformed, non-causal networks are inherently more complex. While we believe that this is largely true, and a sufficient motivation to prefer causal over non-causal models, it is not universally true. There are true causal models which are more complex than alternative admissible models, although the latter cannot be arrived at by Chickering's rule. We shall see an example of that below. Our real reason for preferring causal models to their probabilistic imitators is that most of our interest in Bayesian networks arises from an interest in understanding the stochastic universe around us, which comes in the form of numerous (partially) independent causal systems. We generate models to explain, predict and manipulate these systems, which models map directly onto those systems — that is what it means to have a true (causal) theory of any such system. Alternative models which arise from arbitrary redirections of causal arcs may have some computational interest, but they have little or no interest in scientific explanation or engineering application. The causal models are metaphysically and epistemologically primary; the alternative models arise from mathematical games.

So, to answer the question heading this section: most Bayesian networks are not causal models; however, most *interesting* Bayesian networks are.

4 Causal Discovery and Statistical Indistinguishability

Causal discovery algorithms aim to find the causal model responsible for generating some available data, usually in the form of a set of joint observations over the variables being modeled. Since 1990, when the first general discovery algorithms were invented, many more have been developed. Causal discovery algorithms include the original IC [53], PC [47], K2 [11], BDe/BGe [21], GES [8], and CaMML [54].

Automated causal discovery is analogous to scientific discovery. It is typical for scientific discovery to aim at finding some unique theory to explain some data, only to find instead some larger set of theories, all of which can equally well explain the available data. Philosophers of science call this the problem of **underdetermination**: any finite set of data can be explained by some large group of theories; adding some new data may well rule out a subset of the theories, while the set of theories remaining consistent with the expanded data is still large, perhaps infinitely large. Karl Popper [41] based his theory of scientific method on this observation, calling it Falsificationism: rather than verifying theories as true, he suggested that science progresses by falsifying theories that fail to accommodate new data; science progresses in a way vaguely similar to evolution, with unfit theories dying out and multiple fit theories remaining.[2]

Causal discovery can proceed in much the same way. We can begin with some hypothetical set of dags (the model space) that could provide causal explanations of reality and the real probability distribution, P_R, to be explained.[3] The discovery process then repeatedly finds probabilistic dependencies in P_R, crossing off all of those dags which cannot model the dependencies. In a simplified form, the original Verma-Pearl IC algorithm can be expressed as:

1. **Step I** Put an undirected link between any two variables X and Y if and only if

   for every set of variables $\mathbf{S}$ s.t. $X, Y \notin \mathbf{S}$

   $$X \not\perp\!\!\!\perp Y | \mathbf{S}$$

   I.e., X and Y are directly connected if and only if they are always conditionally dependent.
2. **Step II** For every undirected structure $X - Z - Y$ (where X and Y are not themselves directly connected) orient the arcs $X \rightarrow Z \leftarrow Y$ if and only if

   $$X \not\perp\!\!\!\perp Y | \mathbf{S}$$

[2] In later years Popper made this analogy more explicit, developing what he called evolutionary epistemology [42].

[3] Of course, we usually have access to P_R only through statistical samples; we are simplifying the description here.

for **every** $\mathbf{S}$ s.t. $X, Y \notin \mathbf{S}$ and $Z \in \mathbf{S}$.

> I.e., we have an uncovered collision if and only if the ends are always conditionally dependent upon the middle.

Following Steps I and II, a **Step III** checks for arc directions forced by further considerations, such as avoiding the introduction of cycles and uncovered collisions not revealed by P_R. Collectively, these steps suffice to determine what Verma and Pearl called a **pattern**, namely, a set of dags all of which share the same skeleton (arc structure, disregarding orientations) and uncovered collisions.

The IC algorithm is both the first causal discovery algorithm and the simplest to understand; it is not, however, practical in that it relies upon direct access to P_R via some oracle. Spirtes et al. [47] redesigned it to be more practical, in their PC algorithm, by finding some algorithmic efficiencies and, more importantly, by replacing the oracle with statistical significance tests for dependencies. Because of its simplicity, PC is now widely available, for example, in the machine learning toolbox, Weka [55]. An alternative approach to the discovery problem is to generate a global Bayesian or information-theoretic score for candidate models and then to search the model space attempting to optimize the score. Examples of this are the K2, BDe/BGe and CaMML algorithms cited above. The main difference between these two approaches, generally dubbed constraint-based and metric discovery respectively, is that the constraint learners test for dependencies in isolation, whereas the metric learners score the models based upon their overall ability to represent a pattern of dependencies.

Considerations arising from the Verma-Pearl algorithm led directly to a theory of the observational equivalence, or statistical indistinguishability, of models. It turns out that, given the assumption of a dependency oracle, the IC algorithm is in some sense optimal: Verma and Pearl proved that, if we restrict ourselves to observational data, no algorithm can improve upon its power to use the oracle's answers to determine causal structure. We can formalize (strong) statistical indistinguishability so:

Definition 2 (Strong Indistinguishability)

$$\forall \theta_1 \exists \theta_2 [P_{M_1(\theta_1)} = P_{M_2(\theta_2)}]$$
$$\textit{and vice versa}$$

That is, any probability distribution representable by M_1, via some parameterization θ_1, is also representable by M_2 via some other parameterization, and vice versa. What Verma and Pearl [53] proved then is that two models are strongly indistinguishable if and only if they are in the same pattern. The set of patterns over the model space form a partition, with patterns being equivalence classes of dags. In consequence of these results, most researchers have focused upon the learning of patterns (or, equivalently, Markov equivalence classes), dismissing the idea of discovering the causal models themselves as quixotic. Thus, for example, the GES [8] stands for "Greedy Equivalence Search"; it explicitly eschews any attempt to determine which dag within the best equivalence class of dags might be the true causal model.

By strong indistinguishability, the dags within each equivalence class exactly share the set of probability distributions which they are capable of representing. So, in other words, in the hands of most researchers, causal discovery as been turned from the search for causal models into, once again, the weaker and easier search for probability distributions. Of course, not attempting the impossible is generally good advice, so perhaps these researchers are simply to be commended for their good sense. If the problem of underdetermination is unsolvable, there is no point in attempting to solve it.

On the other hand, we might consider what human scientists do when observational data alone fail to help us, when multiple theories clash, but *not* over the observational data. Suppose, for example, we were confronted with a largely isolated system that could equally well be one of these two causal models (which are in a common pattern):

$$\begin{array}{l} \textit{Smoking} \rightarrow \textit{Cancer} \\ \textit{Smoking} \leftarrow \textit{Cancer} \end{array}$$

Underdetermination may halt scientific discovery, but not so near to the starting point as this! Commonly, of course, we would have more to go on than joint observations of these two variables. For example, we would know that *Smoking* generally precedes *Cancer* rather than the other way around. Or, we would expand our observations, looking, say, for mutagenic mechanisms articulating smoking and cancer. But, what would we do if we had no such background knowledge and our technology allowed for no such mechanistic search? We would *not* simply throw up our hands in despair and describe both hypotheses as equally good. We would experiment by intervening upon *Smoking* (disregarding ethical or practical issues).

Merely finding representations for our probability distributions does not exhaust our scientific ambitions, not by a long way. Probabilities suffice for prediction in the light of evidence (probabilistic updating); but if we wish to understand and explain our physical systems, or predict the impact of interventions upon them, nothing short of the causal model itself will do.

5 A Loss of Faith

Before exploring the value of interventional data for the discovery process, we shall consider another difficulty arising for causal modeling and for causal discovery in particular, namely a lack of faith. This has become one of the central arguments against the causal discovery research program.

The probabilistic causality research program in philosophy of science aims to find probabilistic criteria for causal claims [50, 46, 52]. The underlying intuition is that, contrary to much of the philosophical tradition, causal processes are fundamentally probabilistic, rather than deterministic. The probabilistic dependencies we measure by collecting sample observations over related variables are produced not merely by our ignorance but also, in some cases, directly by the causal processes under study. A variety of powerful arguments have been brought in support of this approach to understanding causality. Perhaps the most basic point is that our philosophy of science must at least *allow* the world to be indeterministic. The question is, after all, synthetic, rather than analytic: indeterminism is logically possible, so deciding whether our world is indeterministic requires empirical inquiry beyond any philosophizing we may do. Purely analytic philosophical argument cannot establish the truth of determinism. Turning to empirical means to resolve the question, we will immediately notice that our best fundamental theory of physics, quantum physics, is indeterministic, at least in its most direct interpretation.

If, then, causal processes give rise to probabilistic structure, we should be able to learn what causal processes are active in the world by an inverse inference from the probabilistic structures to the underlying causal structure. In other words, the probabilistic causality theory underwrites the causal discovery research program. This is one reason why the causal discovery of Bayesian networks has attracted the attention of philosophers of science, both supporters and detractors of probabilistic causality.

Patrick Suppes' account of probabilistic causation begins with what he calls prima facie causation [50]. C is a **prima facie cause** of E if the event type of C is positively related to the event type of E (ignoring an additional temporal precedence requirement); i.e., those C such that $P(C|E) - P(C) > 0$. We are then to filter out spurious causes, which are any that can be screened off by a common ancestor of the prima facie cause and the purported effect. What remains are the genuine causes of E. The essence of this is the identification of potential causes by way of probabilistic dependence.[4] Of course, this is also how the IC algorithm operates: both Step I and Step II posit causal structure to explain observed dependencies. And every other method of automated causal discovery also takes probabilistic dependence as its starting point; and they all assume that once all probabilistic dependencies have been causally explained, then there is no more work for causality to do. In other words, they assume the model to be discovered is *faithful*: corresponding to every d-connection in a causal model there must be a probabilistic dependence.

[4] Suppes, as do many other advocates of probabilistic causality, directs attention to causes which *raise* the probability of their effects and away from those which *lower* that probability — or, as Hausman [20] puts it, to the positive (promoting) *causal role* of the candidate cause. We are not much concerned with causal roles (whether promotion or inhibition) here, as such matters are primarily aimed at accounting for ordinary language behavior.

This assumption is precisely what is wrong with causal discovery, according to Nancy Cartwright [4] and other skeptics.

In the simplest case, where $C \longrightarrow E$ is the *only* causal process posited by a model, it is hard to see how a presupposition of faithfulness can be contested. Such a simple causal process which left no probabilistic reflection would be as supernatural as a probabilistic reflection produced by nothing at all. In any case, insisting upon the possibility of an unfaithful structure between understanding and reality *here* leaves inexplicable the ability to infer the causal structure in any circumstance.

But there are many situations where we should and must prefer an unfaithful model. One kind of unfaithfulness is where transitivity of probabilistic dependence fails. If causality is somehow based upon the kind of causal processes investigated by Salmon [45] and Dowe [12], processes which are capable of carrying information from one space-time region to another (Salmon-Dowe processes for short), then it seems somehow causality *ought* to be transitive. Salmon-Dowe processes are clearly composable: when ball A strikes B and B strikes C, the subprocesses composed form a larger process from A to C. No doubt this kind of Newtonian example lies behind the widespread intuition that causality must be transitive. We share the intuition that causal processes are somehow foundational for causal structure, and that they can be composed transitively; unfortunately for any simple analysis, causal structure itself is not transitive. One of many examples from Hitchcock [23] will make this clear: suppose there is a hiker on a mountain side and at just the wrong time a boulder dislodges and comes flying towards her; however, observing the boulder, she ducks at the right moment, and the boulder sails harmlessly past; the hiker survives. This is represented graphically in Figure 3. We are to suppose that if the boulder dislodges, the hiker will duck and survive, and that if the boulder doesn't dislodge, she will again survive. In this case, there is no sense in which the boulder makes a difference to survival. It would be perverse to say that the boulder has caused the hiker to survive, or to generalize and assert that in this and relevantly similar cases falling boulders cause hikers to survive. While causal processes, and their inherent transitivity, are one part of the story of causality, making a difference, probabilistic dependence, is equally part of that story, a part which here fails dramatically.

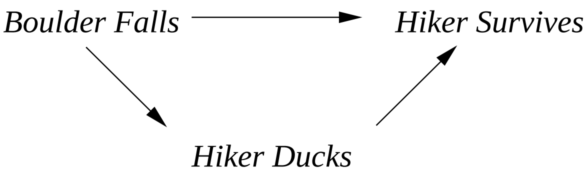

Fig. 3. The hiker surviving.

Hiddleston [22] claims that intransitivity in all cases can be attributed to the fact that there are multiple causal paths; when we look at component causal effects in isolation, such things cannot happen. He is mistaken. An example of Richard Neapolitan [36] makes this clear: finesteride reduces DHT (a kind of testosterone) levels in rats; and low DHT can cause erectile dysfunction. However, finesteride doesn't reduce DHT levels sufficiently for erectile dysfunction to ensue (in at least one study); in other words, there is a threshold above which variations in DHT have no effect on dysfunction. Graphically, this is simply represented by:

Finesteride → *DHT* → *Dysfunction*

Since, there is no probabilistic dependency between finesteride and erectile dysfunction, we have a failure of transitivity in a simple chain. We can equally have failures of transitivity in simple collisions [38].[5]

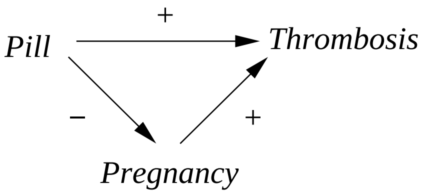

Fig. 4. Neutral Hesslow.

Another kind of unfaithfulness does indeed arise by multiple paths, where *individual arcs* can fail the faithfulness test. These include "Simpson's paradox" type cases, where two variables are directly related, but also indirectly related through a third variable. The difficulty is most easily seen in linear models, but generalizes to discrete models. Take a linear version of Hesslow's example of the relation between the *Pill*, *Pregnancy* and *Thrombosis* (Figure 4). In particular, suppose that the causal strengths along the two paths from *Pill* to *Thrombosis* exactly balance, so that there is no net correlation between *Pill* and *Thrombosis.* And yet, the above model, by stipulation, is the

[5] Hiddleston's mistake lies in an overly-simple account of causal power, for in linear models, and the small generalization thereof that Hiddleston addresses, causality is indeed transitive. Such models are incapable of representing threshold effects, as is required for the finesteride case.

We note also that some would claim that all cases of intransitive causation will be eliminated in some future state of scientific understanding: by further investigation of the causal mechanisms, and consequent increase in detail in the causal model, all apparently intransitive causal chains will turn into (sets of) transitive causal chains. This may or may not be so. Regardless, it is an a posteriori claim, and one which a theory of causality should not presuppose.

true causal model. Well, in that case we have a failure of faithfulness, since we have a direct causal arc from *Pill* to *Thrombosis* without any correlation wanting to be explained by it. In fact, causal discovery algorithms in this case will generally not return Figure 4, but rather the simpler model (assuming no temporal information is provided!) of Figure 5. This simpler model has all and only the probabilistic dependencies of the original, given the scenario described.

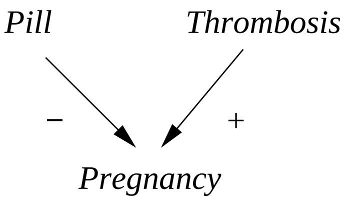

Fig. 5. Faithful Hesslow.

A plausible response to this kind of example, the response of Spirtes et al. [48], is to point out that it depends upon a precise parameterization of the model. If the parameters were even slightly different, a non-zero correlation would result, and faithfulness would be saved. In measure theory (which provides the set-theoretic foundations for probability theory) such a circumstance is described as having *measure zero* — with the implication that the probability of this circumstance arising randomly is zero (see [48], Theorem 3.2). Accepting the possibility of zero-probability Simpson-type cases implies simply that we *can* go awry, that causal discovery is fallible. But no advocate of causal discovery can reasonably be construed as having claimed infallibility.[6] The question at issue cannot be the metaphysical claim that faithfulness *must* always be maintained — otherwise, how could we ever come to admit that it had been violated? Rather, it is a methodological proposal that, until we have good reason to come to doubt that we can find a faithful model, we should assume that we can. In the thrombosis case, with precisely counterbalancing paths, we begin with knowledge that the true model is not faithful, so we are obliged to abandon faithfulness.

Cartwright objects to this kind of saving maneuver. She claims that the "measure zero" cases are far more common than this argument suggests. In particular, she points out that many systems we wish to understand are artificial, rather than natural, and that in many of these we specifically want to cancel out deleterious effects. In such cases we can anticipate that the cancel-

[6] Cartwright notwithstanding: "Bayes-net methods. . . will bootstrap from facts about dependencies and independencies to causal hypotheses—and, claim the advocates, *never get it wrong*" [4, p. 254] (italics ours). Here, Cartwright's strawman has it wrong.

ing out will be done by introducing third variables associated with both cause and effect, and so introducing *by design* a “measure-zero” case. In addition, there are many natural cases involving negative feedback where we might expect an equilibrium to be reached in which an approximate probabilistic independency is achieved. For example, suppose that in some community the use of sun screen is observed to be unrelated to skin cancer. Yet the possible causal explanation that sun screen is simply ineffective may be implausible. A more likely causal explanation could be that there is a feedback process such that the people using the sun screen expose themselves to more sunlight, since their skin takes longer to burn. If the people modulate their use of sun screen according to their exposure to the sun, then their total UV exposure would remain the same. Again, Steel [49] has pointed out that there are many cases of biological redundancy in DNA, such that if the allele at one locus is mutated, the genetic character will still be expressed due to a backup allele; in all such cases the mutation and the genetic expression will fail the faithfulness test. As Steel emphasizes, the point of all these cases is that the measure-zero premise fails to imply the probability zero conclusion: the system parameters have not been generated “at random” but as a result of intelligent or evolutionary design, leading to unfaithfulness.

If these cases posed some insurmountable burden for causal discovery algorithms, this would surely justify skepticism about automating causal discovery, for clearly we humans have no insurmountable difficulties in learning that the sun causes skin cancer, etc., even if these relations are also not easy to learn. However, it turns out there are equally clear, if again practically difficult, means available for machines to discover these same facts.

6 Intervention

So, it is time to see what interventions can do for causal discovery.

The concept of causal intervention and its uses in making sense of causal models have been receiving more attention recently, as, for example, in [56]. Indeed, both Pearl [40] and Spirtes et al. [48] treat intervention in some detail and provide a foundation for much of our work, and yet they have not applied interventions to resolve the problems raised by faithlessness and statistical indistinguishability. We now indicate how these are resolvable using interventions, first dealing with loss of faith. In empirical science, when observational data fail to differentiate between competing hypotheses, a likely response is to go beyond observation and experimentally intervene in nature: if we hold fixed known alternative causes of cancer and apply and withold a candidate carcinogen to experimental and control groups respectively, we can be in a position to resolve the issue of what the true causal model is, whether or not it is faithful to observable dependency structures. Intervention and experiment would seem to have the power to resolve our conflict between truth, on the one hand, and simpler, faithless models, on the other.

In order to explore the epistemological power of intervention, we will consider a particular kind of intervention, with some ideal features. Much of the literature idealizes every possible feature (e.g., the do-calculus [40], and the manipulation theorem of [48]): interventions are themselves uncaused (they are root nodes in a causal model), and so multiple interventions are uncorrelated; interventions impact upon exactly one variable in the original model; interventions are deterministic, definitely resulting in the intervened upon variable adopting a unique state. Such extreme idealization is not a promising starting point for a *general* theory of causal modeling, and the actual interventions available to us often fall well short of the ideal (cf. [27]). For our purposes here, however, we shall adopt all of them except the last: by default our interventions influence their target variables, but do not to the extreme of cutting off all influence of their prior parents (but, should they do so, we shall indicate this by calling the interventions *perfect*).[7] The extreme, perfect interventions can be represented in Bayesian networks simply by setting the target variable to the desired state and cutting all of its inbound arcs. Our less perfect interventions cannot be so represented; existing parents retain their arcs to the target variable. So, the natural representation of our interventions is by augmenting our dags with new intervention variables.

It is worth noting, as an aside, that this approach makes clear the difference between intervention and observation. Since most existing tools *only* provide explicit means for probabilistic updating under observation,[8] some have attempted to understand the effect of causal interventions upon a variable X simply by setting X to some desired value in a Bayesian network and updating. Such updates, however, can radically mis-estimate the effects of interventions. For example, consider Figure 6. *HT40* describes blood pressure (hypertension) at age 40, *HT50* the same at age 50 and *CHD50+* the coronary status of a subject from ages 50 to 60. We may be contemplating an intervention to reduce blood pressure at age 50, which may well reduce the chances of a heart attack over then next ten years. If, however, we were hypothetically to *observe* the same reduced level of blood pressure at age 50 as we might achieve by intervention, the resulting risk of heart attack would be even lower. Technically, that is because there are two d-connecting paths from *HT50* to *CHD50+*, whereas with a perfect intervention the link from *HT40* to *HT50* is cut, so there is only one path from *HT50* to *CHD50+*.[9] What this means non-technically is that should we observe someone at 50 with lower blood pressure, that implies the she or he also had lower blood pressure at age 40, which, entirely independently of blood pressure at 50, has an effect on *CHD50+*; whereas if we *intervene* at 50 we are not thereby gaining any

[7] We will assume that our interventions have *some* positive probability of affecting the target variable; indeed, we shall assume the same for every parent variable.

[8] For a description of a tool, *The Causal Reckoner*, that does more, see [27].

[9] And with an imperfect intervention, although the link from *HT40* to *HT50* is not cut, the influence of *HT40* on *CHD50+* is reduced.

new information about *HT40*. In short, in order to determine the effects of intervention, simply using observations and updating with standard Bayesian net tools is radically wrong.

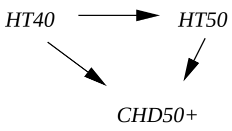

Fig. 6. Hypertension and coronary heart disease.

So, we prefer to represent interventions quite explicitly, by putting new intervention variables into our graphs. Furthermore, in order to test the limits of what we can learn from intervention we consider *fully* augmented models, meaning those where *every* original variable X gets a new intervention parent I_X, doubling the number of variables. In the case of the two Hesslow models, faithless (true) and faithful (false), full augmentation results in Figure 7.

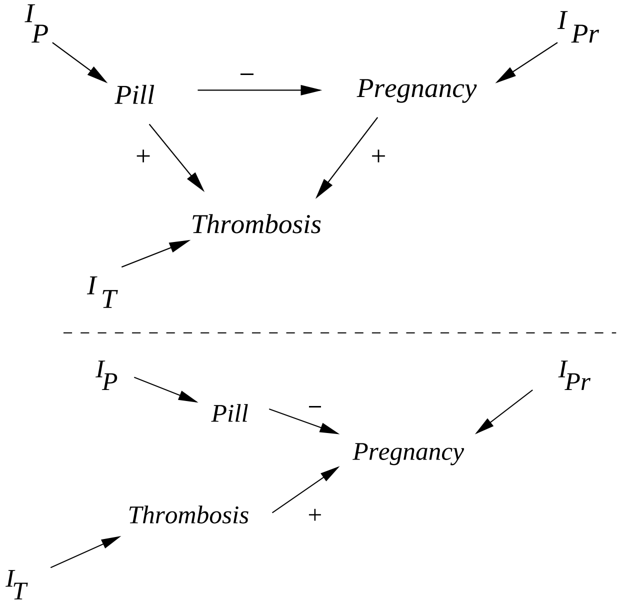

Fig. 7. Augmented Hesslow models: the faithless, but true, causal model (top); the faithful, but false, model (bottom).

What we suggest, then, is that the argument over whether faithfulness is an acceptable assumption for causal discovery is simply misdirected, ignoring the power of interventions. Faithfulness is not the issue; the real issue is **admissibility under augmentation**: A causal model M is admissible under augmentation if and only if its fully augmented model M' is capable of representing the system's fully augmented probability distribution.

Spirtes, Glymour and Scheines (SGS) [48] proved a theorem about augmentation which already suggests the value of interventions for distinguishing between causal models:

SGS Theorem 4.6 *No two distinct causal models that are strongly indistinguishable remain so under intervention.*[10]

This has a trivial corollary:

Corollary 1. *Under intervention, no two distinct causal models are strongly indistinguishable.*

These results tell us that when we augment models, even strongly indistinguishable models, we can then find *some* probability distribution that will distinguish them. Unfortunately, this does not provide much practical guidance. In particular, it tells us nothing about the value of intervention when we are dealing with a specific physical system and its *actual* probability distribution, as we typically are in science. But we can also apply intervention and augmentation to answering questions about particular physical systems rather than the set of all possible physical systems.

First, we can show that the neutral Hesslow system M_1 of Figure 4 can be distinguished from its faithless imposter M_2 of Figure 5. Under these circumstances, the probability distributions asserted by these models are identical; i.e., $P_{M_1(\theta_1)} = P_{M_2(\theta_2)}$. Then:

Theorem 2 (Distinguishability under Imperfect Intervention; [30]). *If $P_{M_1(\theta_1)} = P_{M_2(\theta_2)}$, then under imperfect interventions $P_{M_1'(\theta_1)} \neq P_{M_2'(\theta_2)}$ (where M_1 and M_2 are the respective structures of Figures 4 and 5).*

The proof is in [30] and is related to that of SGS Theorem 4.6 (we also proved there the same result for perfect interventions). The notable difference is that we begin with a particular probability distribution, that of the true Hesslow model, and a particular inability to distinguish two models using that distribution.[11] Using Wright's path modeling rules [57] we are able to find the

[10] Note that where we write of augmenting models with intervention variables, Sprites et al. [48] talk about "rigid indistinguishability", which amounts to the same thing.

[11] In SGS terminology, this is an application of the idea of rigid distinguishability to models which are weakly indistinguishable, that is, having some probability distribution which they can both represent.

needed difference between the augmented models, when there is none in their unaugmented originals. This difference arises from the introduction of new uncovered collisions under augmentation: every intervention variable induces a collision with its target variable and any pre-existing parent variable. As the Verma-Pearl algorithm is sensitive to such dependency structures (and likewise every other causal discovery algorithm), the dependency structures of M_1 and M_2, while parameterized to be identical under observation, cannot be so parameterized under intervention.

Although this theorem is particular to cases isomorphic to the neutral (and linear) Hesslow case, the result is of much wider interest than that suggests, since the neutral Hesslow structure is the only way in which a true linear causal model can be unfaithful to the probability distribution which it generates, through balancing multiple causal paths of influence. Perhaps of more interest will be interventions upon discrete causal models, which are more commonly the center of attention in causal discovery and which certainly introduce more complexity than do linear models. In particular, in non-linear cases there are many more opportunities for intransitivities to arise, both across multiple paths and, unlike linear models, across isolated paths. Nevertheless, similar theoretical results are available for discrete causal models, although, as their description and proof are more complicated, we skip them here (see instead [38]).

But the power of intervention to reveal causal structure goes well beyond these preliminary theorems [29]. It turns out that a comprehensive regime of interventions has the power to eliminate all but the true causal model from consideration. In other words, despite the fact that strong indistinguishability can leave a very large number of causal models equally capable of accommodating observational data, interventional data can guarantee the elimination of all models but one, the truth (this is the combined effect of Theorems 5 and 6 of [29]). Again, these results are generalizable to discrete models, with some additional restrictions upon the interventions needed to cope with the more complex environment [38]. Current research, at both Carnegie Mellon University and Monash University, is aimed at determining optimal approaches to gathering interventional data to aid the causal discovery process. None of these results are really surprising: they are, after all, inspired by the observation that human scientists dig beneath the readily available observations by similar means.

Neither the presumption of faithfulness by causal discovery algorithms nor their inability to penetrate beneath the surface of strong statistical indistinguishability offer any reason to dismiss the causal interpretation of Bayesian networks nor their automated discovery. The common view to the contrary is plausible only when ignoring the possibility of extending causal modeling and causal discovery by intervening in physical systems. Even if such interventions are expensive, or even presently technologically impossible, that is no principled reason for rejecting causal interpretation. The history of human science is replete with examples of competing theories remaining indistinguishable for centuries, or millenia, before technological advances have decisively found in favor of one over another. Witness Copernican and Ptolemaic astronomy, continental drift versus static geological theory, evolution theory versus static species theory. All of these are decisively settled today. In no case was the truth of no interest prior to the technological advances which allowed their settlement: on the contrary, interest in settling the issues has typically driven those very technological advances.

7 Causal Explanation

So far, we find that the causal interpretation of Bayesian nets is a reasonable one. At least, the arguments thrown up against it have failed to show otherwise, which is less than a clear justification for adopting a causal interpretation, but more than nothing. A principled justification would perhaps rely upon a serious exploration of the metaphysics of causation. We shall not attempt that, although in the next sections we will direct interested readers to some of the relevant current discussions in the philosophy of causation. A pragmatic justification is more to our taste: having dispelled the known objections, the success of causal modeling and causal discovery in developing an application technology for AI that is usable and used, and which presupposes the causal interpretation, should settle the argument. Norsys Corp., developer of the Bayesian network tool *Netica*, lists a considerable variety of real-world applications:

- *Agricultural Yield* for predicting the results of agricultural interventions [3]
- *PROCAM* model for predicting coronary heart disease risk [51]
- *Wildlife Viability* for predicting species viability in response to resource management decisions [34]
- *Risk Assessment Fusion* combining multiple expert opinions into one risk assessment [1]

There is much more work that can be done in making sense of Bayesian networks in causal terms. One such effort is to provide an account of the explanatory power of a cause for some effect. For example, to what extent can we attribute the current prevalence of lung cancer to smoking? This question about type causality can also be particularized in a question about token causality, as in: Was Rolah McCabe's terminal lung cancer due to her smoking cigarettes produced by British American Tobacco? We defer such questions about token causation to the section §9.

The type causality question is an old issue in the philosophy of science; what makes it also a new issue is the prospect of tying such an account to Bayesian networks, so that we might have computer assisted causal reasoning. Such an attempt is already on offer in the work of Cheng and Glymour [6, 15]. They define a concept of *causal power* for binomial variables, which, in the case of causes which promote their effects (rather than inhibit them, which takes a different equation), is:

$$p_c = \frac{P(e|c) - P(e|\neg c)}{1 - P(e|\neg c)}$$

The numerator corresponds to the probabilistic dependence used as a criterion by Suppes of prima facie causation. Cheng and Glymour get the effect of Suppes' filtering out of spurious cases of causation by imposing structural restrictions on the Bayesian networks which source the probabilities required to define p_c. So, the result is a causal power measure for genuine causation which is proportional to positive probabilistic dependence and inversely proportional to probability of the effect when the cause is absent.

This is a plausible start on explanatory power in Bayesian networks; unfortunately there is no finish. Cheng and Glymour's definition is restricted to binomial networks with parent variables which fail to interact (as, for example, an XOR interacts).[12] Furthermore, as Glymour [14] notes, the restrictions they impose on allowable Bayesian networks are equivalent to requiring them to be noisy-OR networks. The result is that for these, and also for the extension of Hiddleston [22], causal power is *transitive*, whereas we have already seen in Section §5 that probabilistic dependence over causal links (and, hence, a key ingredient for any useful definition of causal power) is *intransitive*. A different account is wanted.

We prefer to identify causal power with an information-theoretic measure related to mutual information [25]. The mutual information of C for E (or vice versa) is a measure of how much you learn about one variable when the other variable is observed. Hope and Korb's causal power differs from mutual information per se in a number of ways. First, it is asymmetric; it is required that it be attributed to the cause, rather than the effect, according to the ancestral relations indicated by the causal model. Also, the relevant probabilities are relativized to a context of interest. That is, any causal question is raised in some context and a measure of causal power needs to be sensitive to that context. For example, the causal power of smoking for lung cancer may be fairly high in some populations, but it is non-existent amongst those who already suffer from lung cancer. Finally, the causal power of C is measured according to a hypothetical perfect intervention upon C and not based upon mutual information computed by observations of C. Thus, if Sir Ronald Fisher's [13] speculative defence of smoking were actually true, i.e., that the true model were

$\mathit{Smoking} \longleftarrow \mathit{Gene} \longrightarrow \mathit{Cancer}$

rather than

$\mathit{Smoking} \longrightarrow \mathit{Cancer}$

[12] However, Novick and Cheng [37] relax this restriction to some extent by considering pairwise parental interactions.

then the causal power of *Smoking* for *Cancer* would be nil, whereas the mutual information between the two is unaffected by the change of structure.

This measure of causal power takes full advantage of Bayesian networks: all relevant interactions between causal variables are automatically taken into account, since the computation of the information-theoretic measure depends upon the underlying computations of the Bayesian network. As a result, for example, all cases of intransitivity found by Hitchcock and others yield end-to-end causal powers of zero, as is desirable.

This causal power theory also shows considerable promise for answering a pressing practical need in the application of Bayesian network technology, namely making the networks easier to interpret. By providing a means to query any such network about the implications of proposed interventions (of any type) it is no longer necessary for users themselves to follow and account for causal influences across multiple paths. Furthermore, by separating questions of causal power from those of probabilistic updating we can reduce the temptation to attempt to find causal explanations using tools which only answer questions about probabilistic updating, confusing any causal story rather than illuminating it.

The concept of causal power is a necessary ingredient for a full account of type causality, i.e., causal relations between types of events, rather than particular evens (token causality, treated in section §9). In some sense, causal Bayesian networks without any variables instantiated provide as full an account of the type causal relations between its variables as could be imagined (assuming they are true models of reality, of course). However, there remains analytical work to do beyond such an observation. In Hesslow's neutral model, for example, does the *Pill* cause *Thrombosis*? The net effect — the net causal power — is, as we have noted, nil. However, type causal questions are typically aimed at determining whether there is *some* way (consistent with any explicitly provided context) for the cause to have an impact on the effect. In order to answer such a question, (type) causal paths need to be considered in isolation, for example the *Pill* → *Thrombosis* path isolated from the path through *Pregnancy*, by fixing the value of the latter variable [23].[13] The type causal question can be answered affirmatively if any such isolated path has a non-zero causal power. Of course, we are commonly interested also in knowing *how important* a cause may be: for that we need the non-zero causal power itself.

[13] In addition to that, type causal questions need to be relativized to an objectively homogeneous context, so that the causal power being computed is *not* an average of disparate powers in distinct contexts, as Cartwright [5] has effectively argued (see also [52]).

8 Causal Processes

Now we will consider how Bayesian networks can help us make sense of token causality: claims about the particular responsibility of particular causal happenings for particular outcomes. We begin by sketching some relevant philosophical background, for we believe the philosophy of causal processes is now needed.

There are two distinct approaches which in recent times have dominated attempts to come to grips with the notion of causality. One is the probabilistic causality research program, already introduced. The other is the attempt to locate a supervenience base for causal relationships in an underlying metaphysics of process, initiated by Salmon [45] and furthered by Dowe [12]. Processes are contiguous regions of space extended through some time interval — i.e., spacetime "worms". Of course, they can't be just any such slice of spacetime; most such slices are causal junk [26]. The Salmon-Dowe research program is largely aimed at coming up with clear criteria that rule out junk, but rule in processes which can sustain causal relationships. Intuitively, we can say legitimate processes are those which can carry information from one spacetime region to another ("mark transmission" is what Salmon called this; Dowe calls it "conserving physical quantities"). Examples are ordinary objects (balls carry around their scratches) and ordinary processes (recipes carry their mistakes through to the end). Non-examples are pseudo-processes and pseudo-objects (e.g., Platonic objects, shadows, the Void of Lewis [33]). Hitchcock [24] rightly points out that, thus far, this account leaves the metaphysics of causal processes unclear. The Salmon-Dowe research program is incomplete. But we know of no reason to believe it is not completable, so for the purposes of our discussion we shall describe as "Salmon-Dowe processes" those which fall under some future completed analysis of this type.

If it is causal processes which ground the probabilistic dependencies between variables, then it must be possible to put the variables within a single model into relation with one another via such processes. This suggests a natural criterion of relevance to require of variables within a single causal model: namely, if two variables appear in a causal model, there must be a sequence of possible or actual causal processes connecting them. This makes precise Hitchcock [23], who vaguely requires that pairs of variables not be "too remote" from each other. Note that we do not demand a possible sequence of causal processes between any two variables, but a sequence of possible processes: it may be, for example, that two events are spacewise separated, yet mediated by a common third event. Nor, of course, do we demand *actual* processes between any event types in the model. Probabilistic dependency is founded upon possibilities, realized and unrealized.[14]

[14] That there are causal processes behind the arcs of causal models suggests the answer to one of the concerns about causal modeling that Nancy Cartwright raises, namely that causal reality may not be made up of discrete token events,

The two approaches to understanding causality, dependency and process, have disparate strengths and weaknesses. This disparity has led many to suggest that there is no one concept of causality and that attempts to provide a unified account are confused.[15] While we agree that there may well be various distinct concepts of causality, we are unconvinced that the particular disparity between the dependence and process analyses argues for two concepts of causality. Instead, we propose a causal unification program: that we develop a unifying account that uses the strengths of the one to combat the weaknesses of the other.

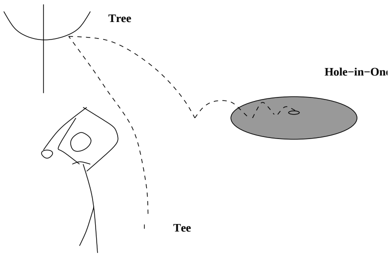

Fig. 8. Rosen's hole-in-one.

Dependency accounts characteristically have difficulties dealing with negative relevance, that is, causes (promoters, in role language) which in some token cases are negatively relevant to the effect (i.e., prevent it), or vice versa. Deborah Rosen [44] produced a nice example of this in response to Suppes [50]. In Figure 8 Rosen has struck a hole-in-one, but in an abnormal way. In particular, by hooking into the tree, she has *lowered* her chance of holing the ball, and yet this very chance-lowering event is the proximal cause of her getting the hole-in-one. The only hope of salvaging probability-raising here, something which all of the dependency accounts mentioned above wanted, is to refine the reference class from that of simply striking the tree to something like striking the tree with a particular spin, momentum, with the tree surface at some exact angle, with such-and-such wind conditions, etc. But the idea that we can always refine this reference class in enough detail to recover a chance-raising reference class is far-fetched. It is what Salmon [46] described as *pseudo-deterministic faith.*[16] In any case, as Salmon also pointed

but perhaps continuous processes instead [4]. Well, we expect that reality is made up of token processes, whether discrete or continuous. Discrete Bayesian networks are a convenient way of modeling them, and the variables we choose are convenient and useful abstractions. They need to be tied to the underlying reality in certain ways, but they certainly do not need to be exhaustive descriptions of that reality.

[15] Hitchcock has suggested this, e.g., in Hitchcock (2004a, b); see also [16].

[16] Note that the escape by contrasting striking the tree with missing it fails on at least two counts. Of course, missing the tree, given the hook, is a contrast class

out, we can always generate chance-lowering causes in games, or find them in quantum-mechanical scenarios, where there is no option for refinement. Take Salmon's "cascade" [46], where a quantum-mechanical system undergoes state changes with the probabilities indicated in Figure 9. Every time such a system reaches state d via state c rather than state b, it has done so through a chance-lowering event of transition into state c. By construction (in case this is a game, by nature otherwise) there is no refinement of the intermediate state which would make the final transition to state d more probable than the untaken alternative path through b; hence, probability-raising alone cannot account for causality here.

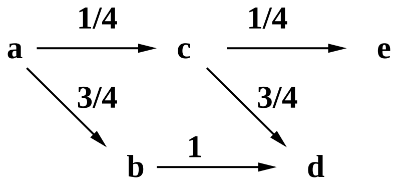

Fig. 9. Salmon's quantum-mechanical cascade.

Salmon's way out was to bite the bullet: he asserted that the best we can do in such cases is to locate the event we wish to explain causally (transition to d) in the causal nexus.[17] If the transition to d occurs through b, then everyone is happy to make the causal attribution; but if it has occurred through c, then it is no less caused, it is simply caused in a less probable way. Insisting on the *universal* availability of promoting causes (probability raising) is tantamount to the pseudo-deterministic faith he denounced [45, chapter 4]. Instead of reliance upon the universal availability of promoters, Salmon asked us to rely upon the universal availability of Salmon-Dowe causal processes leading from each state to the next. This seems the only available move for retaining irreducible state transitions within the causal order.

Assuming, per above, that the metaphysics of process has been completed, the problem remains for Salmon's move that it is insufficient. For one thing, as we saw above, causal processes are composable: if we can carry information from one end to the other along two processes, then if we connect the processes, we can carry the (or, at any rate, some) information from the composite beginning to the composite end. But the many cases of end-to-end probabilistic

with a *lower* probability of success than hitting the tree. But we are attempting to understand causality *relative* to a given causal model. And this maneuver introduces a new variable, namely *how* the ball is hit (or, perhaps, its general direction), so the maneuver is strictly evasive. Secondly, if we are going to countenance new variables, we can just introduce a local rule for this hole: just behind the tree is a large net; landing in the net also counts as holing the ball.

[17] In Salmon's terms, within an objectively homogeneous reference class.

independency need to be accommodated; the possibility of causal intransitivity needs to be compatible with our criteria of causality. Hence, invoking causal processes cannot suffice.

A pure causal process account has other problems as well. Whereas probability raising clearly itself is too strong a criterion, missing minimally every least probable outcome within a non-trivial range of outcomes, simply invoking causal process is clearly too weak a criterion. In some sense the Holists are right that everything is connected to everything else; at any rate, everything within a lightcone of something else is likely to have a causal process or potential process relating the two. But while it makes sense to assert that the sun causes skin cancer, it makes little sense to say that the sun causes recovery from skin cancer. Yet from the sun stream causal processes to all such events, indeed to every event on earth. Salmon's account of 1984 lacked distinction.

It is only in adding back probabilistic dependencies that we can find the lacking distinction. Positive dependencies, of course, have difficulties dealing with negative relevance; processes do not. Processes alone cannot distinguish relevant from irrelevant connections; probabilistic dependencies can. Plausibly what is wanted is an account of causal relevance in terms of processes-which-make-a-relevant-probabilistic-difference. Two accounts which provide this are those of Menzies [35] and Twardy and Korb [52]. What we will do here, however, is apply these two concepts of causal process and difference making to making sense of causal responsibility.

9 Causal Responsibility

Many of the more pressing questions that arise about causality concern questions of responsibility, legal or moral. These are questions about particular facts, that is, particular events and their particular causal relationships. Or, in other words, questions about token causality ("actual causality") rather than type causality.[18] Incidentally, much of the philosophical literature on causality focuses on token causality and stories told about token causality; we shall treat some of them here.

It has always been clear that type and token causality, while distinct, are related, but the relationship is itself not clear. Causal modeling with Bayesian networks provides an opportunity for getting that relationship clear, by providing an opportunity to establish criteria for both based upon the same model. Our analysis aims at putting type-token causality into a kind of general-to-particular relationship. And what we will do here is to outline a

[18] To be sure, any satisfying account of either legal or moral responsibility goes beyond a satisfying account of token causality alone, since it will have to incorporate a treatment of legal or moral principles and their application to causal questions. We will not enter into such matters here.

plausible account of token causality, one that needs work to be complete, but appears to us to be aimed in the right direction.

Our treatment is based upon a presumption that we have in hand the right causal model. That is, what our analysis asserts is, or is not, a cause of some particular event, depends upon the causal model assumed to be true; the token causality analysis itself does not provide guidance in finding that true model. We will instead rely upon certain principles of model building which arise from the prior discussion, although we do not defend them explicitly (for a more complete defence see [30]):

Principle 1 (Intervention): *Variables in a causal model must be intervenable.*

Principle 2 (Distinction): *Every pair of variables in a causal model must have a physically possible intervention which, in some physically possible context, affects the distribution of one variable without affecting that of the other.*

Principle 3 (Process): *If two variables appear in a causal model, there must be a sequence of possible or actual causal processes connecting them.*

The first efforts to provide a Bayesian-net based analysis of token causality were those of Hitchcock [23] and Halpern and Pearl [17, 18]. Hitchcock's is a simplification of [17], which is arguably superior in some ways but more complex than we care to deal with here. The difficulties we will point out with Hitchcock's treatment carry through transitively to that of Halpern and Pearl.

Consider again the case of Hitchcock's hiker (Figure 3). Clearly, what we want to say is that boulders do cause death in such circumstances, if only because human responses are fallible, so the type relations are right in that model — each arc corresponds to a type causal relation that manifests itself in a probabilistic dependency under the right circumstances.[19] But in the particular case — to be sure, idealistically (deterministically) described — the boulder's fall does not affect survival in any way, because there is no probabilistic dependency between the two.

Hitchcock [23] describes two plausible criteria for token (actual) causality. Both of them look at component effects, by isolating some causal path of interest. The first is very simple. Let's call it **H1** (following [22]).

> **H1:** $C = c$ **actually caused** $E = e$ if and only if both $C = c$ and $E = e$ occurred and when we iterate through all $\Phi_i \in \text{Paths}(C, E)$, for some such Φ_i if we block all the alternative paths by fixing them at their actually observed values, there is a probabilistic dependence between C and E.

[19] Of course, we would never say "Boulders falling cause survival." But that's because in our speech acts causal role ordinarily leaks into causal attributions. We are not here interested in a theory of ordinary language utterances about causality.

In application to Hitchcock's hiker's survival, this works perfectly. Considering the direct path *Boulder* ➝ *Survival*, we must fix *Duck* at true, when there is no probabilistic dependency. The second path (through *Duck*) doesn't need to be considered, since there is no variable mediating the path alternative to it, so there is no question of blocking it.

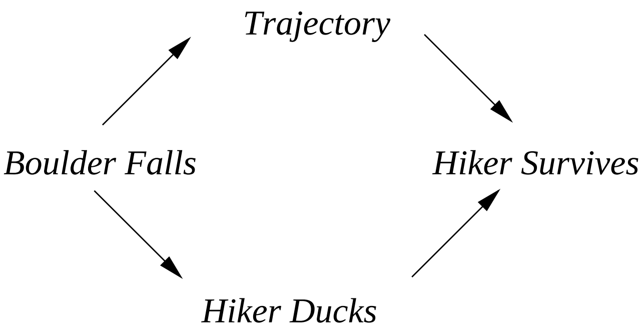

Fig. 10. The hiker surviving some more.

The second path could, of course, be considered if we embed the model of Figure 3 in a larger model with a variable that mediates *Boulder* and *Survival.* We could model the trajectory of the boulder, giving us Figure 10. In this case, H1 gets the wrong answer, since we are now obliged to fix *Trajectory* and discover that there is now a probabilistic dependency between *Boulder* and *Survival.* In particular, if the boulder *doesn't* fall, but somehow regardless achieves its original trajectory, then the hiker won't have ducked and will end up dead. Hitchcock's response to this possibility is to say that the introduction of *Trajectory* requires a "sophisticated philosophical imagination" — we have to be able to imagine the boulder miraculously appearing on collision course without any of the usual preliminaries — and so an account of actual causation for the ordinary world needn't be concerned with it. Hiddleston objects to this as an ad hoc maneuver: he suspects that variables will be called miraculous when and only when they cause trouble for our analysis. However, he is mistaken. Our Intervention Principle makes perfectly good sense of Hitchcock's response. Either *Trajectory* is intervenable (independently of *Boulder*) or it is not. If it is not, then modeling it is a mistake, and H1's verdict in that case is irrelevant. If it is intervenable, then there must be a possible causal process for manipulating its value. A possible example would be: build a shunt aimed at the hiker through which we can let fly another boulder. For the purposes of the story, we can keep it camouflaged, so the hiker has no chance to react to it. All of this is possible, or near enough. But in order to introduce this variable, and render it intervenable, we have to make the original story unrecognizable. Hitchcock's criterion, just as much as ours to follow, is model-relative. The fact that it gives different answers to different models is unsurprising; the only relevant question is what answer it gives to the right model.

This reveals some of the useful work our model-building principles do in accounting for actual causation, even before considering the details of any explicit criterion.

H1 handles a variety of examples without difficulty. For example, it copes with the ordinary cases of pre-emption which cause problems for dependency theories. Thus, in Figure 11 if a supervisor assassin fires at the victim if and only if the trainee assassin doesn't fire and, idealistically again, neither the trainee nor supervisor can miss, then an account requiring end-to-end dependency, such as Lewis's original counterfactual analysis of causation [31], fails. In particular, should the trainee fire, this action will not be considered the cause of the victim's death, since there is no dependency. In the face of such counterexamples, Lewis adopted a step-wise dependency of states of the bullet as it traverses the distance to the victim. Although there is no end-to-end dependency, if we take the transitive closure of step-by-step dependencies, we find end-to-end causation. We find this objectionable on two counts: first, as we have seen, causation is not transitive; second, finding the intermediate dependencies requires generating intermediate variables, and so altering the causal story in unacceptable ways. Hitchcock's H1, on the other hand, has it easy here: we simply observe that the supervisor did not fire and that under this circumstance there is a dependency between the trainee's action and the victim's health.

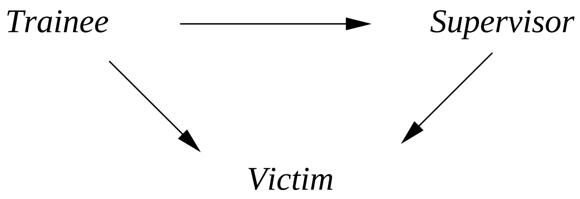

Fig. 11. Pre-emptive assassination.

This is an example of pre-emption by "early cutting". Pre-emption can also occur through late cutting. If Billy and Suzy are each throwing a rock at a bottle (and, as usual, they cannot miss) and if Suzy throws slightly earlier than Billy, then Suzy causes the bottle to shatter and Billy does not (see Figure 12), again despite the fact that there is no end-to-end dependency. In this case, however, there is also no step-wise dependency for Lewis to draw upon: at the very last step, where the bottle shatters, the dependency will always fail, because Billy's rock is on its way.

Hitchcock's H1 fails to accommodate Suzy's throw, because the end-to-end dependency fails under the actual circumstances. So, Hitchcock resorts to

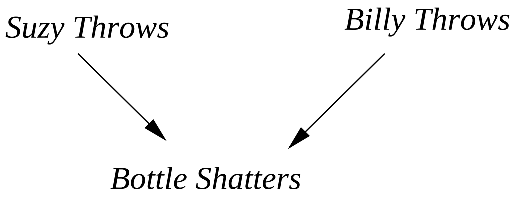

Fig. 12. Pre-emptive bottle smashing.

counterfactual contexts to cope (in this he is following the account of Halpern and Pearl [17]).[20] For these contexts Hitchcock only allows counterfactual circumstances which would not change the values of any of the variables on the causal path under consideration. Any variable off that path will have a range of values which have no impact on the causal path, minimally that value which it actually took. Such values are said to be in the **redundancy range** (RR) for that path. Then the new criterion, **H2**, is:

> **H2:** $C = c$ **actually caused** $E = e$ if and only if both $C = c$ and $E = e$ occurred and when we iterate through all $\Phi_i \in \text{Paths}(C, E)$, for some such Φ_i there is a set of variables $\mathbf{W}$ s.t., when fixed at values in their redundancy ranges relative to Φ_i, there is a probabilistic dependence between C and E.

Since actual values are trivially within the RR, the prior (positive) successes of H1 remain successes for H2. With Suzy and Billy, it's clear that Billy's throwing or not are both within the redundancy range, and the dependency upon Suzy's throw reappears when we consider what happens when Billy's throw is absent. This seems a very tidy solution.

However, Hiddleston [22] offers an example which H2 cannot handle, as usual an example concerning potential violent death. Suppose the king's guard, fearing an attack, pours an antidote to poison in the king's coffee. The assassin, however, fails to make an appearance; there is no poison in the coffee. The king drinks his coffee and survives. Did the antidote cause the king to survive? That is no more plausible than the claim that the boulder falling has caused the hiker to survive; however, H2 makes this claim, since *Poison* being true is in the redundancy range.[21] Interestingly, H1 gets this story right, since the poison is then forced to be absent, when the depen-

[20] Lewis [32] also resorted to counterfactuality, replacing sequences of dependencies with sequences of hypothetical dependencies ("quasi-dependencies"). Incidentally, Halpern and Pearl [17] analyse this case using a temporally expanded (dynamic) network; however, the complexities involved would not reward our treating it here.

[21] It might be pointed out that the model here is incomplete, and an intermediate node which registers the combined state of poison and antidote would push *Poison=true* out of the redundancy range. But that's an ineffective response, unless

dency of survival on antidote goes away. Hiddleston concludes that H1 was just the right criterion all along, but needed to be supplemented with Patricia Cheng's (and Clark Glymour's) theory of causal models and process theory [6, 15, 14]. We agree with his general idea: examining dependencies under actual instantiations of context variables is the right way to approach actual causality. Cheng's causal model theory, however, is far too restrictive, as we noted above.

9.1 An Algorithm for Assessing Token Causation

So, we now present our alternative account of actual causation in the form of an "algorithm" for assessing whether $C = c$ actually caused $E = e$, given that both events occurred. Our steps are hardly computationally primitive, but opportunities for refining them and making them clearer are surely available.

Step 1

Build the right causal model M.

Of course, this is a complicated step, possibly involving causal discovery, expert consultation, advancing the science of relevant domains, and so forth. All of our model-building rules apply. This (hopefully) leads to the right causal model, describing the right type causal relationships. So, this account of token causality starts from the type causal model and gets more specific from there.

This step, applying the model-building principles, circumvents a number of problems that have arisen in the literature. For example, we know that *Time* should not be invoked as a variable (it violates the Intervention Principle) and that problem-defining constraints should likewise be excluded (because of the Distinction Principle). We also know that the imaginative introduction of intermediate events to save some kind of step-wise dependency across a causal chain is (normally) illegitimate. So, despite being open-ended, this "step" is not vacuous.

Step 2

Select and instantiate an actual context O (we will designate the model M in context O by M/O).

Typically, this involves selecting a set O of variables in M and fixing them at their observed values. Often the causal question itself sets the context for us, when selecting the context is trivial. For example, someone might ask, "Given that no poison was added to the coffee, did the antidote cause the king's survival?" Indeed, that appears to be exactly what Hiddleston was asking in the above example.

in fact *no* model of the structure offered by Hiddleston is possible. However, we can always construct such a model, as a game, for example.

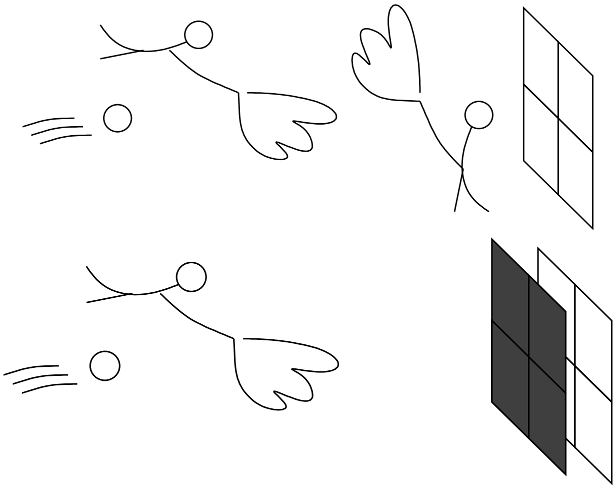

Fig. 13. Backup fielders, imperfect and perfect.

A striking example of context-setting is this (from [9]). Suzy and Billy are playing left and center field in a baseball game. The batter hits a long drive between the two, directly at a kitchen window. Suzy races for the ball. Billy races for the ball. Suzy's in front and Billy's behind. Suzy catches the ball. Did Suzy prevent the window from being smashed? Not a hard question. However, now suppose we replace Billy with a movable metal wall, the size of a house. We position the wall in front of the window. The ball is hit, Suzy races and catches it in front of the wall. Did Suzy prevent the window from being smashed? This is no harder than the first question, but produces the opposite answer. The question, of course, is how can our criterion for actual causation reproduce this switch, when the two cases are such close parallels of each other. What changes is the context in which the causal question gets asked. Here the changed context is not reflected in the values of the variables which get observed — neither the wall nor Billy are allowed to stop the ball; rather, the changed context is in the probability structure.[22] Billy is a fallible outfielder; the wall, for practical purposes, is an infallible outfielder. The infallibility of the wall leaves no possible connecting process between Suzy's fielding and the window.[23]

[22] The causal question being relative to an observational context presupposes that it is also relative to the causal model, including parameters, in which the context is set.

[23] An alternative analysis of this would be to say that since the wall *is* infallible, it effectively has only one state, and so is not a variable at all. Whether we really should remove the variable depends upon whether or not we wish to take seriously the possibility of it altering state, even if only by an explicit intervention.

The idea of a connecting process arose already in the Process Principle, in that there must be a sequence of possible connecting processes between any two variables in the model. Specifically, by **connecting process** we mean any Salmon-Dowe process such that under the relevant circumstances (e.g., O) C makes a probabilistic difference to E. We suggest beyond the Process Principle a somewhat more demanding model building principle for individual arcs is in order:

Principle 4 (Connecting Process): *For every arc $C \rightarrow E$ in M/O there must be a possible value for C such that there is a connecting process between C and E.*

The baseball example is a case where an *individual* arc fails to have a corresponding connecting process for any value of its causal variable. Such arcs we call **wounded** and remove them:[24]

Step 3

Delete wounded arcs, producing M^*.

In the second baseball case the arc *Suzy Catches* → *Window Shatters* starts out wounded: it's an arc that should never have been added and which no causal discovery algorithm would add. But many cases of wounding arise only when a specific context is instantiated. For example, in bottle smashing (Figure 12), if Suzy doesn't throw, there's nothing wrong with the causal process carrying influence from Billy's throw to the bottle. If we ask about Billy's throw specifically in the context of Suzy having thrown first, however, then there is no connecting process. The arc *Billy Throws* → *Bottle Shatters* is **vulnerable** to that specific context, and, given the context, is wounded.

Until now, in Bayesian network modeling two kinds of relationship between pairs of variables $< C, E >$ have been acknowledged: those for which there is always a probabilistic dependency regardless of context set, when a direct arc must be added between them;[25] and those pairs which are screened off from each other by some context set (possibly empty). But vulnerable arcs are those which are sometimes needed, they connect pairs of the first type above, but also they are sometimes not needed; when they are wounded, the arcs can mediate no possible probabilistic dependency, since they cannot participate

Regardless of whether we deal with the wall in parameters or structure, there will remain no possible dependency between Suzy's catch and the window.

[24] For a more careful discussion of the metaphysics of process and wounding we refer the reader to [19].

[25] This means, for every possible set of context variables O there is *some* instantiation of the context $O = o$ such that $P(E|C, O = o) \neq P(E|O = o)$. This is not to be confused with requiring that for all possible context sets, and *all possible instantiations thereof*, C and E are probabilistically dependent, which is a mistake Cartwright [4] makes.

in any active path. We might say, the arcs flicker on and off, depending upon the actual context.

Strictly speaking, deleting wounded arcs is not necessary for the assessment of token causality, since the next step applies a dependency test which is already sensitive to wounding. Removing the wounded arc simply makes the independency graphic, as faithful models do.

Step 4

Determine whether intervening upon C can make a probabilistic difference to E in the given circumstances M^*/O.

Note that we make no attempt here to isolate any path connecting C and E (in contrast with our treatment of type causality above). In token causation our interest is in identifying whether C actually *does* make a difference within context O; to answer this question we must allow that alternative strands of influence may nullify the affect. Thus, in questions of token causation we should allow for neutralizing alternatives. In neutral Hesslow, for example, we would not attribute token causality to the pill, whether or not thrombosis ensued — unless, of course, the woman's state of pregnancy were fixed as part of the context. Allowing neutralization for a type question is not an option, however, since the type question indicates a desire to know whether there is some possible extension to context which yields a difference-making intervention, which there is in neutral Hesslow. So, in this way, token causality differs from a straightforward particularization of type causality, by being bound to the specific context M^*/O.

Since we are not focused here on causal role, the probabilistic difference identified in Step 4 might well be to *reduce* the probability of E; hence, rather than saying that C actually caused E, it might be better simply to say that C is actually causally relevant to E. As the context in which the token causal question gets raised, O, is enlarged, this criterion becomes more particular to the historical circumstances; as the context O shrinks, this criterion more closely resembles type causal relevance, with the exception noted above.

10 Conclusion

Bayesian network technology and its philosophy have reached an important juncture. The representational and inferential methods have proven themselves in academic practice and are beginning to be taken up widely in industry and scientific research. This, and the difficulty in building them by expert elicitation, has fueled considerable work in the automation of causal discovery, which in turn has prompted a reconsideration of the causal interpretation of the models discovered, by both supporters and skeptics. The friction between the two subcommunities has sparked a variety of ideas, both philosophical disputes and initiatives potentially valuable in application, such as the measures of causal power introduced above. This is a lively and productive time for research in causal modeling, both theoretical and applied.

Acknowledgements

We thank Erik Nyberg, Charles Twardy, Toby Handfield, Graham Oppy and Luke Hope for contributing to work and discussions that we have drawn upon here.

References

1. Marc Bouissou and Nguyen Thuy. Decision making based on expert assessments: Use of belief networks to take into account uncertainty, bias, and weak signals. In *Decision-making aid and control of the risks: Lambda-Mu 13/ESREL 2002 Conference*, Lyon, France, 2002.
2. Luc Bovens and Stephan Hartmann. Bayesian networks and the problem of unreliable instruments. *Philosophy of Science*, 69:29–72, 2002.
3. Jeremy Cain. Planning improvements in natural resources management. Technical report, Centre for Ecology and Hydrology, 2001.
4. N. Cartwright. What is wrong with Bayes nets? *The Monist*, 84:242–264, 2001.
5. Nancy Cartwright. *Nature's Capacities and their Measurement.* Clarendon Press, Oxford, New York, 1989.
6. Patricia W Cheng. From covariation to causation: A causal power theory. *Psychological Review*, 104:367–405, 1997.
7. D. M. Chickering. A tranformational characterization of equivalent Bayesian network structures. In P. Besnard and S. Hanks, editors, *Proc of the 11th Conf on Uncertainty in AI*, pages 87–98, San Francisco, 1995.
8. D. Max Chickering. Optimal structure identification with greedy search. *Machine Leaning Research*, 3:507–559, 2002.
9. John Collins. Preemptive prevention. *Journal of Philosophy*, 97:223–234, 2000.
10. G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.
11. G. F. Cooper and E. Herskovits. A Bayesian method for constructing Bayesian belief networks from databases. In Smets D'Ambrosio and Bonissone, editors, *uai91*, pages 86–94, 1991.
12. Phil Dowe. *Physical Causation.* Cambridge University, New York, 2000.
13. R.A. Fisher. Letter. *British Medical Journal*, pages 297–8, 3 August 1957.
14. Clark Glymour. *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology.* MIT, 2001.
15. Clark Glymour and P W Cheng. Causal mechanism and probability: A normative approach. In M Oaksford and N Chater, editors, *Rational models of cognition.* Oxford, 1998.
16. Ned Hall. Two concepts of causation. In J. Collins, N. Hall, and L. A. Paul, editors, *Causation and Counterfactuals*, pages 225–76. MIT Press, 2004.

17. Joseph Y. Halpern and Judea Pearl. Causes and explanations, part I. *British Journal for the Philosophy of Science*, 56:843–887, 2005.
18. Joseph Y. Halpern and Judea Pearl. Causes and explanations, part II. *British Journal for the Philosophy of Science*, 56:889–911, 2005.
19. Toby Handfield, Graham Oppy, Charles Twardy, and Kevin B. Korb. Probabilistic process causality, 2005. Under Submission.
20. Daniel M. Hausman. Probabilistic causality and causal generalizations. In Ellery Eells and James H. Fetzer, editors, *The Place of Probability in Science*. Open Court, 2005.
21. D. Heckerman, D. Geiger, and D.M. Chickering. Learning Bayesian networks: the combination of knowledge and statistical data. In Lopes de Mantras and David Poole, editors, *Proc of the 10th Conf on Uncertainty in AI*, pages 293–301, San Francisco, 1994.
22. Eric Hiddleston. Causal powers. *British Journal for the Philosophy of Science*, 56:27–59, 2005.
23. Christopher R. Hitchcock. The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, 158(6):273–299, 2001.
24. Christopher R. Hitchcock. Routes, processes and chance-lowering causes. In Philip Dowe and Noordhof, editors, *Cause and Chance*, pages 138–51. Routledge, 2004.
25. L R Hope and K B Korb. An information-theoretic causal power theory. In *Proc of the 18th Australian Joint Conference on AI*, pages 805–811, Sydney, NSW, 2005. Springer.
26. Phil Kitcher. Explanatory unification and the causal structure of the world. In Phil Kitcher and Wesley C. Salmon, editors, *Minnesota Studies in the Philosophy of Science*, volume XIII, pages 410–505. Univ of Minnesota, 1989.
27. K. B. Korb, L. R. Hope, A. E. Nicholson, and K. Axnick. Varieties of causal intervention. In *Pacific Rim International Conference on AI*, pages 322–31, 2004.
28. K. B. Korb and A. E. Nicholson. *Bayesian Artificial Intelligence*. CRC/Chapman and Hall, Boca Raton, FL, 2004.
29. Kevin B Korb and Erik Nyberg. The power of intervention. *Minds and Machines*, 16:289–302, 2006.
30. Kevin B Korb, C R Twardy, T Handfield, and G Oppy. Causal reasoning with causal models. Technical Report 2005/183, School of Computer Science and Software Engineering, Monash University, 2005.
31. David Lewis. Causation. *Journal of Philosophy*, 70:556–67, 1973.
32. David Lewis. *Philosophical Papers, Volume II*. Oxford Univ, 1986.
33. David Lewis. Void and object. In J. Collins, N. Hall, and L. A. Paul, editors, *Causation and Counterfactuals*, pages 277–90. MIT Press, 2004.
34. B G Marcot, R S Holthausen, M G Raphael, M M Rowland, and M J Wisdom. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management*, 153:29–42, 2001.
35. Peter Menzies. Difference making in context. In J. Collins, N. Hall, and L. Paul, editors, *Counterfactuals and Causation*, pages 139–80. MIT Press, 2004.
36. R. E. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, 2003.
37. L R Novick and Patricia W Cheng. Assessing interactive causal influence. *Psychological Review*, 111:455–485, 2004.

38. Erik P Nyberg and Kevin B Korb. Informative interventions. Technical Report 2006/204, School of Information Technology, Monash University, 2006.
39. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, Ca., 1988.
40. Judea Pearl. *Causality: Models, reasoning and inference.* Cambridge, 2000.
41. Karl R. Popper. *The Logic of Scientific Discovery.* Basic Books, New York, 1959. Translation, with new appendices, of Logik der Forschung (1934), Vienna.
42. Karl R. Popper. *Objective Knowledge: An Evolutionary Approach.* Oxford University, 1972.
43. H. Reichenbach. *The Direction of Time.* Univ of California, Berkeley, 1956.
44. Deborah Rosen. In defense of a probabilistic theory of causality. *Philosophy of Science*, 45:604–13, 1978.
45. W.C. Salmon. *Scientific Explanation and the Causal Structure of the World.* Princeton, 1984.
46. Wesley Salmon. Probabilistic causality. *Pacific Philosophical Quarterly*, pages 50–74, 1980.
47. P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search.* Number 81 in Lecture Notes in Statistics. Springer Verlag, 1993.
48. P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search.* MIT Press, second edition, 2000.
49. Daniel Steel. Homogeneity, selection and the faithfulness condition. *Minds and Machines*, 16:303–317, 2006.
50. Patrick Suppes. *A Probabilistic Theory of Causality.* North Holland, Amsterdam, 1970.
51. C R Twardy, A E Nicholson, and K B Korb. Knowledge engineering cardiovascular Bayesian networks from the literature. Technical Report 2005/170, Clayton School of IT, Monash University, 2005.
52. Charles Twardy and Kevin B Korb. A criterion of probabilistic causation. *Philosophy of Science*, 71:241–262, 2004.
53. T. S. Verma and J. Pearl. Equivalence and synthesis of causal models. In Smets D'Ambrosio and Bonissone, editors, *Proc of the Sixth Conference on Uncertainty in AI*, pages 255–68, 1991.
54. Chris S. Wallace and Kevin B. Korb. Learning linear causal models by MML sampling. In A. Gammerman, editor, *Causal Models and Intelligent Data Management.* Springer-Verlag, 1999.
55. I H Witten and E Frank. *Data mining: Practical machine learning tools and techniques.* Morgan Kaufmann, 2nd edition, 2005.
56. James Woodward. *Making Things Happen.* Oxford Univ, 2003.
57. S. Wright. The method of path coefficients. *Annals of Mathematical Statistics*, 5(3):161–215, Sep. 1934.