MONIST; APRIL 01, 2001

WHAT IS WRONG WITH BAYES NETS?[*]

By Nancy Cartwright
London School of Economics and University of California, San Diego

1. The basic question: Can we get to causality via Bayes nets?

Probability is a guide to life partly because it is a guide to causality. Work over the last two decades using Bayes nets supposes that probability is a very sure guide to causality. I think not, and I shall argue that here. Almost all the objections I list are well-known. But I have come to see them in a different light by reflecting again on the original work in this area by Wolfgang Spohn and his recent defense of it in a paper titled "Bayesian Nets Are All There Is to Causality".[1]

Bayes nets are directed acyclic graphs that represent probabilistic independencies among an ordered set of variables. The parents of a variable X are the minimal set of predecessors that render X independent of all its other predecessors. If the variables are temporally (or causally) ordered, we can read the very same graph as a graph of the (generic-level) causal relations among the quantities represented, it is maintained. This commits us to the causal Markov condition described below, which is a relative of Reichenbach's claim that conditioning on common causes will render joint effects independent of other another. It is also usual to add an assumption called faithfulness or stability as well as to assume that all underlying systems of causal laws are deterministic (plus the causal minimality condition, which I will not discuss). With these assumptions in hand there are a variety of algorithms for inferring causal relations from independencies. These I will loosely call "Bayes-nets methods".

In criticizing the inference of causes from Bayes nets it is usual to list the objections I note. Is this just an arbitrary list? And why should one have expected any connection between the two to begin with? After all, Bayes nets encode information about probabilistic independencies. Causality, if it has any connection with probability, would seem to be related to probabilistic dependence.

The answers to the two questions are related. When we see why there might be a connection between causality and independence, we see why there will be a list of objections. The answer to "why" is not one that will sound surprising, but I want to focus on it because working through the argument will show that we have been looking at probability and causality in the wrong way. Probabilities may be a guide to causes, but they are, I shall argue, like symptoms of a disease: there is no general formula to get from the symptom to the disease.

2. The call for explanation

It is usual to suppose that once the right set of assumptions are made about the causal systems under study, we can read information about causes from a Bayes net that satisfies those assumptions. Wolfgang Spohn maintains that if there is a tight connection like this between

Bayes nets and an independent notion of causation, there should be a general reason for this. He cannot find one; so he proposes that the notion of causation at stake is not independent. The probabilistic patterns of a Bayes net is our concept of causation: "it is the structure of suitably refined Bayesian nets which decides about how the causal dependencies run."[2]

I agree with Spohn that if there is a tight connection, there should be a reason for it. The alternative is what Gerd Buchdahl called a "brute force" connection, one which holds in nature but has no "deeper" explanation. There are such brute-force connections between concepts we use in science. These are what we record in fundamental laws. And some of them involve relational concepts like" ... causes...." For instance, if the allowed energy configuration for a system in relation to its environment is represented by a specific Hamiltonian, say H, then whatever the system's current state (say phi), the system will evolve in time according to -i/h deltaphi/deltat = Hphi.

But I do not think there is a brute force connection in the case of Bayes nets and causation. That is primarily because there is a reason for the connection, a good reason. The problem is that the reason does not justify a tight connection. The reason lets us see why the connection will hold when it does, but it also allows us to see how loose the connection between the two is. For simplicity I will stick to yes-no causes and effects in the subsequent discussion. We are looking for an equivalence between causal connections and Bayes nets. I will start with causation and see first how--and when--we can get from causation to the probability relations pictured in Bayes nets.

3. From causation to probabilistic dependence

3a. Where have all the caveats gone?

Causes produce their effects; they make them happen.[3] So, in the right kind of population we can expect that there will be a higher frequency of the effect (E) when the cause (C) is present than when it is absent; and conversely for preventatives. What kind of populations are "the right kinds"? Populations in which the requisite causal process sometimes operates unimpeded and its doing so is not correlated with other processes that mask the increase in probability, such as the presence of a process preventing the effect or the absence of another positive process.

Here are some of the conditions that we know need to be satisfied: the necessary triggering and helping causes for C must operate together sometimes in the population, and their joint operation should not be probabilistically dependent on that of other causes or preventatives, nor of C itself. Nor is the operation of C to produce E probabilistically dependent on the operation of other causes or preventatives of E.

The trouble with Bayes nets is that they ignore all these caveats. When Bayes nets are used as causal graphs, effects are probabilistically dependent upon each of their causes. That's it. Nothing can mask this. The assumptions about causality made by the Bayes-net approach go all the way back to the first, then ground-breaking, probabilistic analysis of causality by Patrick Suppes.[4] Suppes begins with prima facie causation: any earlier factor that is correlated with an effect is a prima facie cause of that effect. Real causes are ones that survive the same independence tests

that are required in the Bayes net. But nothing gets to be a candidate for a cause unless it is correlated with the putative effect.

Twenty to thirty years later we have the Bayes-net approach, in all essentials equivalent to the original formulation proposed by Suppes when the subject first began. It is as if Simpson's paradox and causal decision theory never existed. Nor decades of practice by econometricians and other social scientists, who plot not simple regressions but partial regressions. Nor the widely deployed definition proposed by Granger in 1969, which looks for probabilistic dependence only after conditioning on the entire past history of the cause--which ensures that all the other causes up to the time of the putative cause will be held fixed.

The demand that effects always be probabilistically dependent on each cause follows in the Bayes-net approach from the assumption that Peter Spirtes, Clark Glymour and Richard Scheines call faithfulness.[5] Judea Pearl calls it stability.[6] The assumption is necessary to the approach. Without it the procedures developed by Pearl and by Spirtes, et al., cannot get very far with the discovery of causal connections; and the proofs that assure us that they will not make mistakes if sample sizes are large enough will not go through. For Spohn it matters because he argues that causal connections are the connections marked out on God's big Bayes net once the variables have been temporally ordered. With the faithfulness/stability assumption, the causal connections are unique; but they are seldom unique otherwise.

For those readers who are not already deeply into this discussion let me rehearse the standard objections to the assumption that all genuine causes are prima facie causes. First there is Simpson's Paradox: facts about probabilistic dependency can be reversed in moving from populations to subpopulations. For example, factor X may be positively dependent or negatively dependent or independent of Y in a population, but still be any of these three in all partitions of the population along the values of a third factor Z, if Z is itself probabilistically dependent on X and Y. Z may, for instance, be a preventative of Y; because of its correlation with X, the presence of X does not, after all, increase the frequency of Y's in the population as my opening argument suggests.

It is typical in social science to sidestep this problem by looking for probabilistic dependence between a putative cause and its effect only in subpopulations in which all possible confounding factors are held fixed. This is apparent in the econometrics concept of Granger causality, where X Granger causes Y if X and Y are probabilistically dependent holding fixed everything that has occurred up to the time of the putative cause.

The same strategy was at the heart of various versions of causal decision theory two decades ago. What is the probability that if I were to do C, E would occur? The conditional probability P(E/C) gives the wrong answer. It could be either way too big or way too small because of the operation of confounding factors; relevant here--it could be zero even though my doing C could have a substantial impact on whether E occurs. One standard proposal (the one I urged[7]) is to set P(C*(This character cannot be represented in ASCII text)E) = P(E/C&K), where K is a state description over the values of a full set of "other" causal factors for E that obtain (or will obtain) in the decision situation.[8] Where we do not know the values of the factors in K we should

average over all possible values using our best estimate of the probability that they will occur: summation of P(E/C&K[sub j])P(K[sub j]).

Exactly the same formula has reappeared among the Bayes-nets causal theorists. Judea Pearl has recently produced a very fine and detailed account of counterfactuals and their probabilities, based on Bayes nets. According to Pearl, the probability of y if we were to "set" X = x[sub j] is summation of P(y/x[sub j] & parents of x[sub j])P(parents of x[sub j]).[9] Despite Pearl's endorsement of stability, from this formula it looks as if a factor can have a high degree of causal efficacy even though on his account it is not really a cause at all because it is not prima facie a cause. I take it that Pearl does not take this to be a problem because he thinks cases where "stability" is violated involve "'pathological' parameterizations"[sup 10] and are not in the range he will address.

The second kind of case usually cited in which genuine causes are not prima facie causes is when one and the same cause has different kinds of influence on the effect. The different influences may cancel each other. The easiest version of this to handle is when a given factor acts as both cause and preventative of the effect, by different routes. G. Hesslow's birth-control pills[11] are the canonical philosophical example. The pills are a positive cause of thrombosis. On the other hand, they prevent pregnancy, which is itself a cause of thrombosis. Given the right weights for the three processes, the net effect of the pills on the frequency of thrombosis can be zero.

If we suspect that cancellations of this kind are occurring, we can confirm our suspicions by looking at the probabilities of thrombosis, given the pills in populations in which factors from the separate causal routes between pills and thrombosis are held fixed. But this is no comfort to the Bayesnets theorist.

Even this strategy is not available where there are no routes between the cause and its effects. We have two kinds of trouble with routes. The first is a worry that I share with Glenn Shafer.[12] Let us accept that in every case of singular causation there is a temporally continuous process connecting the cause with the effect. That does not guarantee that there will always be a vertex between any two other vertices in God's great causal graph. That is because the graphs are graphs of causal laws that hold between event-types. Every token of the cause-type may in actuality be connected by a continuous process with the effect token without there being some chain of laws in between: C[sub 1] causes C[sub 2], ..., C[sub n] causes E. For instance, the signal from the cause to the effect may piggyback on causal processes that the cause in question does not initiate. This is particularly likely where there are a lot of processes with the necessary spatiotemporal relations already available for use in conveying the influence from the cause to the effect.[13]

Roman Frigg offers a number of examples.[14] He explains that what is wanted are cases for which there is

A generic law without contiguity: On the generic level cause and effect do not exhibit contiguity, neither in space nor in time.

No unique causal chain: There is no unique causal chain that connects cause and effect. That is, on the concrete level the connection between the two can be realised in many different ways.

No unique transmission of causal information: The causal information may be transmitted in different ways. (In addition the kind of physical and/or institutional structures that guarantee the capacity of the cause to bring about its effect may be totally different from those that guarantee that the causal message is transmitted, i.e., the causal law and the individual chains connecting the cause and effect may result from different structures.)

One of Frigg's examples involves as cause Person B getting an HIV virus from another person and as effect, that B dies later on. He tells us

1. The infection with the HIV virus leads in most cases to death. But a long period of time elapses between these two events ...

2. The infection with the HIV virus leads (in most cases) to the outbreak of AIDS, i.e., the destruction of the immune system. This in turn can lead to death by a variety of different routes. To mention just a few: diarrhoea (various pathogenes possible); encephalitis with brain atrophy; neuropathy; pneumonia; ringworm (various types); meningitis; herpes simplex; tuberculosis; and fever.

3. The causal information can initially be transmitted on various different paths as well: Sexual contact (vaginal, oral, anal), exchange of blood (blood transfusions, use of dirty needles, injuries in the hospital), communication from the mother to the child.[15]

Frigg also offers examples of death by malaria and from exposure to strong radiation, and of the democratic election of an individual as president of a country resulting in that person's becoming president, the receipt of a court order causing someone to appear in court and the ordering of a plane ticket causing a person to receive a ticket.

The other is a problem we all sweep under the rug: the representations in a causal graph are discrete; for every vertex there is always a predecessor and a successor. Is causality really like that? If it is then we can have causes of causes with mixed influences on the effect, directly, not by different routes. So the device of holding fixed vertices along the various routes is not even available to provide us with a way of using facts about probabilistic dependencies and independencies to test whether a factor is really causally inefficacious rather than having mixed influence on the effect.

3b. Can the caveats be ignored?

What justification do Bayes-nets theorists give for ignoring all these caveats and insisting that all causes must appear as causes at the first crude look? Spirtes, Glymour and Scheines discuss Simpson's Paradox at length. They present two graphs, the first embodying Simpson's paradox; the second is a graph that by contrast is "faithful" to all the independencies assumed in the paradoxical case--i.e., as in Suppes's original formulation, all causes are prima facie causes:[16]

They then invite us to

[s]uppose for a moment that we ignore the interpretation that Simpson gave to the variables in his example.... Were we to find A and C are independent but dependent conditional on B, the Faithfulness Condition requires that if any causal structure obtains, it is structure (ii). Still, structure (i) is logically possible,[17] and if the variables had the significance Simpson gives them we would of course prefer it. But if prior knowledge does not require structure (i), what do we lose by applying the Faithfulness Condition; what, in other words, do we lose by excluding causal structures that are not faithful to the distribution [i.e., that allow genuine causes that do not appear as causes prima facie]?[18]

I assume that this passage is meant as a defense of the faithfulness condition since it appears at the end of the long exposition of Simpson's Paradox in the section in which they introduce faithfulness as an axiom[19] and just before the only other remark that could be construed as a defense of this axiom in the face of Simpson's Paradox. But what is the defense? The answer to their question is obviously: What we lose is getting the causal structure right.

Perhaps they mean to suggest that when we do not know anything, it is more reasonable to plump for structure (ii) than for structure (i). But what is the argument for that? I respond with a truism: when you don't know, you don't know; and it is often dangerous to speculate. If we have no idea what the variables stand for, let alone how they operate, we are not in a position to make a bet with any degree of credibility. "Ah yes," I am sometimes told, "but what if you had to bet?" Well, tell me more about the context in which I am forced to bet--a psychological experiment perhaps?--and I may be able to tell you which bet I would plump for.

Perhaps, however, Spirtes, Glymour and Scheines are speaking sloppily. They did not mean "What do we lose?" but rather "How often will we lose?" For immediately after this they report that "[i]n the linear case, the parameter values--of the linear coefficients and exogenous variances of a structure--form a real space, and the set of points in this space that create vanishing partial correlations not implied by the Markov condition [i.e., that violate faithfulness] have Lebesgue measure zero.[20]

This is surely intended as an argument in favor of faithfulness--and it is frequently cited as being so intended--though I am not sure exactly what the conclusion is that it is supposed to support. I gather we are to conclude that it is unlikely that any causal system to which we consider applying our probabilistic methods will involve genuine causes that are not prima facie causes as well.

But this conclusion would follow only if there were some plausible way to connect a Lebesgue measure over a space of ordered n-tuples of real numbers with the way in which parameters are chosen or arise naturally for the causal systems that we will be studying. I have never seen such a connection proposed; that, I think, is because there is no possible, plausible story to be told. Moreover, were some connection mooted, we should keep in mind that it could not bear directly on the question of how any actual parameter-value is chosen because, as we all know, any specific point in the space will have measure zero. So we not only need a story that connects a Lebesgue measure over a space of n-tuples of real numbers with how real parameter values arise, but we need a method that selects as a question to be addressed before values are chosen: shall values occur that satisfy faithfulness or not?

Not only is the theorem about Lebesgue measure not relevant to the issue of whether all causes are prima facie causes. I think it is an irresponsible interjection into the discussion. Getting it right about the causal structure of a real system in front of us is often a matter of great importance. It is not appropriate to offer the authority of formalism over serious consideration of what are the best assumptions to make about the structure at hand.

Judea Pearl argues somewhat differently about the choice of parameter values. He uses the term stability for the condition that insists that effects be probabilistically dependent on their causes even before confounding factors are conditioned on. Here is what he says in its entirety:

Some structures may admit peculiar parameterizations that would render them indistinguishable from many other minimal models that have totally disparate structures. For example, consider a binary variable C that takes the value 1 whenever the outcomes of two fair coins (A and B) are the same and takes the value 0 otherwise. In the trivariate distribution generated by this parameterization, each pair of variables is marginally independent yet is dependent conditioning on the third variable. Such a dependence pattern may in fact be generated by three minimal causal structures, each depicting one of the variables as causally dependent on the other two, but there is no way to decide among the three. In order to rule out such "pathological" parameterizations, we impose a restriction on the distribution called stability....

This restriction conveys the assumption that all the independencies imbedded in [the probability distribution] P are stable; that is, they are implied by the structure of the model D and hence remain invariant to any changes in the parameters [of D]. In our example only the correct structure (namely, A right arrow C left arrow B) will retain its independence pattern in the face of changing parameterizations--say, when one of the coins becomes slightly biased.[21]

We can see here two points of view that Pearl takes that make stability seem plausible to him. First, Pearl thinks causal systems should be decidable. It is clearly a criticism of the systems described that "there is no way to decide among the three". This attitude is revealed in discussions of other topics as well. For instance, as we shall see below, I reject the causal Markov condition. Pearl objects that by so doing I make questions about the causal structure and about the truth of certain counterfactuals unanswerable.[22]

Unanswerable given what information? Immediately after the section defining "stability" Pearl tells us, "With the added assumption of stability, every distribution has a unique minimal causal structure ..., as long as there are no hidden variables."[23] Clearly he intends that the questions he is concerned about should be answerable given an order for the full set of causally relevant variables and the probability distribution over them. But so far as I can see, once we have given up the idea that there is something wrong with the notion of cause so that it has to be reduced away, there is no good reason to suppose that probabilities should be able to answer all questions about causality for us. (Nor am I sure that Pearl insists they should; for it is unclear whether he thinks all causal systems are stable or takes the more modest line that his methods are capable of providing answers to all his questions only for systems that are stable.)

The other point of view that matters for Pearl's claims about stability is the point of view of the engineer--which he is. It is apparent from the passage that Pearl thinks of causal structures as in some sense coming first: they get fixed, but then the parameter values can vary. But of course a causal system comes with both its structure and its parameters--you can't have one without the other.

I think the way to put the issue that makes sense of the idea of "structure first" is in terms of the kinds of operation we typically perform on the kinds of engineered devices Pearl generally has in mind. Think of a toaster. Its parts and their arrangement are fixed. We may bend the position of the trip plate a little, or of the expanding metal strip which it will meet, in order to keep the brownness of the toast calibrated with the settings on the brownness control. The values of the parameters do not matter so long as the basic causal structure does not break down; indeed the values are just the kind of thing we expect to drift over time. But we would have a legitimate cause of complaint if the same were true of the structure within the first year we owned the toaster.

That is fine for a toaster. But for other situations the parameters may matter equally with the structure, or more so. If birth-control pills do cause thrombosis we may work very hard to weaken the strength with which they do so, at least to the point where people who take the pills are no worse off than those who do not. Indeed we may take this as an important aim--we are more obliged to get the effects to cancel out than we are to continue to spend money and research time to reduce the risk of thrombosis among pill-takers below that of non-pill-takers. Getting the cancellation that stability/faithfulness prohibits is important to us.

This brings us to what seems to me a real oddity in the whole idea of stability/faithfulness. Probabilities and causal structures constrain each other. If the probability is fixed, then nature--and we--are not allowed to build certain kinds of causal structures. For instance, if we have the three binary variables, A, B and C, as in Pearl's example, with a probability in which they are pairwise independent and have to create a causal arrangement (or lack of!) among the three, we are prevented from building just the one Pearl describes. Or, to think of it with causal structure first, as Pearl generally seems to, if C is to take value 1 or 0 depending on whether the outcome of the flip of two coins is the same or not, we are prevented from using fair coins and must introduce at least a little bias.

I come, finally, to the question of whether we should in fact expect to see a lot of causes that are not causes prima facie. A good many of the systems to which we think of applying the methods advocated by Bayesnets theorists are constructed systems. Either highly designed, like a toaster or an army admissions test, or a mix of intentional design, historical influence and unintended consequences, as in various socio-economic examples. In these cases cancellation of the effects of a given cause, either by encouraging the action of other factors or by encouraging the contrary operation of the cause itself, can be an important aim, particularly where the effect is deleterious. It will often be a lot easier to design for, or encourage the emergence of, cancellation than it is to eliminate the cause of the unwanted effect, or less costly or more beneficial overall (as in my discussion of birth-control pills). There is no good reason to assume that our aims are almost always frustrated.[24]

This is a view that Kevin Hoover also stresses in his work on causality in macroeconomics. He considers a macroeconomic example in which "agency can result in constraints appearing in the data that [violate faithfulness]."[25] He concludes

Spines et al.... acknowledge the possibility that particular parameter values might result in violations of faithfulness, but they dismiss their importance as having "measure zero". But this will not do for macroeconomics. It fails to account for the fact that in macroeconomic and other control contexts, the policymaker aims to set parameter values in just such a way as to make this supposedly measure-zero situation occur. To the degree that policy is successful, such situations are common, not infinitely rare.[26]

Perhaps, however, the issue will be made: can we ever really expect exact cancellation? After all, to get an arrow in a Bayes-net causal graph, any degree of dependence between cause and effect will do. After we have the arrow in, we need not be misled by the smallness of the dependence to think the influence is small. For we can then insist on measuring degree of efficacy by the formula above that I and other causal-decision theorists proposed and that Pearl endorses for P(C *(This character cannot be represented in ASCII text) E).

One reason we may think exact cancellations are rare is that actually getting any really precise value we aim for is rare. In a recent discussion of instrumentalism, Elliott Sober[27] talks about a comparison of the heights of corn plants in two populations. One thing we know, he claims, is that they are not really equal. Still, that is the working hypothesis. I take it that one of the reasons he thinks we know this is that "exactly equal" is very precise; and any very precise prediction is very likely to be wrong in an imprecise discipline.

This raises some very difficult issues about modelling and reality, especially for probabilities. We design a device to set the difference between two quantities at zero; tests for quality-control show that, within bounds of experimental error, we succeeded; and we model the difference as zero. Should we think it "really" is zero? It is not certain the question makes sense, even when we are thinking of, say, a difference between the length of two strips in a single designated device. It becomes particularly problematic when we are thinking about a difference of two probabilities in a population. Is the increase in probability of thrombosis on taking birth control pills exactly offset by the decrease via pregnancy prevention in British women between the ages of 20 and 35 in the period from 1980 to 1990? All the conventional issues about what we intend by talking about the true probability become especially acute here.

Some, I think, we can sidestep, particularly when we are thinking about the application of the Bayes-nets approach to causality as opposed to the philosophical issue about substitutability raised by Spohn. For we are going to be using these methods in doing real social, medical and engineering science, using real data.[28] And here it is not unusual for our best estimates from the data to render two quantities probabilistically independent where estimates of appropriate partial conditional probabilities--as well perhaps as our background knowledge or even other kinds of tests we have conducted for the relevant causal connections--suggest the result is due to cancellation. In this case we either have to insist that the probabilities are not those our best estimates indicate or forsake the commitment to faithfulness.

Before leaving this section I should repeat an old point, for completeness. Sometimes it is argued that Bayes-nets methods should supplement what we know. So if we do have independent evidence of cancellation, we should use it and not insist on faithfulness. But where we do not have such information we should assume faithfulness. As I indicated earlier, this strategy is ill-founded; indeed, I think irresponsible. Where we don't know, we don't know. When we have to proceed with little information we should make the best evaluation we can for the case at hand--and hedge our bets heavily; we should not proceed with false confidence having plumped either for or against some specific hypothesis--like faithfulness--for how the given system works when we really have no idea.

4. From probabilistic dependence to causality.

If we have a hypothesis that C causes E, we can use what we have just reviewed to test it, via the hypothetico-deductive method. But that is a method that we know to be more accurate at rejecting hypotheses than confirming them. Bayes-nets methods promise more: they will bootstrap from facts about dependencies and independencies to causal hypotheses--and, claim the advocates, never get it wrong.

Again, as Spohn argues, if there really is this tight connection, there ought to be an argument for why it obtains. And there is. Again, we can see from looking at the argument why the inference from and independencies sometimes works, and why it will not work all the time. As with the other direction of inference, there is an argument for the connection and the argument itself makes clear that the connection is not tight.

What kinds of circumstances can be responsible for a probabilistic dependence between A and B? Lots of things. The fact that A causes B is among them: Causes produce their effects; they make them happen. So, in the right kind of population we can expect that there will be a higher frequency of the effect (E) when the cause (C) is present than when it is absent; and conversely for preventatives. With caveats.

What else? Here are a number of things, all discussed in the literature: 1) A and B may have a common cause or a common preventative or correlated causes or correlated preventatives, where either the causes are deterministic or the action of producing B is independent of the action producing A. 2) A and B may cooperate to produce an effect. In populations where the effect is either heavily present or heavily absent, A and B may be dependent on each other. 3) When two populations governed by different systems of causal laws or exhibiting different probability distributions are mixed together, the resulting population may not satisfy the causal Markov condition even though each of the sub-populations do. (This is analogous to Simpson's paradox reversals.) 4) A and B may be quantities with the same kind of temporal evolution, both monotonically increasing, say. Then the value of A at t will be probabilistically dependent on the value of B at t. 5) A and B may be produced as product and by-product from a probabilistic cause.

Let us look at each in turn and at what the defenders of Bayes nets have to say about them. I begin with 1) which is the case that advocates of Bayes-nets methods acknowledge and try to deal with squarely--assuming the underlying system is deterministic.

3a) Why factors may be dependent: 1) Common causes (where nature is deterministic).

Following Judea Pearl?[29] let us call the total effect of all those causes of X that are omitted from V and which combine with the direct causes in V of X to form a set of causes sufficient to fix the value of X, a random disturbance factor for X.[30] Bayes-net methods are applied only to special sets of variables: sets V such that for each X in V, the random disturbance factor for X is probabilistically independent of that for every other variable in V. In such a variable set we can prove that the causal Markov condition will be satisfied.[31]

The causal Markov condition, along with the assumption that all causes are prima facie causes, lies at the heart of the Bayes-net methods. It tells us that a variable will be probabilistically independent of every other variable except its own effects once all of its direct causes have been conditioned on. So we eliminate cases where a dependence between A and B is due to reason 1) by requiring that the dependence persist once we have conditioned on the parents of A.

Everyone acknowledges that some constraint like this is necessary. You cannot get directly from dependence to causation; you at least have to first hold fixed the causal parents or something equivalent, then look for dependence. So in the remaining sections when I talk about the route from dependence to causation, I mean dependence conditional on a set of causal parents. To claim that this is enough to ensure a causal connection is to maintain the causal Markov condition.

My description of the restriction on the variable set V is rather long-winded. The first reason is to avoid a small problem of characterization. What I have called "random disturbance factors" are sometimes called "exogenous" factors. There are various concepts of exogeneity. This usage obviously refers to the one in which exogenous factors are not caused by any variables in the system. Pearl clearly assumes that is true of random disturbance factors. But the proof requires more, for it is possible for all exogenous causes of one variable to be independent of those for another without the disturbance terms themselves being independent. That's because it is possible for a function of X and Y to be dependent on Z even if the three factors are pairwise independent. So it is not enough that the exogenous causes for a variable be independent of those for other variables: the proof needs their net effects to be independent.

The second reason is that some of the other terminology used in the discussion is unclear. Often we are told, as by Spirtes, Glymour and Scheines,[32] that the methods will be applied only to sets that are causally sufficient, adding the bold assumption that as a matter of empirical fact, this will ensure the necessary independencies among the disturbance factors. But what is causal sufficiency? Spirtes, Glymour and Scheines tell us, "We require for causal sufficiency of V for a population that if X is not in V and is a common cause of two or more variables in V, that the joint probability of all variables in V be the same on each value of X that occurs in the population."[33]

Let us assume that C is a common cause of A and B if C is a cause of A and a cause of B. The problem then is that this definition is too demanding. Every cause of a common cause is itself a common cause. These could go back in time ad infinitum. And for any system for which there is

a temporally continuous process connecting a cause with an effect at the type-level, between each common cause and an earlier one there will be infinitely more. If we apply the methods only to variable sets that get them all in, we will not apply them at all. What we want to get in are all the last ones--the ones as close to both effects as possible.[34] But it will take some effort to formulate that properly. Spirtes, Glymour and Scheines are particularly hampered here because they restrict their definitions to facts about causally correct representations rather than talking about causal relations in the world.

Spirtes, Glymour and Scheines avoid this problem by offering a different characterization. They define, "We say that a variable X is a common cause of variables Y and Z if and only if X is a direct cause of Y relative to {X,Y,Z} and a direct cause of Z relative to { X,Y,Z } ."[35] And for direct cause: "C is a direct cause of A relative to V just in case C is a member of some set C included in V/{A} such that (i) the events in C are causes of A, (ii) the events in C, were they to occur, would cause A no matter whether the events in V/({ A } union of C} were or were not to occur, and (iii) no proper subset of C satisfies (i) and (ii)." [36]

The variables in V are seldom sufficient to fix the value of an effect. So how can fixing whether the events in C occur or not ensure that A occurs? So let us add into the set C the random disturbance factor for A. But in fact, it looks as if we have to assume as fixed all exogenous causes, or at least the "last" one if that makes sense, since it will not help to fix one but allow temporally subsequent ones to vary. We also need to add that quantities occurring between C and A in nature's objective graph must be assumed to take on the values dictated by C. And so forth.

I do not know how to formulate all this correctly. But it needs to be done if the notion of causal sufficiency is to be used. Quite reasonably the advertisements for Bayes-nets methods make much of the fact that the subject is formal and precise: we can prove theorems about manipulation, abut efficient conditioning sets for measuring the size of a causal effect, about the certainty of the results of the algorithms when applied to systems satisfying specific conditions, etc. But this is all pseudo-rationalism if we do not provide coherent characterizations of the concepts we are using.

The trouble with the characterization of "causal sufficiency" arises from the fact that for Spirtes, Glymour and Scheines the notion of direct cause is relative to the choice of a particular variable set. Spohn's talk of the set of "all variables needed for a complete description of empirical reality",[37] temporally ordered, avoids this; Pearl, too, because he supposes that the underlying system is a set of deterministic causal laws on a finite set of causally ordered variables. No-one to my knowledge has a good account of causal sufficiency for dense sets of effects, for instance, for the kinds of systems studied by time-series analysis. As I remarked, Spirtes, Glymour and Scheines talk only of correct causal representations?[38] That not only allows them to appear to avoid metaphysics, as Spohn and Pearl clearly do not, but also leaves an opening for supposing that the underlying metaphysics is continuous. But the advantages are illusory if we cannot produce adequate definitions.

Before proceeding to look at the list of factors that undermine the causal Markov condition, I should comment on one recent defense of it. Daniel Husman and James Woodward[39] offer a proof of the condition alternative to the proof by Pearl and Verma. Central to their discussion is a

concept they call modularity: each separate effect under study should be manipulable without disturbing any other. They claim that, given certain other conditions (such as the existence for each effect of a cause not in the variable set under consideration), the causal Markov condition is "the flip side of" modularity. This would be a good defense if it were true. For we need not agree with Hausman and Woodward that all causal systems must be modular; we could nevertheless (supposing their other conditions are met) assume the causal Markov condition whenever we assume modularity.

The trouble is that the proof does not bear the interpretation they put on it. For given their other conditions, both modularity and the causal Markov condition follow separately. One is not the flip side of the other, both are the result of the conditions they suppose at the start. And these conditions are at any rate strong enough to call the Pearl and Verma proof of the causal Markov condition into play.[40]

3b. Other reasons why factors may be dependent.

2) When two causes cooperate to produce one effect, they will be mutually dependent in a population homogeneous with respect to that effect. These kinds of cases are common in practice. Data is hard to come by. We collect it for one reason, but need to utilize it for many others. Imagine for example that we have data on patients from a given hospital, where one disease, D, is especially prevalent. But we are interested in another condition, B. Unbeknownst to us B cooperates with A in the production of D, so A and B are dependent in our population (even once we have conditioned on all the parents of A in a causally sufficient variable set). We erroneously infer that A causes B in this kind of population.

The problem here is not with the sample size. We can imagine that the sample is large and the frequencies are indicative of the "true" probabilities for the population involved. The problem for the causal Markov condition is with the choice of population. We all know that to study the relation between A and B we should not use populations like this. But how should we--properly--characterize populations "like this"?

3) Mixing. Even if we assume the causal Markov condition for populations where the probabilities of the effect are fixed by the causal history, for mixed populations cooperating causes can still be correlated if the proportion of the effect is determined by some external factor rather than the causal history. Spirtes, Glymour and Scheines tell us that there are no cases of mixing for causally sufficient variable sets: "When a cause of membership in a sub-population is rightly regarded as a common cause of the variables in V, the Causal Markov Condition is not violated in a mixed population."[41] I think this is a bad idea: the "variables" that are the "cause of membership in a sub-population" will often look nothing like variables --they don't vary in any reasonable way and there is no reason to think there is a probability distribution over them; and even if we did count them as variables, it looks as if we would have to count them as common causes of every variable in V to ensure restoration of the causal Markov condition.[42]

4) Many quantities change in the same direction in time. There will thus be a probabilistic dependence between them. Social scientists solve this problem by detrending before they look for dependencies. Spirtes, Glymour and Scheines maintain that there is no problem to solve.

They use their previous solution to the problem of mixing plus a bold claim: "If we consider a series in which variable A increases with time, then A and B will be correlated in the population formed from all the units-at-times, even though A and B have no causal connection. Any such combined population is obviously a mixture of populations given by the time values."[43]

Like others[42] I find this claim ungrounded. Moreover it seems to me to be in tension with their commitment to determinism--which is important to them since in deterministic systems the causal Markov condition is bound to be true if only we add enough into the set of parents. Their idea I take it is that there will be different probability distributions across the causes operating at each time slice, hence mixing. But consider deterministic models in physics. These I take it are important for Spirtes, Glymour and Scheines because these are what make many people sympathetic to their claim that all macroscopic processes are deterministic. Any two systems moving inertially will have their positions correlated, and they have exactly the same causes operating at each instant with the same probability distribution over them, namely--none.

5) Products and by-products are mutually dependent, and when causes act purely probabilistically, no amount of conditioning on parents will eliminate the dependence. Perhaps then there isn't any purely probabilistic causation --that would save the causal Markov condition. But that is a big metaphysical gamble, especially in the face of the fact that for the kind of variables for which Bayes-nets methods are being sold, we seldom are able to formulate even a reasonable probabilistic model, let alone a deterministic one. We can of course stick to the metaphysical insistence that everything must be deterministic. I think this claim is unwarranted, but I will say no more about the matter here since I have written much about it elsewhere.[45]

5. Analysis

Dependence could be due to causation. But there are lots of other reasons for it as well. Bayes-nets methods stress one--the operation of common causes--and tell us how to deal with it when the underlying system is deterministic. The other four reasons standard in the philosophical literature are badly handled or made light of. And what about other reasons? Have we listed them all?

The reasons I listed are prevalent not only in the philosophical literature. They are standard fare in courses on social-science methodology, along with lots of other cautions about the use of probabilities to infer causes in even experimental or quasi-experimental contexts. And they are not handled so badly there. In part the failures in the philosophical discussion arise from the requirement that the connection, whatever it is, be tight. We look for a claim of the form: A causes B iff A and B are probabilistically dependent in populations satisfying X. Then X is hard to formulate in the kind of vocabulary we need for formal proofs and precise characterizations.

But why should we think the connection is tight? As Spohn says, if it is tight there ought to be a reason. There is, as I have argued, a reason for the connection between probabilistic dependence and causality, but the very reason shows that the connection is not tight. Causes can increase the probability of their effects; but they need not. And for the other way around: an increase in probability can be due to a causal connection; but lots of other things can be responsible as well.

I think we are still suffering under the presumptions of the old Hume program. First, we don't like modalities, especially strange ones. My breakfast cereal box says: "Shredded Wheat can help maintain the Health of your Heart." In the same sense, causes can increase the probability of their effects. Distressed at this odd modality,[46] we try to render this as a claim that causes will increase the probability of their effects, given X; then we straggle to formulate X. Second, we cannot get loose from the idea that causes need associations to make them legitimate. So we want some "if and only if" with probabilities on the right, even if we grudgingly have to use some causal concepts to get the right-hand side filled in properly. I think we are looking at the issue entirely the wrong way. The connection between causes and probabilities is not like that. It is, rather, like the connection between a disease and one of its symptoms. The disease can cause the symptom, but it need not; and the same symptom can result from a great many different diseases.

This is why the philosopher's strategy is bad. We believe there must be some "if and only if" and so are inclined to make light of cases that do not fit. The advice from my course on methods in the social sciences is better: "If you see a probabilistic dependence and are inclined to infer a causal connection from it, think hard. Consider the other possible reasons that that dependence might occur and eliminate them one by one. And when you are all done, remember--your conclusion is no more certain than your confidence that you have eliminated all the possible alternatives."

NOTES

* Research for this paper was supported by a grant from the Latsis Foundation, for which I am extremely grateful. The work is part of the Measurement in Physics and Economics Research Project at LSE.

1. Spohn, forthcoming.

2. Spohn, forthcoming, p. 10.

3. When we deal with quantities of more than two values, there are other possibilities; e.g., a cause may raise the level of the effect.

4. Suppes 1970.

5. Spirtes, Glymour and Scheines 1993.

6. Pearl 1999.

7. Cartwright 1983, "Causal Laws and Effective Strategies."

8. What counts as a complete set of factors is not so easy to characterize for probabilistic causality. (For one definition, see Cartwright 1989, p. 112.) The task is easy if we are allowed to help ourselves to the notion of the objective system of causal laws governing a population, as

Pearl and Spohn and I do (see Cartwright, forthcoming). In that case K ranges over all the parents of E, barring C, relative to God's great causal graph for the population.

9. Pearl 2000, p. 73.

10. Pearl 2000, p. 48.

11. Hesslow 1976.

12. Shafer 1996.

13. For further discussion, see Cartwright 1999, chs. 5 and 7.

14. Frigg 2000.

15. Frigg 2000, p. 1.

16. Spirtes, Glymour and Scheines 1993, p. 68.

17. I suppose they mean by "logically" possible that it is consistent with the other assumptions they wish to make about causal laws and probabilities.

18. Spirtes, Glymour and Scheines 1993, pp. 67-68.

19. In fact this is not literally true since the section, though headed "Axioms", only introduces a definition of faithfulness and does not make any claims about it. It is clear from the various sales pitches they make for their methods, however, that they take it to be a condition true of almost all causal systems.

20. Spirtes, Glymour and Scheines 1993, p. 68.

21. Pearl 2000, p. 48.

22. UCSD Philosophy of Economics Seminar, May 1999.

23. Pearl 2000, p. 49.

24. For further discussion see Cartwright 1999, ch. 2, and 2000.

25. Hoover forthcoming, pp. 7-33.

26. Hoover forthcoming, pp. 7-35.

27. Sober 1999.

28. For an example of an attempt to use the Spirtes, Glymour and Scheines methods on real economic data in economics, see Swanson and Granger 1997. Their struggles there are particularly relevant to my point in this paragraph. Which of the low partial correlations observed in their data should be taken to indicate that the "true" partial correlation is zero? They consider various alternatives choices among the lowest observed partial correlations and show that different choices give rise to different causal structures.

29. Pearl 2000, p. 44.

30. These are often designated u[sub x] when variables in V are designated x, y,. ...

31. Cf. Verma and Pearl, 1991 or Pearl, 2000, p. 30.

32. Cf. Spirtes, Glymour and Scheines 1993, p. 54.

33. 1993, p. 45.

34. Note that where A cright arrowB and B cright arrowC and B cright arrowD, if C and D are independent conditioning on B, they need not be independent conditioning on A if P(B/A) is not equal to 1. So B must be included if the causal Markov condition is to be satisfied.

35. Spirtes, Glymour and Scheines 1993, p. 44.

36. Ibid., p. 43.

37. Spohn forthcoming, p. 11.

38. This is my account. They say what they do. For instance, in their section "Axioms," they provide no axioms bur only definitions (of the causal minimality condition, the faithfulness condition, and the causal Markov condition). But I take it their claims are: every correct causal graph over a causally sufficient set of variables satisfies these conditions.

39. Hausman and Woodward 1999.

40. For a full discussion see Cartwright forthcoming b).

41. Spirites, Glymour and Scheines 1993, p. 60 (italics in the original).

42. For details see Cartwright 1999, ch. 5 and Cartwright 2000.

43. Spirtes, Glymour and Scheines 1993, p. 63.

44. Cf. Berkovitz forthcoming and Hoover forthcoming, ch. 7, and Sober 1988, pp. 161-62 and 2000.

45. See Cartwright 1997, Cartwright 1999, ch. 5 and Cartwright 2000.

46. I offer a treatment--though not yet really satisfactory--of these kinds of modality in Cartwright 1999.

REFERENCES

Berkovitz, J. forthcoming, "The Many principles of the Common Cause," in Reports on Philosophy, Issue on the Common Cause Principle.

Buchdahl, G. 1969, Metaphysics and the Philosophy of Science, Oxford: Blackwell.

Cartwright, N. 1983, How the Laws of Physics Lie, Oxford: Oxford University Press.

----- 1989, Nature's Capacities and Their Measurement, Oxford: Clarendon Press.

----- 1997, "What Is the Causal Structure," in Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences, Vaughn R. McKim and Stephen P. Turner (eds.), University of Notre Dame Press, 1997.

----- 1999, The Dappled World, A Study of the Boundaries of Science, Cambridge: Cambridge University Press.

-----. 2000, "Measuring Causes: Invariance, Modularity and the Causal Markov Condition," Measurement in Physics and Economics Discussion Paper Series. Centre for the Philosophy of Natural and Social Science, London School of Economics.

----- forthcoming, "Two Concepts of Invariance, Two Concepts of Intervention and Two Concepts of What it Is to Be Causally Correct," Philosophy of Science.

----- forthcoming, "Against the Causal Markov Condition," CPNSS Discussion Paper Series: Measurement in Physics and Economics, London School of Economics.

Frigg, R. 2000, "Examples of Non-Contiguous Causation," MS, CPNSS, London School of Economics.

Granger, C. 1969, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," Econometrica 37, pp. 424-38.

Hausman, D. and Woodward, J. 1999, "Independence, Invariance and the Causal Markov Condition," British Journal for Philosophy of Science 50, pp. 521-83.

Hesslow, G. 1976, "Discussion: Two Notes on the Probabilistic Approach to Causality," Philosophy of Science 43, pp. 290-92.

Hoover, K. forthcoming, Causality in Macroeconomics, Cambridge: Cambridge University Press.

Pearl, J. 1999, Causality, Cambridge: Cambridge University Press.

Pearl, J. and Verma, T. 1991, "A Theory of Inferred Causation," in Allen, J. A., R. Fikes and E. Sandewall (eds.), Principles of Knowledge, Representation, and Reasoning, pp. 441-52. San Mateo, CA: Morgan Kaufmanns.

Spirtes, P., Glymour, C., and Scheines, R. 1993, Causation, Prediction and Search. New York: Springer Verlag.

Swanson, N. and Granger, C. 1997, "Impulse Response Functions Based on Causal Approach to Residual Orthogonalization in Vector Autoregressions," Journal of American Statistical Association 92, pp. 357-67.

Shafer, G. 1996, The Art of Causal Conjecture, Cambridge, MA: M.I.T. Press.

----- 1988, Reconstructing the Past, Cambridge, MA: M.I.T. Press.

----- 1999, "Instrumentalism Revisited," Critica 91, pp. 3-39.

----- 2000, "Venetian Sea Levels, British Bread Prices, and the Principle of the Common Cause," MS., Department of Philosophy, Logic and Scientific Method, London School of Economics.

Spohn, W. forthcoming, "Bayesian Nets Are All There Is to Causality," in: D. Costantini, M. C. Galvotti, P. Suppes (eds.), "Stochastic Dependence and Causality," Stanford, CA; CSLI Publications.

Suppes, P. 1970, A Probabilistic Theory of Causality, Amsterdam: North Holland.

~~~~~~~~
~~~~~~~~