

Bayesian Nets Are All There Is To Causal Dependence

Wolfgang Spohn
Fachgruppe Philosophie
Universität Konstanz
D-78457 Konstanz

1. Introduction

There are too many theories of causation to get into the focus of a small paper. But there are two in which I have a natural interest since they look almost the same: namely the theory of Clark Glymour, Peter Spirtes, and Richard Scheines, so vigorously developed since 1983¹ and most richly stated in Spirtes et al. (1993) (whence I shall refer to it as the SGS theory), and my own theory, published since 1978 in a somewhat irregular way. They look almost the same, but the underlying conceptions turn out to be quite dissimilar. Hence, the original idea for this paper was a modest one: simply to compare the philosophical basics of the two theories. However, no paper without a thesis! Therefore I have sharpened my comparison to the thesis written right into the title.

The plan of the paper is simple. Section 2 sets out the formal theory of Bayesian nets in an almost informal way, and section 3 analyses the philosophical differences hidden in the common grounds. Section 4 briefly extends the comparison to the treatment of actions or interventions.

2. Causal Graphs and Bayesian Nets

Whenever we want to conduct a causal analysis in a given empirical field, we have to start by conceptually structuring this field. This is usually done by specifying a frame or a set U of variables characterizing the field. Each variable $A \in U$

¹ The acknowledgments of Glymour et al. (1987) report that the work on that book took about four years.

can take some value from the set of its possible values. Thus, by specifying a value for each variable in U we specify some possible small world, some way how the empirical field characterized by the frame U may realize.

Variables should be conceived here as specific and not as generic variables. A generic variable would be something like social status or annual income which may take different values for different persons at different times. However, it is hard to find any causal order among generic variables. One then finds causal circles – high social status tends to generate high annual income, and vice versa –, and one even finds apparent self-causation – social status tends to reproduce itself.

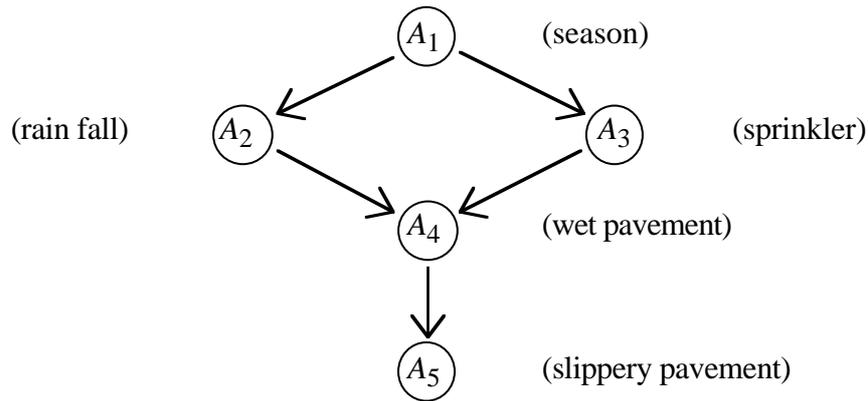
By contrast, a specific variable is something like my social status today or my annual income in 1998, not conceived as it actually is, which is given by some particular figure, but conceived as something which may take any value, say, between 0 and 1 billion Euros. There is a proper causal order among specific variables. For instance, there is no self-causation. If my social status today is high, it tends to be high tomorrow as well (though there is no guarantee, see the sudden fall of politicians), but this is a causal relation between two different specific variables.

Indeed, the causal structure within the frame U of specific variables is neatly captured by a causal graph over U which is nothing but a *DAG*, a *directed acyclic graph* $\langle U, E \rangle$ with U being its set of nodes and E being its set of edges. That the graph is directed means that its edges are directed, i.e. that E is an asymmetric relation over U , and that it is acyclic means that the directed edges don't form circles, i.e. that even the transitive closure of E is asymmetric.

Let me give a standard example (used by Pearl 1998 and elsewhere): U consists of five variables:

- A_1 : season of a given year (spring, summer, fall, winter)
- A_2 : rain fall during the season (yes, no)
- A_3 : sprinkler during season (on, off)
- A_4 : wet pavement (yes, no)
- A_5 : slippery pavement (yes, no)

which we might plausibly arrange into the following DAG (if the variables refer to some place in Southern California):



The DAG $\langle U, E \rangle$ becomes a *causal graph*, if the edges in E are given a causal interpretation, i.e. if an edge $A \rightarrow B$ is interpreted as stating that A is directly influencing B , or that B is directly causally dependent on A , within the given frame U . Thus, so far the DAGs simply express the formal properties of direct causal dependence.

Specific variables have a specific temporal location. Hence, the variables in U are temporally ordered. So I shall add the natural constraint that in any edge $A \rightarrow B$ of a causal graph A temporally precedes B . Some philosophers oppose, but this is not the place to discuss their worries.

The next and crucial step is to introduce probabilities. The frame U generates, as mentioned, a space of possible small worlds the subsets of which may take probabilities according to some probability measure P . In particular, each event of the form $\{A = a\}$, stating that the variable A takes the value a , gets a probability. Accordingly, there is probabilistic dependence and independence among variables. More explicitly, we may define the sets X and $Y \subseteq U$ of variables to be *probabilistically independent* given or *conditional on* the set $Z \subseteq U$, i.e. $X \perp_P Y / Z$, iff for all x, y, z $P(X = x | Y = y, Z = z) = P(X = x | Z = z)$, i.e. iff, given any realization z of Z , any event about X is probabilistically independent of any event about Y .

Following SGS, we can state two conditions concerning a DAG $\langle U, E \rangle$ and a measure P for U , in which $Pa(A)$ denotes the set of parents or immediate predecessors of the node A , $Nd(A)$ denotes the set of non-descendants of A , and $Pr(A)$ denotes the set of nodes temporally preceding A .

There is, first, the *Markov condition* (cf. Spirtes et al. 1993, pp. 53ff.) stating that for each $A \in U$ $A \perp_P Nd(A) / Pa(A)$. i.e. that each variable is independent from all its non-descendants given its parents. If the DAG agrees with the given temporal order this condition is equivalent to the apparently weaker condition that for each A

$\in U \ A \perp_P Pr(A) / Pa(A)$. This condition is also equivalent to the decomposability of P :

$$P(U = u) = \prod_{A \in U} P(A = a \mid Pa(A) = x),$$

where a and x , respectively, are the realizations of A and $Pa(A)$ according to the realization u of U . This decomposability harbors enormous computational advantages so ingeniously exploited by Pearl (1988) and others.

For instance, the above example satisfies the Markov condition iff

$$\begin{aligned} &A_3 \perp_P A_2 / A_1, \\ &A_4 \perp_P A_1 / \{A_2, A_3\}, \text{ and} \\ &A_5 \perp_P \{A_1, A_2, A_3\} / A_4, \end{aligned}$$

or iff, for all a_1, \dots, a_5 realizing A_1, \dots, A_5

$$P(a_1, \dots, a_5) = P(a_1) \cdot P(a_2 \mid a_1) \cdot P(a_3 \mid a_1) \cdot P(a_4 \mid a_2, a_3) \cdot P(a_5 \mid a_4).$$

There is, second, the *minimality condition* (cf. Spirtes et al. 1993, pp. 53f.) stating that no proper subgraph of the DAG $\langle U, E \rangle$ satisfies the Markov condition. Following Pearl (1988, p. 119) a DAG satisfying the Markov and the minimality condition is called a *Bayesian net(work)*. In a Bayesian net, the parents of a node thus form the smallest set of variables for which the relevant conditional independence holds.

For instance, the above example satisfies the minimality condition iff none of the following independencies holds:

$$\begin{aligned} &A_2 \perp_P A_1, \\ &A_3 \perp_P A_1, \\ &A_4 \perp_P A_2 / A_3 \text{ and } A_4 \perp_P A_3 / A_2, \text{ and} \\ &A_5 \perp_P A_4. \end{aligned}$$

SGS further introduce a third condition, *the faithfulness condition* (cf. Spirtes et al. 1993, pp. 56), which is, however, more complicated and slightly less important so that I shall neglect it in the sequel.

So far, I have only introduced two distinct graph-theoretical representations: one of causal dependence between variables and one of conditional probabilistic dependence. However, the core observation of each probabilistic theory of causation is that there is a close connection between causal and probabilistic dependence, that the two representations indeed coincide, i.e. that *each causal graph is a Bayesian net*. Thereby, the Markov and the minimality condition turn into the *causal* Markov and the *causal* minimality condition. This means, to repeat, that the set of variables on which *A directly causally depends* within the frame *U* is *the* smallest set conditional on which *A* is probabilistically independent from all its other non-effects or, equivalently, from all other temporally preceding variables.² This assertion may indeed be used to *define* direct causal dependency within the frame *U*. At least I proposed to do so in Spohn (1976/8, section 3.3, in particular pp.117f.). The definitional equivalence also follows from the assumptions made by SGS.

So far there is perfect agreement between SGS and me. However, there are also differences: first, concerning the development of causal theory, and second, concerning the understanding of the basic theory thus laid out. I shall dwell on the second point, but let me briefly mention the main differences of the first kind.

In my work, I did not use, and did not even think of, any graph-theoretical methods. These methods, graph-theoretic representations of independence relations, so-called d-separation, etc., were essentially introduced and pushed forward by Judea Pearl and his group after around 1985 (cf. Pearl 1988, pp. 132ff.). I am enthusiastic about these methods. They add powerfully to the strength, beauty, and vividness of the theory. Of course they are richly used by SGS. What I did have, however, in Spohn (1976/8, sections 3.2+3, with some variations translated in Spohn 1980), was the above-mentioned probabilistic definition of direct causal dependence and the full theory of conditional probabilistic independence on which this definition and the graph-theoretic methods rest, i.e. the graphoid and the semi-graphoid axioms, including the conjecture of their completeness (refuted by now) and the weaker conjecture of the completeness of the properties of direct causal dependence entailed by them (proved by now).³

Naturally, I wondered how the above account of causal dependence between variables may be founded on an account of causal relations between events or states of affairs or singular propositions. This is obviously philosophically important, but of

² That there is exactly one such set is a consequence of the properties of conditional probabilistic independence.

³ For the conjectures see Spohn (1976/8, p. 105 and p. 119). For the positive and negative results see, e.g., the overview in Spohn (1994).

little use in scientific and statistical methodology, and thus of no concern to SGS. The foundation seemed straightforward: the event $\{A = a\}$ is a *direct cause* of the event $\{B = b\}$ in the possible small world u if and only if both events occur in u , if $\{A = a\}$ precedes $\{B = b\}$, and if $\{A = a\}$ is positively relevant to $\{B = b\}$ according to P under the obtaining circumstances C , which are best identified with the event that all the variables preceding B (and differing from A , of course) take the values they take in u . Thus, the variable B directly causally depends on the variable A iff some event about A is a direct cause of some event about B in some possible small world. For a long time, I was under the influence of the view of Suppes (1970, p. 58) that probabilistic causation cannot be transitive. In Spohn (1990) I changed my mind and started to prefer defining (direct or indirect) causation as the transitive closure of direct causation, though, as explained there, the issue is quite intricate.

Finally, in Spohn (1983, ch. 5 and 6; see also 1988) I have proposed the theory of ranking functions, as they are called nowadays, which yield a perfect deterministic analogue to probability theory, to conditional probabilistic dependence and independence, to the theory of Bayesian nets, and thus to the above account of probabilistic causation, and I have suggested there that this is how deterministic causation should be analyzed.⁴

So I have always moved within the philosophical confines. By contrast, Judea Pearl and his collaborators have done impressive work developing and utilizing the whole theoretical field for the purposes of artificial intelligence in a most detailed and fruitful way. And SGS have done impressive work developing sound statistical methodology on a sound philosophical basis, a different and in many respects much more difficult endeavor which starts to be successful in the big statistical community. Though all this work is addressed, to a large extent, to other departments, it contains a lot of high philosophical interest. But there is no place to further expand on this.

3. About the Causal Import of Bayesian Nets

Let me turn, then, to the interpretational differences between SGS and me which are my main concern. For this purpose, let us look again at the proposed definition: the variable A directly causally depends, within the frame U , on all and only the members of the smallest set of variables in U preceding A conditional on which A is

⁴ A suggestion which I have coherently explained in English only in Spohn (2000b).

probabilistically independent from all other variables in U preceding A . This definition hides two relativizations which deserve closer scrutiny.

First, direct causal dependence is obviously frame-relative according to this definition. The relativization would be acceptable, if it concerned only the direct/indirect distinction: what appears to be a direct causal dependency within a coarse-grained frame may well unfold into a longer causal chain within a more fine-grained frame. In this sense the frame-relativity is also accepted by SGS (cf. Spirtes et al. 1993, pp. 42f.). It's worse, however. The whole notion of causal dependence is frame-relative according to this definition: where there appears to be a direct or an indirect causal dependency within a coarse-grained frame, there may be none within a more fine-grained frame, and vice versa. This consequence seems harder to swallow.

The second relativization is better hidden. The talk of conditional independence refers, of course, to an underlying probability measure. Where does it come from?

It might come from reality, so to speak. This raises the question, of course, how to conceive of objective probabilities – a large question which I want to cut short by simply saying that they should best be understood as chances or propensities. This, however, is obscure enough. I have three reservations about using chances in the present context.

The first reservation is that chances are hard to find. But we want, and do, apply the probabilistic theory of causation almost everywhere, and in particular to fields where it is very unclear whether genuine chances exist. Almost all examples of SGS are from social sciences, medical sciences, etc. Maybe, if basic physics is chancy, everything else in the universe is chancy, too. But if so, we suffer from a complete lack of understanding of the chances, say, in economics or medicine, and whatever the probabilities are we are considering in these fields, they are certainly not suchlike chances.

A further reservation is that I find it very awkward in the meantime to talk of chancy events being caused (as has been most forcefully argued by Railton 1978). The idea behind genuine chances is that of partial determination without further determinability, and the idea behind causation is that of full determination. So, it's rather only the chances of events which are fully determined or caused and not the chancy events themselves. I certainly agree with Papineau (1989, pp. 308 and 320) that we need a probabilistic theory of causality in any case and that it is then largely a matter of terminology whether we should say that something that has raised the chance of an occurring event is among the causes of that event or only among the causes of the chance of that event. Still, my terminological preference is clear.

Mainly, however, my reservation is due to the fact that the above theory would be doomed as an analysis of causation if it starts with the notion of chance. The philosophical point of the enterprise is to elucidate the obscure notion of causal necessitation or full determination, and then the notion of chance or partial determination is presumably part of the package to be elucidated. To analyze the one in terms of the other does not seem helpful. I rather hoped to get a grip somehow on both notions together, on causation and chance.

If objective probabilities are thus to be avoided in the above definition of causal dependence, the only alternative is to use subjective probabilities. This is certainly an option, indeed the one I always preferred. However, it clearly amounts to a further relativization of causation to an epistemic subject or to its epistemic state. The above definition then says not what causal dependence *is*, but only how it is *conceived* by some epistemic subject.

This relativization is certainly in good Humean spirit. But even Hume who maintained it so bravely, was ambiguous and denied it at other places. Likewise, I have never been happy with these relativizations, but I did not get clear about how to get rid of them and what else to say about causation.

For instance, I could not see that the manipulability account of causation is of any help. Whether to explain the notion of something influencing something else by the notion of myself influencing something else or the other way around does not seem to make much of a difference. Moreover, actions, goals, etc. always deemed to me extraneous to the topic of causation. I found no help in the process theory of causation of Salmon (1984). Rich and illuminating as it is, its fundamental distinction between processes and pseudoprocesses leads in a large circle back to counterfactuals. So why not immediately engage into a counterfactual analysis of causation? Alluding to mechanisms is unhelpful since mechanisms seem to be nothing but suitably refined causal chains. The idea of energy transfer seems entirely beside the point when it comes to causation in the social sciences. Postulating a second-order universal of causal necessitation adds little in itself. And so forth.

So, the crucial question persisted: what else to say about causation? Only slowly it dawned upon me that I might, and indeed should, turn the inability to say more into a positive thesis. In a sense which I shall explain below there is nothing more to say about causation than I already did!

By contrast, these relativizations are plainly unacceptable to SGS, and this is, I admit, only common-sensical. They do not want, and do not pretend, to give an analysis of causation. They rather want to develop a theory over some undefined

notion of causation, just as statistics is a big theory over some undefined notion of probability. So, in effect, they develop a theory jointly about causation and probability (cf. Spirtes et al. 1993, pp. 5ff. and 41ff.).

Their attitude, then, is this. Causal dependence, whatever it is, is ubiquitous. However, we are able to model only small parts of empirical reality by tentatively describing them by causal graphs and statistical hypotheses. The basic axiom of this model building is that these causal graphs are Bayesian nets, i.e. satisfy the Markov and the minimality condition introduced above (and also the faithfulness condition). The frame-relative definition of direct causal dependence is thus only an equivalence following from their axiom and has no explicative status. This shows clearly that their underlying conception is quite different from mine.

The natural follow-up question is: why should the axiom hold? SGS do not claim universal validity. The Einstein-Podolsky-Rosen paradox and quantum entanglement in general seem to provide a noticeable exception on which, however, I would like to be silent as well. But this does not diminish the success of the axiom elsewhere. They summarize their defense of the axiom in the following way:

”The basis for the Causal Markov Condition is, first, that it is necessarily true of populations of structurally alike pseudo-indeterministic systems whose exogenous variables are distributed independently, and second, it is supported by almost all of our experience with systems that can be put through repetitive processes and whose fundamental propensities can be tested.” (Spirtes et al. 1993, p. 64)

I am not quite satisfied by this. The first defense points to an interesting and important fact, but defers the issue to deterministic causation. And the second defense shows that we have a lot of intuitive skills and scientific knowledge in order to select appropriate sections of reality. But they continue the summary of their defense:

”Any persuasive case against the condition would have to exhibit macroscopic systems for which it fails and give some powerful reason why we should think the macroscopic natural and social systems for which we wish causal explanations also fail to satisfy the condition. It seems that no such case has been made.”

Indeed, it is interesting how they argue about specific putative counter-examples. Their strategy is always the same: whenever there is a causal graph which is not a Bayesian net, there exists a suitable causal refinement of the original graph which is

a Bayesian net. In the specific cases they discuss I find their argument convincing, for instance, when they reject the interactive forks of Salmon (1984, pp. 168ff.).⁵

But why should this strategy always work (with the disturbing exception already noticed)? Two possible explanations come to my mind. One possibility is that we have an *independent* notion of causation, and using that notion we generally happen to find suitable refined causal graphs which are Bayesian nets. But surely it is incredible that we merely happen to find these refinements. There should be a general reason for this success. Here one might continue in the following way:

Basically, causation is deterministic, and then, given a specific conception of deterministic causation, we can specify very general conditions under which such causal relationships get displayed in Bayesian nets. This is the strategy pursued by Papineau (1985). It is also the strategy behind SGS' theorem that (linear) pseudo-indeterministic systems, i.e. systems with a suitable (linear) deterministic extension in which the exogenous variables are independently distributed, satisfy at least the Markov condition (cf. Spirtes et al. 1993, pp. 58ff.).

This strategy is very illuminating as far as it goes. But I doubt that it works in the end. My reason for my doubt is that I don't believe that we have a workable theory of deterministic causation which could play this independent role. Rather I believe, as already indicated, that all our problems and arguments about probabilistic causation turn up all over again when deterministic causation is at issue.⁶

Hence, I don't think that the strategy presently envisaged works on the basis of deterministic causation. And I do not see any other independent notion of causation for which it has been, or could be, argued that it generally exhibits itself in Bayesian nets. So I am indeed skeptical of the whole approach.

How else might we explain that there always are suitably refined causal graphs which are Bayesian nets? The only other possibility which comes to my mind is to say that there is *no* independent notion of causation to be alluded to, that this *is* our understanding of causation. In other words: it is the structure of suitably refined Bayesian nets which decides about how the causal dependencies run. *We cannot regard B to be causally dependent on A unless we find a sequence of arrows or directed edges running from A to B in a suitably refined Bayesian net and unless, of course, this stays to be so in further refinements.* The last clause shows that the talk

⁵ This rejection is of vital importance to their and my enterprise. If interactive forks were not only an apparently unavoidable, strange exception, as in the EPR paradox, but a perfectly normal and unsurprising phenomenon, as Nancy Cartwright argues again in this volume, then Bayesian nets would lose much of their interest, and my title thesis would simply be wrong.

⁶ See Spohn (2000b) for some substantiation of this claim.

of suitable refinements is unnecessarily vague. In the final analysis it is the all-embrasive Bayesian net representing the whole of reality which decides about how the causal dependencies actually are.

Of course, we are bound to have only a partial group of this all-embrasive Bayesian net. Therefore it is important to have theorems telling under which conditions and to which extent our partial grasp is indicative of the final picture, that is, under which conditions the causal relations in a fine-grained Bayesian net are maintained in coarsenings. The theorem of SGS about pseudo-indeterministic systems is a good example. Clearly, however, the conditions to be specified in such theorems cannot be but assumptions about the shape of the final picture.

These remarks indicate how I propose to get rid of the two relativizations of causal dependence explained above. If the notion of causal dependence is *prima facie* frame-relative, we can eliminate this relativity only by moving into the all-embrasive frame containing all variables needed for a complete description of empirical reality. The all-embrasive Bayesian net, then, does not distribute subjective probabilities over this frame in some arbitrary way. Rather, full information about the maximal frame should be accompanied by full information about the facts, so that subjective probabilities are optimally informed and thus objective at least in the sense proposed by Jeffrey (1965, ch. 12). In this way, the relativization of causal dependence to an epistemic state is eliminated as well.⁷

I am well aware that by referring to the all-embrasive frame and to objective probabilities in this sense I am referring to entirely ill-defined and speculative entities. It is clear, moreover, that all causal theory can only deal with specific frames and specific Bayesian nets and their relations. Still, I find it philosophically inevitable to refer to such ill-defined entities, and the philosophical task is to try to strip them at least of some of their obscurity.

This finally explains my claim that in a sense there is no more to causal dependence than the above definition: this definition *with* its relativizations does all the theoretical work, and the move just proposed to eliminate these relativizations and thus to say what causal dependence really is is only a philosophical appendix adding no substantial theoretical content.

This needs two qualifying remarks. The first remark is that, even in the sense intended here, it is not wholly true that Bayesian nets exhaust all there is to the notion of causal dependence. I have hardly addressed the relation between time and causa-

⁷ Or at least reduced. My vague formulations do not allow conclusions concerning the uniqueness of the objective probabilities thus understood.

tion and not at all the relation between space and causation, and both add considerably to the notion of causal dependence, i.e., to how the all-embrasive Bayesian net has to look in the final analysis. By contrast, I have already expressed my doubts that such notions as action, mechanism, energy transfer, or process further enrich the notion of causal dependence. Anyway, whatever the further aspects of the notion of causal dependence, the theory of Bayesian nets covers its central conceptual content.

The second remark is that one must be very clear about the status of my claim that unrelativized, i.e. actual causal dependence is relativized causal dependence relative to the all-embrasive frame and Jeffreyan objective probabilities. This is very much like the claim of Putnam (1980) that the ideal theory cannot be false. Both assertions are *a priori true*. Something is *a priori true* iff it cannot *turn out* to be otherwise. By contrast, something is necessarily true iff it cannot *be* otherwise. Hence, there is nothing metaphysically necessary about the truth of the ideal theory. The world could easily be different from what the ideal theory says even given the truth of the ideally complete evidence on which it relies. But the world cannot turn out to be different from what the ideal theory says because this theory exhausts all factual and counterfactual means of evidence.

Similarly, causal dependence cannot turn out to be different from what it is in the all-embrasive Bayesian net. But again this is only an epistemological claim, slightly more contentful than Putnam's claim, which has nothing to do with the metaphysics of causation. Indeed, I was completely silent on the latter. If I had wanted to say something about the metaphysics, I should have entered the whole of science, and then, of course, much more could be said.

Let me emphasize once more that I believe exactly the same story to apply to deterministic causation. There, again, Bayesian nets form the conceptual core of causal dependence, the only difference being that Bayesian nets are now constructed not in terms of probability measures, but in terms of ranking functions, their deterministic analogue.

4. Actions and Interventions

When I started to write about causation in Spohn (1976/8), my real interest was decision theory. Therefore action variables were part of my picture from the outset. More precisely, I considered not only a set U of occurrence variables, as I called them for want of a better term, but also a set V of action variables. Thus the frame considered was always $U \cup V$. In decision contexts the task is to find the optimal action, action sequence, or strategy, and once one has found it, one starts executing it (unless weakness of will interferes). Hence, it does not make sense to assume the decision maker to have a probabilistic assessment of his *own* possible actions. For this reason I postulated that a decision model must not explicitly or implicitly contain any probabilities for the action variables in V (and thus took opposition to Jeffrey 1965).⁸ So instead of considering one probability measure P over $U \cup V$ I followed Fishburn (1964, pp. 36ff.), and assumed a family $\{P_\nu\}$ of probability measures over U , parametrized by the possible action sequences ν realizing V , which were to express probabilities of events over U conditional on ν . It is straightforward then to extend the notion of conditional dependence and independence to such a family $\{P_\nu\}$, with the effect that relativized causal dependence can be explained relative to the frame $U \cup V$ in the way sketched above and that a causal graph over $U \cup V$ can be constructed which is a Bayesian net (in a slightly generalized sense). Consequently, all action variables are exogenous variables in that graph (but there may be more), and they introduce an asymmetry into the independence relation since occurrence variables can be (conditionally) independent from action variables, whereas the question whether an action variable is independent from another variable cannot arise simply because no probabilities are assigned to actions.⁹

A natural application of this account is Newcomb's problem, of course, which is basically a problem about the relation between probability and causality. As I observed in Spohn (1978, sect. 5.1), the account just sketched entails that among the four combinations of probabilistic dependence on and independence from action variables on the one hand and causal dependence on and independence from action variables on the other exactly one is impossible, namely the case that something is probabilistically dependent on, but causally independent from the action variables.

⁸ See also our exchange in Spohn (1977) and Jeffrey (1977). I still believe my principle "no probabilities for one's own options" to be correct and full of important consequences. It expresses, for instance, the most basic aspect of the freedom of the will since it exempts the will, i.e. willful actions, from causes, at least in the eyes of the agent. Cf. Spohn (1978, p 193).

⁹ For all this see Spohn (1976/8, sections 3.1+2).

But this, and only this, was the case Nozick (1969) worried about. Accordingly, there is no Newcomb problem, and two-boxing emerges as the only rational option. I still think that this observation is basically sound.¹⁰

When studying causation more closely later on, I neglected action variables for the sake of simplicity. But one can observe a growing interest in the explicit consideration of action variables in the theory of causation and the surrounding statistical and AI literature which certainly relates also to the triumph of the graph-theoretical methods. Thus, a theory of intervention or manipulation has become also a central part of the SGS theory.

Their picture is this (cf. Spirtes et al. 1993, pp. 75ff.). They start with an unmanipulated graph, as they call it, over a frame U . Then they consider one or several manipulations which they represent through a set V of variables enriching the original frame U in such a way that they are exogenous variables in the enriched or combined graph and directly manipulate or act on some variables in U . These intervention variables in V have a zero state which says: "Don't interfere!" or "Let it go!" If they take this state, the original unmanipulated graph stays in force. But if they take another state they enforce a new distribution on the directly manipulated variables irrespective, and thus breaking the force, of the ancestors of the directly manipulated variables in the unmanipulated graph. In the simplest case the new distribution will outright dictate a certain value to the directly manipulated variables. Their so-called manipulation theorem says then how to compute all the probabilities of the manipulated graph from the unmanipulated graph and the new distributions of the directly manipulated variables. All this provides also a nice and precise explanation of the epistemological difference between observing a variable to take a certain value and making it to take that value¹¹ which entail two quite different belief revisions (cf. also Meek, Glymour 1994, pp. 1007ff.).

However, the SGS theory of manipulation strikes me as being essentially equivalent with my old proposal just sketched. I did not distinguish a particular unmanipulated graph or, what comes to the same, a special zero state of the intervention variables, because there is not always a natural zero state – in the Newcomb situation you have to take one or two boxes, you cannot just let it go – and because non-interference or refraining seemed to me to be an action as well. One could, however, distinguish some values of action variables as such zero states in my framework and

¹⁰ Of course, I have become aware that this observation does not exhaust the problem. It is a rich problem indeed, and at least in the iterated Newcomb problem I have converted to a one-boxer. Cf. Spohn (2000a).

¹¹ A distinction which has been observed also by Kyburg (1980).

thus define the unmanipulated graph in the sense of SGS as the subgraph determined by these action variables taking their zero states. Their manipulation theorem then simply states the recursive decomposition of probabilities characteristic of Bayesian nets and their slight generalization to a probability family $\{P_v\}$.¹²

Again, a crucial difference lies in the fact that SGS build a very detailed statistical theory of prediction (of the effects of intervention) on their basic definitions.¹³ Our basic agreement, however, is also displayed in our treatment of Newcomb's problem, where Meek, Glymour (1994, p. 1015) reach the same conclusion as the one I have sketched above.

To sum up: There is a large agreement between SGS and me in the formal basics of a probabilistic theory of causal dependence, including even the extension to actions or interventions. The main difference is that they abstain from any bold statement about what causation is, wisely so for their purposes, whereas I have advanced and argued for the, positive or negative, thesis that from an epistemological point of view the theory of Bayesian nets exhaust, with the caveats mentioned, the theory of causal dependence.

References

- Fishburn, P.C. (1964), *Decision and Value Theory*, New York: Wiley.
- Glymour, C., R. Scheines, P. Spirtes, K. Kelly (1987), *Discovering Causal Structure*, New York: Academic Press.
- Jeffrey, R.C. (1965), *The Logic of Decision*, Chicago: University Press, 2nd ed. 1983.
- Jeffrey, R.C. (1977), "A Note on the Kinematics of Preference", *Erkenntnis* 11, 135-141.
- Kyburg, H.E. jr. (1980), "Acts and Conditional Probabilities", *Theory and Decision* 12, 149-171.
- Meek, C., C. Glymour (1994), "Conditioning and Intervening", *British Journal for the Philosophy of Science* 45, 1001-1021.
- Nozick, R. (1969), "Newcomb's Problem and Two Principles of Choice", in: N Rescher et al. (eds.), *Essays in Honor of Carl G. Hempel*, Dordrecht: Reidel, pp. 114-146.
- Papineau, D. (1985), "Probabilities and Causes", *Journal of Philosophy* 82, 57-74.

¹² The comparison extends to Pearl (1998, sect. 4). which summarizes his work on the role of actions in Bayesian networks. His procedure superficially differs from SGS's. Instead of expanding the original to a manipulated graph he includes action variables in the original graph (which, however, may merely be observed, from outside, as it were), and for representing actions as choices enforcing a certain value of the action variables he mutilates the original graph by cutting out all edges ending in actions variables. The mutilation also leads to a changed probability distribution, the same as the one described by SGS in their manipulation theorem. In Spohn (1978, sect. 5.2) I considered the very same problem – how to turn a theoretically detached view of a set of variables which does not give action variables a special role into a practically relevant view which does respect the special role of actions for the agent? – and I arrived at the very same cutting procedure.

¹³ This remark applies mutatis mutandis to the work of Judea Pearl.

- Papineau, D. (1989), "Pure, Mixed, and Spurious Probabilities and Their Significance for a Reductionist Theory of Causation", in: P. Kitcher, W.C. Salmon (eds.), *Minnesota Studies in the Philosophy of Science, vol. XIII, Scientific Explanation*, Minneapolis: University of Minnesota Press, pp. 307-348.
- Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo: Kaufmann.
- Pearl, J. (1998), "Graphical Models for Probabilistic and Causal Reasoning", in: D.M. Gabbay, P. Smets (eds.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1*, Dordrecht: Kluwer, pp. 367-389.
- Putnam, H. (1980), "Models and Reality", *Journal of Symbolic Logic* 45, 464-482.
- Railton, P. (1978), "A Deductive-Nomological Model of Probabilistic Explanation", *Philosophy of Science* 45, 206-226.
- Salmon, W.C. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton: University Press.
- Spirtes, P., C. Glymour, R. Scheines (1993), *Causation, Prediction, and Search*, Berlin: Springer.
- Spohn, W. (1976/8), *Grundlagen der Entscheidungstheorie*, Dissertation at the University of Munich 1976, published: Kronberg/Ts.: Scriptor, 1978 (references refer to the published version).
- Spohn, W. (1977), "Where Luce and Krantz Do Really Generalize Savage's Decision Model", *Erkenntnis* 11, 113-134.
- Spohn, W. (1980), "Stochastic Independence, Causal Independence, and Shieldability", *Journal of Philosophical Logic* 9, 73-99.
- Spohn, W. (1983), *Eine Theorie der Kausalität*, unpublished Habilitationsschrift, University of Munich.
- Spohn, W. (1988), "Ordinal Conditional Functions. A Dynamic Theory of Epistemic States", in: W.L. Harper, B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics vol. II*, Dordrecht: Kluwer, pp. 105-134.
- Spohn, W. (1990), "Direct and Indirect Causes", *Topoi* 9, 125-145.
- Spohn, W. (1994), "On the Properties of Conditional Independence", in: P. Humphreys (Hg.), *Patrick Suppes: Scientific Philosopher. Vol. 1: Probability and Probabilistic Causality*, Dordrecht: Kluwer, pp. 173-194.
- Spohn, W. (2000a), "A Rationalization of Cooperation in the Iterated Prisoner's Dilemma", in: J. Nida-Rümelin, W. Spohn (eds.), *Practical Rationality, Rules, and Structure*, Dordrecht: Kluwer.
- Spohn, W. (2000b), "Deterministic Causation", in: W. Spohn, M. Ledwig, M. Esfeld (eds.), *Current Issues in Causation*, Paderborn: Mentis.
- Suppes, P. (1970), *A Probabilistic Theory of Causality*, Amsterdam: North-Holland.