

ON THE PROPERTIES OF CONDITIONAL INDEPENDENCE

ABSTRACT. As the paper explains, it is crucial to epistemology in general and to the theory of causation in particular to investigate the properties of conditional independence as completely as possible. The paper summarizes the most important results concerning conditional independence with respect to two important representations of epistemic states, namely (strictly positive) probability measures and natural conditional (or disbelief or ranking) functions. It finally adds some new observations.

In 1978, I submitted my manuscript 'Stochastic Independence, Causal Independence, and Shieldability' to the *Journal of Philosophical Logic*. Having realized that the laws of conditional probabilistic independence are the basis of any theory of probabilistic causality and trying to explain how in my view that theory should be built upon this basis I started the manuscript by proving as many of the laws of conditional and also of unconditional probabilistic independence as I knew and conjectured that the two lists were indeed complete; these conjectures looked very plausible to me after half a year of unsuccessful attempts to prove them. The editor responded that I should rewrite the paper in the light of the referee's comments and attached a report of the utmost value which was four pages long; I have not received so long and careful a review since then. My first guess turned out to be right later on: Pat Suppes was the referee.

Referring to these conjectures Pat Suppes wrote that they "are almost surely incorrect ... It would be very surprising if the properties he mentions would be sufficient ... I think the remarks should be deleted." And so I did.

In this note, now, I would like to return to the issue and to briefly summarize essential parts of the progress made since then. We shall see that Pat Suppes's remark was proved to be mainly correct only ten years later, and we shall see that the progress made is satisfying in essential respects, but still incomplete in the general case. The results to be reported are mainly those of the group around Judea Pearl; but others

also made important contributions, and I can add some news which make the general case less hopeless than it presently looks.

What was the content of the retracted conjectures? It was pure probability theory. Let a probability space $\langle \Omega, \mathcal{A}, P \rangle$ be given; Ω is a space of possibilities, \mathcal{A} an algebra of propositions or events over Ω ,¹ and P a probability measure. Infinity will be irrelevant; thus we may assume the possibility space Ω to be finite, and then it does not matter whether \mathcal{A} is a Boolean or a σ -algebra and whether P is finitely or σ -additive. Just for the sake of completeness let us define probabilistic independence in

DEFINITION 1. Let A , B and C be three events. Then A and B are *independent* w.r.t. P , symbolically $A \perp_P B$, iff $P(A \cap B) = P(A)P(B)$; and if $P(C) \neq 0$, then A and B are *independent conditional on C* w.r.t. P , symbolically $A \perp_P B/C$, iff $P(A \cap B | C) = P(A | C)P(B | C)$.

Conditional independence among events obeys

THEOREM 1.

- (a) If $A \perp_P B/C$, then $B \perp_P A/C$;
- (b) if $P(C) \neq 0$ and $A \cap C \subseteq B \cap C$, then $A \perp_P B/C$ iff $P(A | C) = 0$ or $P(B | C) = 1$;
- (c) if A and A' are disjoint and $A \perp_P B/C$, then $A \cup A' \perp_P B/C$ iff $A' \perp_P B/C$;
- (d) if $A \perp_P C/D$, then $A \perp_P B \cap C/D$ iff $A \perp_P B/C \cap D$.

Any other valid statement I know of which is stated purely in terms of conditional independence and extremal probabilities is entailed by Theorem 1.

However, independence among events is not the most important kind of independence; an independence between two events may be insignificant because it may arise accidentally, so to speak. The firmer and more interesting kind of independence is that between random variables and sets of them, which is my proper concern in the sequel.

So, let a finite set I of random variables on Ω be given; I shall use X , Y and Z to denote single variables and J , K , L and M to denote subsets of I . For any $J = \{X_1, \dots, X_n\} \subseteq I$ let a J -state be any event of the form $\{\omega | X_1(\omega) = x_1, \dots, X_n(\omega) = x_n\}$ specifying one way for X_1, \dots, X_n to be realized. We assume the possible states of disjoint sets of variables to be logically independent, i.e. if, for $k = 1, \dots, n$, A_k

is an X_k -state and if each A_k is a possible event, so is their intersection. Then we may generalize Definition 1 to

DEFINITION 2: Let K , L , and M be three mutually disjoint sets of random variables. Then K and L are *independent* w.r.t. P , symbolically $K \perp_P L$, iff $A \perp_P B$ for all K -states A and L -states B ; and K and L are *independent conditional on M* w.r.t. P , symbolically $K \perp_P L/M$, iff $A \perp_P B/C$ for all K -states A , L -states B , and M -states C with $P(C) \neq 0$.

In informal terms $K \perp_P L/M$ states that given complete knowledge about the variables in M , nothing concerning the variables in L can be inferred or induced from gaining information about the variables in K (and vice versa). By considering only mutually disjoint sets of variables, I neglect per definition the formally trivial, but intuitively somewhat confusing case of self-independence.

Unconditional independence between sets of variables is characterized by

THEOREM 2.

- (U1) If $K \perp_P L$, then $L \perp_P K$ (Symmetry),
- (U2) $K \perp_P \emptyset$ (Trivial Independence),
- (U3) if $K \perp_P L \cup M$, then $K \perp_P L$ (Decomposition),
- (U4) if $K \perp_P L$ and $K \cup L \perp_P M$, then $K \perp_P L \cup M$ (Mixing).

My minor conjecture has been that this characterization is complete, as expressed in

THEOREM 3. *Each binary relation between disjoint subsets of I satisfying (U1–U4) is unconditional independence w.r.t. some probability measure.*

However, this theorem was proved to hold only by Geiger, Paz, and Pearl (1988), and as Milan Studený pointed out to me, by Matúš (1988). This is a remarkable positive result.

Conditional independence is much more intricate, and it will occupy us for the rest of the paper. For its characterization it is useful to introduce

DEFINITION 3. A ternary relation \perp between disjoint subsets of I is called a *semi-graphoid* (over I) iff it satisfies

- (G1) If $K \perp L/M$, then $L \perp K/M$ (Symmetry),
- (G2) $J \perp \emptyset/M$ (Trivial Independence),
- (G3) if $J \perp K \cup L/M$, then $J \perp K/M$ (Decomposition),
- (G4) if $J \perp K \cup L/M$, then $J \perp K/L \cup M$ (Weak Union), and
- (G5) if $J \perp K/L \cup M$ and $J \perp L/M$, then $J \perp K \cup L/M$ (Contraction).²

And it is called a *graphoid* (over I) iff it additionally satisfies

- (G6) if $J \perp K/L \cup M$, and $J \perp L/K \cup M$, then $J \perp K \cup L/M$ (Intersection).

These terms and the labels of the properties are due to Pearl and Paz (1985). They derive from the rich and useful graph theoretic connections of independence notions which, however, need not to be displayed here.

In these terms conditional independence is characterized by

THEOREM 4. For any probability measure P , \perp_P is a semi-graphoid. If P is strictly positive, i.e. if $P(A) = 0$ only for $A = \emptyset$, then \perp_P is a graphoid.³ Moreover, we have for any P :

- (B) if X is a binary variable, i.e. takes only two values, and if $K \perp_P L$ and $K \perp_P L/\{X\}$, then $K \cup \{X\} \perp_P L$ or $K \perp_P L \cup \{X\}$.

My major conjecture was that this characterization is complete as well, i.e. that, more precisely, for any semi-graphoid or graphoid there is, respectively, a probability measure or a strictly positive probability measure such that the semi-graphoid or graphoid is conditional independence with respect to that measure – provided all the variables take more than two values; the proviso is to circumvent the somewhat disturbing special law (B) for binary variables. Pearl and Paz (1985) had the very same conjecture. Alas, we were all wrong.

Before proceeding to this negative result and its partial positive substitutes we have to introduce, however, another very useful model of the graphoid properties of independence. There are subjective probabilities guiding betting behavior and (all?) other actions. But there is also *plain belief*, at least we talk of it all the time; we believe things to be true or to be false, or we suspend opinion. Plain belief cannot be adequately represented in terms of subjective probability. So what to do about it?

One may disregard it; but we had better account for it. A fully conditionalizable and hence dynamizable account of plain belief is, I think, best given in the following terms:

DEFINITION 4. κ is a *natural conditional function* or *NCF*⁴ (over Ω) iff κ is a function from Ω into the set of natural numbers⁵ such that $\kappa^{-1}(0) \neq \emptyset$. An NCF κ is extended to propositions by defining $\kappa(A) = \min\{\kappa(\omega) \mid \omega \in A\}$ for each $A \neq \emptyset$ and $\kappa(\emptyset) = \infty$.⁶ Moreover, for $A \neq \emptyset$ we can define *conditional* κ -values by setting $\kappa(\omega \mid A) = \kappa(\omega) - \kappa(A)$ for $\omega \in A$ and $\kappa(B \mid A) = \min\{\kappa(\omega \mid A) \mid \omega \in A \cap B\} = \kappa(A \cap B) - \kappa(A)$.

An NCF κ is to be interpreted as a *grading of disbelief*. $\kappa(\omega) = 0$ says that ω is not disbelieved, i.e. ω might be realized according to κ ; because not every possibility can be held to be unrealized, Definition 4 requires that $\kappa(\omega) = 0$ for some $\omega \in \Omega$. And $\kappa(\omega) = n > 0$ says that ω is disbelieved with degree n . A proposition is then assigned the minimal degree of disbelief of its members. Thus, if $\kappa(A) = 0$, A is not disbelieved, i.e. A might be true according to κ ; and if $\kappa(A) = n > 0$, A is disbelieved, and $\sim A$ ⁷ believed, with degree n . Hence, precisely the supersets of $\kappa^{-1}(0)$ are plainly believed according to κ . In a similar way, conditional κ -values express conditional belief and disbelief. On that basis, conditionalization rules analogous to those discussed in probabilistic epistemology finally allow for a dynamic theory of plain belief.⁸

NCFs may also be understood in terms of non-standard probability measures, namely as assigning to propositions the infinitesimal order of magnitude of their probability (relative to some given infinitesimal); this maps products of probabilities into sums of κ -values, divisions into subtractions, and sums into minima.⁹ So, in a way, it is no wonder that NCFs behave very similar to probability measures. In particular, if there is conditionalization, there is also independence:

DEFINITION 5. Let K , L and M be three mutually disjoint sets of variables. Then K and L are *independent* w.r.t. to the NCF κ , symbolically $K \perp_{\kappa} L$, iff for all K -states A and L -states B $\kappa(A \cap B) = \kappa(A) + \kappa(B)$. Further, K and L are *independent conditional on M* w.r.t. κ . Symbolically $K \perp_{\kappa} L/M$, iff for all K -states A , L -states B , and M -states C

$$\kappa(A \cap B \mid C) = \kappa(A \mid C) + \kappa(B \mid C).$$

Conditional independence w.r.t. NCFs is characterized by

THEOREM 5. *For any NCF κ , \perp_κ is a graphoid.*¹⁰

There is no need to assume an analogue to strict positivity; NCFs are automatically strictly positive, so to speak (by not taking ∞ as a value). Also, the special law (B) does not hold for NCFs because, whereas $P(A)$ and $P(\sim A)$ are functionally related, $\kappa(A)$ and $\kappa(\sim A)$ are not; there is only the condition that at least one of them must be 0.

What is the use of knowing Theorems 4 and 5 and whether or not they are complete? There is, first, mathematical curiosity to be satisfied. There is, moreover, a general philosophical interest. It is my experience that the notion of independence is one of the most problematic and confusing notions throughout the entire history of philosophy;¹¹ and then it is important to distinguish its various meanings and to uncover their formal properties which are able to guide its quite unreliable use. This has been good philosophical methodology in other cases as well.

But there are more specific reasons for knowing the properties of epistemic independence which we are considering here in the form of probabilistic and NCF-independence. In probabilistic models of epistemic states, which are doubtlessly the most important and successful ones, conditional and unconditional independencies provide a qualitative skeleton for epistemic states which is of vital importance. It is needed in AI for making probability measures computationally manageable. It is needed for describing the influence of experience which strikes the epistemic structure at one point and then spreads along its paths; these are also the paths along which we proceed from suppositions to conclusions. The skeleton is needed to determine which parts are unaffected by local changes in other parts; it thus helps to modularize basically appropriate, but undifferentiated and therefore theoretically unfruitful holistic accounts of epistemic states. And so forth. This fundamental epistemological role of probabilistic independence is displayed in Pearl (1988) in beautiful abundance. The very same remarks apply to any other useful model of epistemic states and its associated independence notion and in particular to NCFs and NCF-independence. It is obvious, then, how important it is to know the formal properties of the various

explications of conditional and unconditional epistemic independence as completely as possible.

There is another specific reason. It is arguable that *causal* dependence and independence is basically a kind of conditional epistemic dependence and independence,¹² a view which I share and tried to support since my (1980).¹³ This entails that one can derive the properties of causal dependence from those of conditional epistemic dependence. Thus, again, it is important to know the latter, both for inferring the properties of causal dependence and for checking the plausibility of these inferences and thus of that view.¹⁴ This reason applies equally to probabilistic and to NCF-independence, since in my view a theory of deterministic causation can be best spelled out in terms of NCFs in strict parallel to a theory of probabilistic causation.¹⁵

The two reasons given are not independent, of course; they rather mesh, because it is quite plausible that the epistemic skeleton is provided not by conditional epistemic dependence in general, but more specifically by causal dependence. In fact, the theory of Bayesian networks is developed just from this assumption.¹⁶

So, what more can be said about conditional epistemic independence beyond Theorems 4 and 5? A lot; indeed there is no point in summarizing here all the known results. Therefore I confine myself to presenting only the results most relevant to the purposes just mentioned.

There are, first, some quite strong partial completeness results. In order to understand them, we need

DEFINITION 6. A *list of total causes*¹⁷ is a set of conditional independence statements for which there is a linear ordering X_1, X_2, \dots of all variables in I such that the list contains for each variable X_k exactly one statement of the form $X_k \perp \{X_1, \dots, X_{k-1}\} - J/J$ for some $J \subseteq \{X_1, \dots, X_{k-1}\}$. A semi-graphoid is then said to be *generated* by a list of total causes iff it is the closure of that list under the semi-graphoid conditions (G1–G5); analogously, graphoids may be generated in this way.

In this special case the semi-graphoid properties suffice to characterize conditional probabilistic independence. This was shown by Geiger and Pearl (1990) by proving

THEOREM 6. *For each semi-graphoid \perp generated by a list of total causes there is a probability measure P such that $\perp_P = \perp$.*¹⁸

Concerning NCFs Hunter (1991) proved the corresponding

THEOREM 7. *For each semi-graphoid \perp generated by a list of total causes there is a NCF $_{\kappa}$ such that $\perp_{\kappa} = \perp$.*¹⁹

Interestingly, this immediately entails the further

THEOREM 8. *Each semi-graphoid generated by a list of total causes is the same as the graphoid generated by that list.*

This is so simply because \perp_{κ} is already a graphoid according to Theorem 5.²⁰ It should also be possible, it seems, to prove this directly without recourse to an interpretation of \perp in terms of probabilities or NCFs.

The appropriateness of the term ‘list of total causes’ and the impact of these theorems gets a bit more perspicuous when we give them a slightly different form. We first need to know how the relation between causal and epistemic dependence is to be specified. The relation for which I have argued in (1980, 1983, and 1990b) is stated in

DEFINITION 7. Assume a linear temporal ordering of the variables in I . Then the variable X is *potentially directly causally relevant* to the variable Y iff X precedes Y and X and Y are epistemically dependent conditional on M , where $M = \{Z \in I \mid Z \text{ precedes } Y \text{ and } Z \neq X\}$, i.e. $\text{not } Y \perp_P X/M$ in the probabilistic case and $\text{not } Y \perp_{\kappa} X/M$ in the deterministic case.

It is quite obvious that this explication grew out of Suppes’ basic work (1970), and its main content is almost self-explanatory. Still it requires some comments.

The definition is relative to a given set I of variables and to a given P or κ . This may seem unwelcome, but it is unavoidable, I think, and I subscribe here to that relativization.²¹ Another issue not discussed here

is how the explication is to be modified when the temporal ordering is assumed to be only weak and not linear so that simultaneous variables are allowed.

Since we are considering causal relations only between variables, it is adequate to speak of causal relevance or dependence; it is only events or facts of which one can say that one *causes* the other. For the same reason the qualification 'potential' is appropriate. The condition that not $Y \perp_P X/M$ only says that there is some possible M -state conditional on which X and Y are dependent; but this need not be the actual one. This makes clear that actual causal relations cannot be studied purely on the level of variables.²²

Moreover, it is obvious that the definition can characterize only direct causal dependence (as relativized to the given set I of variables). If X were only indirectly causally relevant to Y , then the set of variables preceding Y and different from X would contain all the variables mediating between X and Y and would thus render X and Y conditionally independent. The simplest and indeed the right thing to say is that (direct or indirect) causal dependence is the transitive closure of direct causal dependence; but there are counter-arguments which are not easily dissolved.²³

Finally, it must be emphasized that the probabilistic part of the explication is adequate only if P is strictly positive; if it is not, if it contains deterministic ingredients, so to speak, conditional probabilities required for causal judgment may be undefined. In this case, one has, in my view, to resort to Popper measures or rather something more complicated which is, in effect, a combination of probability measures and NCFs.²⁴ This view is related to my more complicated view of deterministic causation. There is a long standing tendency,²⁵ particularly in the statistical literature, to equate deterministic causal dependence with functional dependence in the mathematical sense. This is not false, but it is not the basic sense of deterministic causation and it requires favorable circumstances and some steps of reasoning to reduce causal to functional dependence.²⁶

So let us accept Definition 7 for the sequel, and let us assume that the linear ordering X_1, X_2, \dots of the variables in I is their temporal ordering. On this basis we can introduce

DEFINITION 8. A *list of direct causes*²⁷ is just any set of pairs $\langle X_i, X_k \rangle$ with $i < k$ or of conditional dependence statements corresponding to

such pairs according to Definition 7. A graphoid is then said to be *generated* by a list of direct causes iff it is the closure under the graphoid conditions (G1–G6) of the complementary set of *independence* statements, i.e. of the statements $X_k \perp X_i / \{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{k-1}\}$ for all pairs $\langle X_i, X_k \rangle$ with $i < k$ which are not in the list.

There is an obvious one–one correspondence between lists of total causes and lists of direct causes. If J is the conditioning set of the entry for X_k in a given list of total causes, then J consists just of all the variables directly causally relevant to X_k according to the corresponding list of *direct* causes, i.e. $X_i \in J$ iff $\langle X_i, X_k \rangle$ is in that corresponding list; so J may well be called a *total* cause of X_k .²⁸ The important fact about this correspondence, of which the proponents of Definition 6 were of course fully aware, is that the independence statements associated with corresponding lists mutually imply each other. The entry $X_k \perp \{X_1, \dots, X_{k-1}\} - J / J$ for X_k in a given list of total causes entails with the help of (G4) that $X_k \perp X_i / \{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{k-1}\}$ for each $X_i \in \{X_1, \dots, X_{k-1}\} - J$, i.e. for each X_i with $i < k$ such that $\langle X_i, X_k \rangle$ is not in the corresponding list. Conversely, the latter independencies jointly imply the former with the help of the special graphoid property (G6). These implications entail that the graphoid generated by a list of direct causes is the same as the graphoid generated by the corresponding list of total causes. Combining this observation with Theorems 6, 7, and 8, we finally get

THEOREM 9. *For each graphoid \perp generated by a list of direct causes there is a probability measure P such that $\perp_P = \perp$ and an NCF κ such that $\perp_\kappa = \perp$.*

This entails the further side result that each set containing for all X_i, X_k with $i < k$ either $X_k \perp X_i / \{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{k-1}\}$ or its negation is consistent, i.e. that any hypothesis about a direct causal dependency or independency is logically independent of any set of other such hypotheses. This is noteworthy because it says that causal hypotheses are acceptable and rejectable without logical constraints by the others and thus facilitates them to play the fundamental role attributed to them.

What has been done by deriving Theorem 9 is simply to more fully explicate the intended causal content of Theorems 6 and 7. If the relation between direct causal dependence and conditional epistemic dependence is as specified in Definition 7, these theorems are fully satisfying to the extent to which the important epistemological functions mentioned above are performed just by the causal independencies among the conditional epistemic independencies. In that case, we need to know only the graphoid properties (G1–G6) and can even restrict the use of (G6) to the transition from the given list of direct causes to the corresponding list of total causes.

This, in my view, is the most important partial completeness result. But there are others as well: about so-called saturated independence statements the three relata of which have to add to some fixed set of variables, about special probability measures like normal distributions or binary distributions, for the case that so-called deterministic variables are allowed, with respect to other independence notions, concerning stronger completeness notions, etc. Since there is no point in repeating them here, I refer the reader to Geiger and Pearl (1988), and Geiger (1990), and Geiger, Verma, and Pearl (1990).²⁹

Still it is not wholly clear whether among the conditional epistemic independencies only the causal ones and their implications are epistemologically fundamental. I am not going to discuss that question; but it is clear that in the case of doubt it would be epistemologically important to know further properties of conditional independence and to have a general completeness result. So let us finally turn to this issue.

I already mentioned that the old completeness conjectures about graphoids and semi-graphoids have turned out to be false. The first one to show this and thus to confirm Pat Suppes's remark in his review was Studený (1989a), p. 15. Making essential use of information theoretic means Studený was able to prove the following quite complex property for each probability measure P :

- (S1) $K \perp_P K'$, $L \perp_P L'/K$, $L \perp_P L'/K'$, and $K \perp_P K'/L \cup L'$ if and only if $L \perp_P L'$, $K \perp_P K'/L$, $K \perp_P K'/L'$, and $L \perp_P L'/K \cup K'$.

Surprisingly, this property does not apply to NCFs; a counterexample is given in the Appendix. Thus, the analogy between probability measures and NCFs is not complete. Indeed, it is not clear to me how information theoretic notions may be transferred to NCF theory, nor whether this makes any sense.

In (1989b) Studený mentions a further property which is also about two pairs of sets of variables and the eight independences built from them and which holds only for strictly positive probability measures:

(S2) if $K \perp K' / L \cup L'$, $K \perp K' / L$, $K \perp K' / L'$, $L \perp L' / K \cup K'$, $L \perp L' / K$, and $L \perp L' / K'$, then $K \perp K'$ and $L \perp L'$.

Of course, (S1) and (S2) also hold when all statements are further conditionalized on some constant set M of variables. The fact that these new properties were found after several years of unsuccessful search was certainly a slight sensation.

But it got even worse. Having tasted the flavor of the new properties, Geiger and Pearl (1988, p. 41) mention a whole new family of properties applying to strictly positive probability measures, namely:

(S3) Let $J = \{X_0, \dots, X_{2n-1}\}$, and suppose that for each $i = 1, \dots, 2n$:

(a) $X_i \perp J - \{X_{i-1}, X_i, X_{i+1}\} / \{X_{i-1}, X_{i+1}\}$,

(b) $X_{i-1} \perp X_{i+1} / X_i$,

(c) $X_{i-1} \perp X_{i+1}$ (subscripts taken modulo $2n$).

Then $\perp \{X_0, X_2, \dots, X_{2n-2}\}$ and $\perp \{X_1, X_3, \dots, X_{2n-1}\}$. Here, $\perp K$ means that K is totally independent, i.e. that each member of K is independent of the rest of K .

This is an awfully complicated condition, and apparently one must even fear that it provides a new, logically independent property for each n .

Studený (1992, p. 382), presents a further, simpler family of properties applying to *all* probability measures:

(S4) Let $\{X_0, \dots, X_{n-1}\}$ be n variables ($n \geq 4$). Then $X_0 \perp X_i / X_{i+1}$ holds for all $i = 1, \dots, n-1$ if and only if $X_0 \perp X_{i+1} / X_i$ holds for all $i = 1, \dots, n-1$ (subscripts taken modulo n).

As is proved in the Appendix, (S4) applies as well to NCFs. The fear concerning (S3) is in fact realized with respect to (S4). For, Studený (1992, p. 383), has a beautiful proof of

THEOREM 10. (S4) *for n is logically independent of all properties of probabilistic conditional independence of the form "if $K_1 \perp L_1 / M_1, \dots, K_{r-1} \perp L_{r-1} / M_{r-1}$, then $K_r \perp L_r / M_r$ " with $r < n$.*

Studeny concludes from this theorem that there is no finite axiomatization of all the properties of probabilistic conditional independence.³⁰

Thus the situation looks pretty confusing, if not hopeless. But two observations slightly reduce the confusion. First, there is still another property which holds for all strictly positive probability measures as well as for all NCFs:

(S5) If $L \perp L' / K \cup K' \cup M$ and if three of the four independencies $K \perp K' / M$, $K \perp K' / L \cup M$, $K \perp K' / L' \cup M$, and $K \perp K' / L \cup L' \cup M$ hold, then the fourth holds as well.

This time I have explicitly added a constant conditioning set M . The proof of (S5) is given in the appendix. Clearly, (S5) is a generalization of (S2).

Secondly, we have

THEOREM 11. *For each n , (S3) is entailed by (G1–G6) and (S5).*

Again, the proof is given in the Appendix. Hence, (S5) deprives (S3) of its logical independence, and we need not worry about (S3) any more.

To summarize: what we have so far for probabilistic conditional independence are the old semi-graphoid conditions (G1–G5), the special law (B) for binary variables, and the new properties (S1) and (S4), the latter being in fact a whole family. Additionally we have (G6) and (S5) applying only to strictly positive measures.³¹ NCFs behave in a somewhat simpler way; conditional independence with respect to them is uniformly characterized by (G1–G6), (S4), and (S5), since we do not have to distinguish different kinds of NCFs. So the situation is still quite complicated; and having been wrong once, I shall not dare to make a new conjecture.

However, a completeness theorem has not yet been shown to be impossible. If such a theorem were feasible, it seems more likely that it would be provable with respect to NCFs, since NCFs are mathematically much simpler than probability measures. In particular, since the condition (B) does not apply to NCFs, one may expect that the models one has to find in a completeness proof can be restricted to binary variables. Such a result concerning NCFs may even help with the probabilistic case, provided the alleged homomorphism between NCFs and infinitesimal probability measures still holds.

Indeed, there is hope – at least in principle. The following claims look at least plausible: for each finite set I of variables taking only finitely many values a number n can be effectively specified so that any finite truth-functional compound p of conditional independence statements built from subsets of I is made false by an NCF taking values only up to n , if p is made false by some NCF at all. And similarly, for each such set I a number n can be effectively specified so that any such statement p is made false by a probability measure taking only fractions with denominators up to n as values, if p is made false by some probability measure at all. If these claims were true, then one would have to check only finitely many NCFs or probability measures in order to determine the validity of such a statement p . Thus, the set of all valid statements of this form would be decidable and hence, a fortiori, recursively axiomatizable. As Studený admits, this would not contradict Theorem 10, since finite axiomatizability is only a special case of recursive axiomatizability. So, if these claims were true, the existence of a complete characterization of conditional independence would be assured. One would only have to find it.

APPENDIX

A COUNTER-EXAMPLE TO (S1) FOR NCFs. Let X, Y, Z , and W be four binary variables assuming, respectively, the states A, A', B, B', C, C', D , and D' ; and let the NCF κ be defined for the 16 conjunctions built from these states by the following matrix:

κ	$C \cap D$	$C \cap D'$	$C' \cap D$	$C' \cap D'$
$A \cap B$	0	1	1	2
$A \cap B'$	2	6	6	10
$A' \cap B$	8	5	5	2
$A' \cap B'$	7	7	7	7

We then have $Z \perp_{\kappa} W$, $X \perp_{\kappa} Y/Z$, $X \perp_{\kappa} Y/W$, and $Z \perp_{\kappa} W/\{X, Y\}$, but nothing of $X \perp_{\kappa} Y$, $Z \perp_{\kappa} W/X$, $Z \perp_{\kappa} W/Y$, and $X \perp_{\kappa} Y/\{Z, W\}$.

Proof of (S4) for NCFs. I state the proof only for $n = 4$ binary variables X_0, X_1, X_2 , and X_3 assuming, respectively, the states A, A', B, B', C, C', D , and D' . Let an NCF κ be given and define

$$\begin{aligned} E_1 &= A \cap B \cap C \cap D, & F_1 &= B \cap C \cap D, \\ E_2 &= A \cap B \cap C \cap D', & F_2 &= B \cap C \cap D', \\ E_3 &= A \cap B \cap C' \cap D, & F_3 &= B \cap C' \cap D, \\ E_4 &= A \cap B \cap C' \cap D', & F_4 &= B \cap C' \cap D', \\ E_5 &= A \cap B' \cap C \cap D, & F_5 &= B' \cap C \cap D, \\ E_6 &= A \cap B' \cap C \cap D', & F_6 &= B' \cap C \cap D', \\ E_7 &= A \cap B' \cap C' \cap D, & F_7 &= B' \cap C' \cap D, \\ E_8 &= A \cap B' \cap C' \cap D', & F_8 &= B' \cap C' \cap D', \end{aligned}$$

$$\text{and } x_{rs} = \min(\kappa(E_r), \kappa(E_s)) - \min(\kappa(F_r), \kappa(F_s)).$$

Then $X_0 \perp_{\kappa} X_1 / X_2$ implies that

$$\begin{aligned} \kappa(A \mid C) &= \kappa(A \mid B \cap C) \\ &= \kappa(A \cap B \cap C) - \kappa(B \cap C) = x_{12} \\ &= \kappa(A \mid B' \cap C) \\ &= \kappa(A \cap B' \cap C) - \kappa(B' \cap C) = x_{56}. \end{aligned}$$

and similarly that

$$\kappa(A \mid C') = x_{34} = x_{78}.$$

In the same way, $X_0 \perp_{\kappa} X_2 / X_3$ and $X_0 \perp_{\kappa} X_3 / X_1$ imply that

$$\begin{aligned} \kappa(A \mid D) &= x_{15} = x_{37}. \\ \kappa(A \mid D') &= x_{26} = x_{48}. \\ \kappa(A \mid B) &= x_{13} = x_{24}. \\ \kappa(A \mid B') &= x_{57} = x_{68}. \end{aligned}$$

Clearly, we cannot say anything about the order of the κ -values of the E_r and the F_r except that $\kappa(E_r) \geq \kappa(F_r)$ for each r . But whatever this order is, we find that

$$x_{12} = x_{56} \text{ is between } x_{15} \text{ and } x_{26}$$

(this is to include the case where the first is equal to one of the latter two), and similarly that

$$\begin{aligned}
 x_{34} &= x_{78} \text{ is between } x_{37} \text{ and } x_{48}, \\
 x_{15} &= x_{37} \text{ is between } x_{13} \text{ and } x_{57}, \\
 x_{26} &= x_{48} \text{ is between } x_{24} \text{ and } x_{68}, \\
 x_{13} &= x_{24} \text{ is between } x_{12} \text{ and } x_{34}, \text{ and} \\
 x_{57} &= x_{68} \text{ is between } x_{56} \text{ and } x_{78}.
 \end{aligned}$$

As is easily confirmed, these inequalities imply that the x_{rs} considered are in fact all the same. So we have arrived at the conclusion that

$$\begin{aligned}
 \kappa(A \mid B) &= \kappa(A \mid B') = \kappa(A \mid C) = \kappa(A \mid C') \\
 &= \kappa(A \mid D) = \kappa(A \mid D') = \kappa(A).
 \end{aligned}$$

Since the same reasoning applies to A' , we have $X_0 \perp_{\kappa} X_i$ for $i = 1, 2, 3$. (G5) applied to this and the premises yields $X_0 \perp_{\kappa} \{X_i, X_j\}$ for $i \neq j = 1, 2, 3$; and from this the desired result finally follows with the help of (G4).

I assume that the same proof goes through, though in a still clumsier way, when more than four and not necessarily binary variables are considered.

Proof of (S5). Let A, A', B, B' , and C vary over states taken, respectively, by K, K', L, L' , and M . The first premise says that

$$\begin{aligned}
 (*) \quad & P(A \cap A' \cap B \cap B' \mid C) P(A \cap A' \mid C) \\
 &= P(A \cap A' \cap B \mid C) P(A \cap A' \cap B' \mid C).
 \end{aligned}$$

Now suppose that, say, the first three of the four mentioned independencies hold for P . This means that the last three terms of (*) factor out with respect to A and A' . Hence, (*) implies that $P(A \cap A' \cap B \cap B' \mid C)$ factors out in the same way; i.e. for each given B and B' there are functions $f_{BB'}$ and $g_{BB'}$ such that for all A and A' $P(A \cap A' \mid B \cap B' \cap C) = f_{BB'}(A)g_{BB'}(A')$. Summing up over all A' or A , respectively, shows that $f_{BB'}$ and $g_{BB'}$ can be chosen in a normalized way so that $f_{BB'}(A) = P(A \mid B \cap B' \cap C)$ and $g_{BB'}(A') = P(A' \mid B \cap B' \cap C)$. Hence $K \perp_P K' / L \cup L' \cup M$. Mutatis mutandis, the same proof applies, if three others of the four independencies are given; in fact, as Milan Studený has pointed out to me, strict positivity of P is needed only in these other cases. The analogous proof goes through for NCFs.

Proof of Theorem 11. We have to show that the conclusion of (S3) follows from the premises of (S3) with the help of (G1–G6) and (S5). So assume (a), (b), (c) of (S3). Take subscripts always modulo $2n$, and let us simplify the notation by writing, e.g., $i \perp J - \{i-1, i, i+1\} / \{i-1, i+1\}$ for $X_i \perp J - \{X_{i-1}, X_i, X_{i+1}\} / \{X_{i-1}, X_{i+1}\}$. The assumptions first imply:

- (d) $\{i+1, \dots, j-1\} \perp \{j+1, \dots, i-1\} / \{i, j\}$ for all $i, j = 0, \dots, 2n-1$ with $i+2 \leq j$.

This can be proved by induction: for $j = i+2$, (d) is the same as (a). Now suppose that (d) holds for all i and $j = i+k$ ($k \geq 2$). Thus we have

$$\{i+1, \dots, i+k-1\} \perp \{i+k+1, \dots, i-1\} / \{i, i+k\}$$

and

$$\{i+2, \dots, i+k\} \perp \{i+k+2, \dots, i\} / \{i+1, i+k+1\}.$$

These two independencies imply with the help of (G4) and (G6)

$$\{i+1, \dots, i+k\} \perp \{i+k+2, \dots, i-1\} / \{i, i+k+1\}.$$

Thus, (d) holds also for i and $j = i+k+1$.

Moreover, (a)–(d) entail:

- (e) $i-1 \perp i+1/j$ for each $j \neq i-1, i, i+1$.

For, (a) implies $i \perp j / \{i-1, i+1\}$, (d) implies $i-1 \perp i+1 / \{i, j\}$, (b) says that $i-1 \perp i+1/i$, (c) says that $i-1 \perp i+1$, and these four independencies yield (e) with the help of (S5).

Now let k be an even index, $K = \{k, k+2, \dots, k+2m\}$ for some m , $L = \{k+2m+2, \dots, k-2\}$ the set of the remaining even indices, $i = k-1$, and $j = k+2m+1$. The next step is to show that

- (f) $K \perp L/j$.

This can be seen in the following way: for each $l \in L$, we have $l \perp l+2/j$ because of (e) and $l \perp \{l+4, \dots, k-2\} \cup K / \{l+2, j\}$ because of (d) (weakened by (G3)). These two independencies imply $l \perp \{l+2, \dots, k-2\} \cup K / j$ with the help of (G5) and thus $l \perp K / \{j, l+2, \dots, k-2\}$ with (G4). Repeated application of (G5) to this for each $l \in L$ yields (f).

Finally, we can infer for K , L , i , and j as just specified:

(g) $K \perp L$.

For, (a) implies $i \perp j / K \cup L$ with (G3+G4), (d) implies $K \perp L / \{i, j\}$ with (G3), and (f) says that $K \perp L / j$ and, by reversing the roles of i and j , also that $K \perp L / i$. All this entails (g) with the help of (S5). Since (g) holds for each even index k , the even numbered variables are completely independent. The same reasoning applies, of course, to the odd indices.

ACKNOWLEDGEMENTS

I am grateful to Judea Pearl, Dan Hunter, Azaria Paz, and in particular to Milan Studený for various helpful comments on a first draft of this paper.

*Abteilung Philosophie,
Universität Bielefeld,
D-33615 Bielefeld,
Germany*

NOTES

¹ The precise nature of the objects of probability, whether objectively or subjectively interpreted, is a deep and philosophically highly important issue which is still very open. Here it is irrelevant, and so I use 'event' and 'proposition' interchangeably.

² As Studený (1989a) has noted, (G3–G5) may be integrated into one property: $J \perp K \cup L / M$ iff $J \perp K / L \cup M$ and $J \perp L / M$.

³ This is proved in Dawid (1979) and (1980), and in my (1980) where strict positivity is replaced by a slightly weaker condition and where extensions to infinitely many variables may be found.

⁴ This is the label which I originally used in my (1988) in order to avoid false associations. More suggestive terms are 'disbelief function' (Shenoy, 1991) and 'ranking function' (Goldschmidt and Pearl, 1992).

⁵ In (1988) I allowed the range to consist of ordinal numbers and therefore talked of ordinal conditional functions; but this generality is not needed here.

⁶ Setting $\kappa(\emptyset) = \infty$ is a reasonable convention. But ∞ should not be assigned to possibilities and consistent propositions in my view because it causes all rules of belief revision to fail.

⁷ $\sim A$ of course denotes the complement of A relative to Ω .

⁸ A fuller motivation and explanation of this concept and the ensuing theory may be found in my (1988).

⁹ See my (1988), Section 7, and my (1990a), Theorem 2.

¹⁰ For proof see my (1988), Section 6, and Hunter (1991).

¹¹ For instance, the various forms of idealism stand and fall with a (in my view still lacking) satisfying explanation of what 'dependent' precisely means in their slogan 'the world is mind-dependent'.

¹² This view is an old one, of course. It was Hume's revolutionary thesis that causation is an epistemological notion, which he so convincingly elaborated in his (1739), Book I, Part 3. The importance of conditionalizing was emphasized at least by Mill who explicitly relativized causal statements to the obtaining circumstances or *ceteris paribus* clauses; cf. his (1843), Vol. I, Book III, Ch. 5.

¹³ Cf. in particular my (1990b). One may say that causal dependence is rather a kind of conditional metaphysical dependence. I do not object, but I think the proper view is the following: causal dependence as conceived by a given subject is definable in terms of conditional epistemic dependence according to the epistemic state of that subject; and objective causal dependence can be best understood as an objectification of that epistemic dependence; cf. my (1993).

¹⁴ This importance is well demonstrated in the remarkably rich statistical theory of causal inference and prediction presented in Glymour *et al.* (1991) and Spirtes *et al.* (1992), who build their theory on just two properties of causal dependence, the screening off and the common cause principle, generalized later on to the Causal Markov and the Causal Minimality condition (cf. Glymour *et al.*, 1991, pp. 156ff., and Spirtes *et al.*, 1993, pp. 53ff.). Starting one step earlier, namely from the definition of direct causal dependence mentioned below, one may derive both the screening off and the common cause principle; cf. my (1994), Section 4. Spirtes *et al.* do not take that step, I think, because the consequences are safer than the definition from which they are derived.

¹⁵ This view is to some extent elaborated in my (1991), Section 3.

¹⁶ Cf. Pearl (1988), Section 3.3.

¹⁷ The usual term is 'causal list'. Here I prefer to speak of lists of total causes, because, according to the explication given later on, this is what they are in contrast to the lists of direct causes introduced below.

¹⁸ To be precise: Geiger and Pearl (1990) showed for this special case that conditional probabilistic independence is completely characterized by so-called d-separability, a notion applying to directed acyclic graphs; it is not introduced here, but it is utterly useful for graphically presenting and computationally managing conditional independence; cf. Pearl (1988), pp. 119ff., and Spirtes *et al.* (1993), pp. 71ff. And Verma and Pearl (1990) showed that for a given list of total causes the relation of d-separability is identical with the semi-graphoid generated by the list. For an alternative characterization of d-separation see also Lauritzen *et al.* (1990).

¹⁹ In his proof, Hunter (1991) also proceeds via the equivalence of semi-graphoids and d-separability in this case; but the rest of the proof is quite different, because the probabilistic construction used by Geiger and Pearl (1990) does not carry over to NCFs.

²⁰ Theorem 8 follows only from Theorem 7 and not from Theorem 6. If we look at the proof of the latter by Geiger and Pearl (1990), we find that the probability measure which they construct in order to prove their existence claim is not strictly positive so that its independence relation is not a graphoid. Spirtes *et al.* (1993) have another proof of Geiger and Pearl's theorem in which they construct a suitable strictly positive measure, i.e. joint multivariate normal distribution; cf. their Lemma 3.5.7., pp. 390ff. So this

yields a probabilistic way to Theorem 8.

²¹ In the case of I , the obvious move to get rid of that relativization is to consider the ill-defined set of all variables whatsoever. Mathematically, this can be approximated by studying and comparing the causal relations within smaller and larger sets of variables. Verma and Pearl (1991), Pearl and Verma (1991), and Spirtes *et al.* (1993), Ch. 6, 10, and 11, have found several remarkable results about the way in which statistical data about a smaller set can be indicative of causal relations within a larger set of variables.

²² In my (1990b) I proceeded on the more fundamental level of events or states of affairs on the basis of which the causal relations between variables can be defined.

²³ Cf. my (1990b).

²⁴ Cf. my (1988), Section 7.

²⁵ Remember the old polemics of Russell (1912/13).

²⁶ Cf. my (1993).

²⁷ Where 'cause' is now short for 'potential causal dependence'.

²⁸ Apart from the fact already mentioned that one should not apply the term 'cause' to variables.

²⁹ If I do not want to discuss special probability measures and their independence relations, why am I considering strictly positive probability measures? Because they are not very special and because they have a distinguished role with respect to both purposes motivating the search for the properties of conditional independence. If we are interested in causal independence, then, as already mentioned, only strictly positive measures are suitable for a probabilistic theory of causality; and if we consider the wider epistemological purposes, strictly positive measures are particular suitable as well, as is clear at least since Carnap (1950) postulated strict positivity in his regularity axiom.

³⁰ In his auxiliary Lemma 1 (p. 380) Studený (1992) constructs suitable models with the help of measures which are not strictly positive. Therefore, his proof of Theorem 10 does not show that it holds also for conditional independence w.r.t. strictly positive measures. For the same reason it does not seem to carry over to NCFs.

³¹ Studený (1993), Section 5, mentions two further properties:

If $J \perp K/L$, $J \perp L/K$, $J \perp K/M$, and $L \perp M$, then $J \perp K \cup L$; and
if $J \perp L/K$, $K \perp L/J$, $J \perp K/M$, and $L \perp M$, then $J \cup K \perp L$.

These are clearly weakenings of (G6), but in contrast to (G6) they hold for *all* probability measures.

REFERENCES

- Carnap, R.: 1950, *Logical Foundations of Probability*, Chicago University Press, Chicago.
- Dawid, A. P.: 1979, 'Conditional Independence in Statistical Theory', *Journal of the Royal Statistical Society B*, **41**, 1–31.
- Dawid, A. P.: 1980, 'Conditional Independence for Statistical Operations', *Annals of Statistics*, **8**, 598–617.
- Geiger, D.: 1990, *Graphoids: A Qualitative Framework for Probabilistic Inference*, Ph.D. Thesis, University of California, Los Angeles.

- Geiger, D., Paz, A., and Pearl, J.: 1988, 'Axioms and Algorithms for Inferences Involving Probabilistic Independence', Technical Report R-119, Cognitive Systems Laboratory, University of California, Los Angeles. Also in: *Information and Computation*, **91** (1991) 128-141.
- Geiger, D. and Pearl, J.: 1988, 'Logical and Algorithmic Properties of Conditional Independence and Qualitative Independence', Technical Report R-97-IIL, Cognitive Systems Laboratory, University of California, Los Angeles, forthcoming in *Annals of Statistics*.
- Geiger, D. and Pearl, J.: 1990, 'On the Logic of Causal Models', in: Shachter *et al.* (1990), pp. 3-14.
- Geiger, D., Verma, T., and Pearl, J.: 1990, 'Identifying Independence in Bayesian Networks', *Networks*, **20**, 507-534.
- Glymour, C., Spirtes, P., and Scheines, R.: 1991, 'Causal Inference', in: W. Spohn (Ed.), *Erkenntnis Orientated: A Centennial Volume for Rudolf Carnap and Hans Reichenbach*, Kluwer, Dordrecht, pp. 151-189.
- Goldszmidt, M. and Pearl, J.: 1992, 'Rank-Based Systems: A Simple Approach to Belief Revision, Belief Update, and Reasoning About Evidence and Actions', in: *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, Ma.
- Hume, D.: 1739, *A Treatise Concerning Human Nature*.
- Hunter, D.: 1991, 'Graphoids, Semi-Graphoids, and Ordinal Conditional Functions', *International Journal of Approximate Reasoning*, **5**, 489-504.
- Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H.-G.: 1990, 'Independence Properties of Directed Markov Fields', *Networks*, **20**, 491-505.
- Matúš, F.: 1988, *Independence and Radon Projection on Compact Groups* (in Slovak), Ph.D. Thesis, Prague.
- Mill, J. S.: 1843, *System of Logic*.
- Pearl, J.: 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufman, San Mateo, Ca.
- Pearl, J. and Paz, A.: 1985, 'Graphoids: A Graph-Based Logic for Reasoning About Relevance Relations', Technical Report R-53-L, Cognitive Systems Laboratory, University of California, Los Angeles: also in: B. Du Boulay, D. Hogg, and L. Steels (Eds.), *Advances in Artificial Intelligence II*, Elsevier, Amsterdam 1987, pp. 357-363.
- Pearl, J. and Verma, T.: 1991, 'A Theory of Inferred Causation', in: J. A. Allen, R. Fikes, and E. Sandewall (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, Morgan Kaufman, San Mateo, Ca., pp. 441-452.
- Russell, B.: 1912/13, 'On the Notion of Cause', *Proceedings of the Aristotelian Society*, **13**, 1-26. Also in: B. Russell, *Mysticism and Logic*, George Allen & Unwin, London 1963, pp. 132-151.
- Shachter, R. D., Levitt, T. S., Lemmer, J., and Kanal, L. N. (Eds.): 1990, *Uncertainty in Artificial Intelligence 4*, Elsevier, Amsterdam.
- Shenoy, P. P.: 1991, 'On Spohn's Rule for Revision of Beliefs', *International Journal of Approximate Reasoning*, **5**, 149-181.
- Spirtes, P., Glymour, C., and Scheines, R.: 1993, *Causation, Prediction, and Search*, Springer, New York.

- Spohn, W.: 1980, 'Stochastic Independence, Causal Independence, and Shieldability', *Journal of Philosophical Logic*, **9**, 73–99.
- Spohn, W.: 1983, 'Deterministic and Probabilistic Reasons and Causes', in: C. G. Hempel, H. Putnam, and W. K. Essler (Eds.), *Methodology, Epistemology, and Philosophy of Science. Essays in Honour of Wolfgang Stegmüller on the Occasion of his 60th Birthday*, Reidel, Dordrecht, pp. 371–396.
- Spohn, W.: 1988, 'Ordinal Conditional Functions. A Dynamic Theory of Epistemic States', in: W. L. Harper and B. Skyrms (Eds.), *Causation in Decision, Belief Change, and Statistics*, Vol. II, Kluwer, Dordrecht, pp. 105–134.
- Spohn, W.: 1990a, 'A General Non-Probabilistic Theory of Inductive Reasoning', in: Shachter *et al.* (1990), pp. 149–158.
- Spohn, W.: 1990b, 'Direct and Indirect Causes', *Topoi*, **9**, 125–145.
- Spohn, W.: 1991, 'A Reason for Explanation: Explanations Provide Stable Reasons', in: W. Spohn, B. C. van Fraassen, and B. Skyrms (Eds.), *Existence and Explanation. Essays Presented in Honor of Karel Lambert*, Kluwer, Dordrecht, pp. 165–196.
- Spohn, W.: 1993, 'Causal Laws are Objectifications of Inductive Schemes', in: J. Dubucs (Ed.), *Philosophy of Probability*, Kluwer, Dordrecht, pp. 223–255.
- Spohn, W.: 1994, 'On Reichenbach's Principle of the Common Cause', in: W. C. Salmon and G. Wolters (Eds.), *Logic, Language, and the Structure of Scientific Theories*, University of Pittsburgh Press, Pittsburgh, pp. 215–239.
- Studený, M.: 1989a, 'Multiinformation and the Problem of Characterization of Conditional Independence Relations', *Problems of Control and Information Theory*, **18**, 3–16.
- Studený, M.: 1989b, 'Attempts at Axiomatic Description of Conditional Independence', *Kybernetika* **25**, Suppl., 72–80.
- Studený, M.: 1992, 'Conditional Independence Relations Have No Finite Complete Characterization', in: *Transactions of the 11th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Held at Prague 1990*, Vol. B, Academia, Prague, pp. 377–396.
- Studený, M.: 1993, 'Structural Semigraphoids', *International Journal of General Systems*, forthcoming.
- Suppes, P.: 1970, *A Probabilistic Theory of Causality*, North-Holland, Amsterdam.
- Verma, T. and Pearl, J.: 1990, 'Causal Networks: Semantics and Expressiveness', in: Shachter *et al.* (1990), pp. 69–76.
- Verma, T. and Pearl, J.: 1991, 'Equivalence and Synthesis of Causal Models', in: P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 6*, Elsevier, Amsterdam, pp. 255–268.

COMMENTS BY PATRICK SUPPES

Wolfgang Spohn has given a thorough discussion from many angles of the current analyses of conditional independence, as well as independence in the usual probabilistic sense. I shall not comment on the

additional analysis he gives of what he calls natural conditional functions, for I am not yet persuaded that they are really needed for epistemic purposes.

The positive results summarized at the end of the paper in Theorem 9, and the remarks following this theorem, give what is in many respects as satisfactory a characterization as can be expected of independence or conditional independence for finite sets of random variables, although some complications are mentioned below. What is particularly nice is that Theorem 11 eliminates the need for condition (S3) which constitutes an axiom schema for increasingly large finite sets of variables.

I just have two comments to make.

Finite Axiomatizability of Independence for Events. Following in the tradition of my earlier referee's report, I am still skeptical that we are able to give a finite axiomatization in first-order logic of independence for events. I would gather from the statement following Theorem 1 that Spohn also is still of an unsettled mind about this. It was not really the aim of his paper to discuss the conditions for independence of events and so I shall not say more about it here. I was, however, a little surprised that he did not mention the thorough discussion in Terrence Fine's *Theories of Probability* (1973).

Various Senses of Independence. When we turn from events to random variables, as Spohn does, we enter a much larger arena for the discussion of independence. As a simple example of this literature, I mention the various concepts of dependence introduced in Lehmann (1966), which were discussed in some detail in my 1970 monograph on causality. I recall here that three notions of dependence, and therefore associated notions of independence, in order of increasing strength are quadrant dependence, regress dependence and positive likelihood ratio dependence. There is a large statistical literature on these matters published in the last 20 years. What is interesting to note is that the material Spohn draws on from the literature on the related concepts of uncertainty, probabilistic causality, and the like in artificial intelligence is rather separate from the statistical literature on various measures of strength of dependence. Both kinds of research are desirable. They both show how rich is the notion of dependence, either in a probabilistic sense or

in a causal sense, and in fact really how pluralistic the subject can be made by introducing plausible concepts of weaker or greater strength. Generalizations to cover these other cases would be of interest.

REFERENCES

- Fine, Terrence: 1973, *Theories of Probability*, New York, Academic Press.
Lehmann, E. L.: 1966, 'Some Concepts of Dependence', *Annals of Mathematical Statistics*, **37**, 1137–1153.
Suppes, P.: 1970, *A Probabilistic Theory of Causality*, Amsterdam, North Holland.