First publ. in: Causalità e modelli probabilistici / A cura di Maria Carla Galavotti e Guido Gambetta. - Bologna: CLUEB, 1983. - S. 69-87

Causalità e modelli probabilistici,
a cura di M.C. Galavotti e G. Gambetta,
CLUEB, Bologna © 1983.

Wolfgang Spohn

Probabilistic causality: from Hume via Suppes to Granger

1. *Introduction*

There are subjects which have a sufficiently concrete bearing to be of interest to particular sciences and which are sufficiently general to be of interest to philosophy. Such a subject is causality. Unfortunately, the philosophers' and the scientists' contributions to such subjects tend to be incongruous. This is due to their somewhat differing methods of approaching the truth or, to put it less pretentiously, to their somewhat differing acceptance behaviour. Scientists are the more inclined to accept an idea, the more concretely it can be put to work and the more successful its concrete applications are, whereas philosophers are the more inclined to accept an idea, the better it stands their procedure of exposing it to more and more sophisticated arguments and counterarguments. These are differing demands, and an idea may well fulfill one, but not the other. Of course, this is an idealization of the contrast, but much of it is true of the subject of causality. There, much work is done in a departmentalized way without good knowledge of what is going on in the other department; and if philosophers and scientists are looking over their fence, then philosophers are likely to see scientists trying to make ideas applicable which they think to have been proven to be too coarse or even untenable, and scientists are likely to wonder about the philosophers' strange process of reasoning which produces complicated explications of causality of which they can hardly make any concrete use.

This state of affairs is undesirable and, I think, also unnecessary. I am confident that there is a theory of causality that meets both, the philosophers' and the scientists', demands. However, this is not the place to present and defend such a theory. Instead, I shall present a line of thought which starts with Hume's most rudimentary explication of the concept of causation and ends up with Granger's definition of causality designed for econometrical and statistical applications[1], the most important link being the probabilistic theory of causality by Suppes (1970). If this line is thought to be stringent, it shows that there is a firm continuous bridging between the work of (some) philosophers and that of (some) scientists. It is also the line along which I think a theory of causality satisfying philosophers and scientists may be developed.

2. *The philosophically most preferred explication of causality and its probabilistic analogue*

Let us start with a well known quotation from Hume: "We may define a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second.* Or in other words *where, if the first object had not been, the second never had existed.*"[2] It has often been noticed[3] that this is not one definition in two different wordings; rather, it presents two different explications of causation. For a long time, attention has been focused on the first explanation, which has become elaborated to the so-called regularity theory of causation. But now, as more and more inadequacies of the regularity theory have become apparent under closer scrutiny[4], interest seems to have almost completely turned to the second explanation, which may

[1] Cf. e.g. Granger (1980a). Among all econometric approaches to causality known to me Granger's is the one which fits best into the presented line of thought. I hope that this is a good end point also from the econometrician's point of view. This hope is supported by the fact that Granger's approach seems, for all I know, to be the most widely discussed and thus the most promising approach within econometrics.

[2] Hume (1975), Sect. VII, Part II.

[3] E.g. by Lewis (1973b), p. 556f.

[4] Cf. e.g. Mackie (1974), ch. 3, in particular pp. 81-87, or the hints given by Lewis (1973b), p. 556f.

be called the counterfactual analysis of causation.[5] Let us state it in somewhat more modern terms:

(1) An event A is a cause of an event B if and only if:
 (a) both, A and B, happened, and A earlier than B,
 (b) if A had not happened, B had not happened, too.

Two provisos are to be added to (1). First, (1) is not to mean, of course, that only past events can cause one another; the only reason why (1) refers only to past events is that somewhat strange and clumsy formulations are required for speaking counterfactually about the future. Secondly, (1) is not to dogmatically exclude the possibilities of simultaneous or even backwards causation. Rather, the intent is only to leave aside these intricate possibilities and to restrict our discussion to the normal case, where the cause in fact precedes the effect; this normal case is difficult enough to analyze.

Apart from these points, it is hard to deny the correctness of (1). However, this only means that the counterfactual (1b) is as much in need of explication as is the concept of causation. I cannot remove the main unclarity here, but some improvements on (1) are easily made:

(2) An event A is a cause of an event B if and only if:
 (a) both, A and B, happen, and A earlier than B,
 (b) under the obtaining circumstances, A is a necessary and/or sufficient condition for B.

(2) is presumably the most widely accepted preliminary analysis of causation.[6]. It improves (1) in two respects:

First, the "and/or" in (2b) points to a problem hidden in (1). According to (1b), A is only a necessary condition of B; without A B could not have happened. (1b) does not say that, when A happened, B was bound to happen, i.e. that A is a sufficient condition

[5] Usually, only the regularity theory is associated with Hume. Thus it is worth noting that it seems that Hume was well aware of this ambiguity, that he in fact entertained both analyses of causality and held them to be compatible. This is argued by Beauchamp, Rosenberg (1981), ch. 1.

[6] Cf. the impressive list of philosophers entertaining (2) given by Sosa (1975), p. 1.

of B. Because of this, (1) is questionable, and there is in fact an extensive discussion about how the vagueness of the "and/or" should be resolved, i.e. whether causes are necessary conditions or sufficient conditions or both.[7]

Secondly, and more importantly, (2) makes explicit the dependence of the cause-effect-relation on the obtaining circumstances. This dependence is present also in (1), but only implicitly. This will become clearer by considering a simple (and not original) example: I have just stroken a match, and it has ignited. Certainly, we would say that the first is a cause of the second, and also that, if I had not stroken the match, it would not have ignited. But we do not thereby imply that, whenever a match is (not) stroken, it does (not) ignite, ot that the striking of that match is in itself a necessary or sufficient condition of its igniting. Rather, our confidence in the causal and in the counterfactual assertion about that match rests on our beliefs about the given situation, e.g. that the match was a normal match with a normal sulphur head, that a sufficient amount of oxygen was present, etc. If these beliefs turn out to be false, we may withdraw these assertions; we would do so, for instance, if we would learn that the head of the match contained a tiny time bomb. The lesson is a general one: The truth of a counterfactual or a causal statement about a certain situation does not only depend upon the antecedens and the consequens of the counterfactual or upon what is described as cause and effect; it also depends upon the surrounding facts. And this dependence is made explicit in (2b).[8]

Exactly what is to belong to the obtaining circumstances, is, however, not an easy question. It seems to do no harm, if everything that happened before the cause is counted as belonging to the obtaining circumstances, although these would then contain a lot of irrelevant things. Certainly, some things being the case in the time between cause and effect should be included in the obtaining circumstances, e.g. standing conditions like the presence of oxygen in our match example. Equally certainly, some things being the case in the time between cause and effect must not be included in the obtaining circumstances. For instance, the fact that the match reached its ignition temperature shortly after being stroken and shortly before igniting must not belong to the obtaining circumstances of our case, because, *given* this fact, the striking of the match would – contrary to our intuitions – not be a necessary condition of its ignition. This raises the suspicion that "the obtaining circumstances" is in fact a causal notion that the circumstances pertinent to a certain cause A and its effect B consist only of things which are *causally* independent of A. If this is true, the analysis (2) is threatened by circularity. This is a serious, but, as I think, solvable difficulty. However, let us pass over this problem now; for our line of thought it is more important to be aware of the dependence of a causal relation on the obtaining circumstances than to have solved the problems hidden in this dependence. (But we shall return to this difficulty in section 4, where a hint at a possible solution is given.)

[7] Cf. again Sosa (1975), p. 1, or Mackie (1974), ch. 2, a good deal of which is about this point.

[8] For this point see Mackie (1974), ch. 2, and also Lewis (1973a), sect. 3.2. Lewis' book is certainly still the most influential philosophical treatment of counterfactuals. However, the conception of counterfactuals which I prefer most is the one presented in Ellis (1979).

The main difficulty with (2) is, of course, to say exactly what is meant by a necessary or a sufficient condition. Here opinions are divided. It is clear that "necessary" and "sufficient" do here not mean "logically necessary" and "logically sufficient". Many think a regularity account of conditionship to be appropriate here, i.e. that "under the obtaining circumstances, A is a sufficient condition for B" means "there are true laws $L_1, \ldots, L_n$ such that B is logically implied by A, the obtaining circumstances, and $L_1, \ldots, L_n$". This is certainly not too bad an explication, but if the arguments against the regularity theory of causation referred to above are compelling, it cannot be completely adequate.[9] However, we need not worry now about this grave unclarity; the sequel will be independent of it.

There is another point to be emphasized which is probably tacitly understood in (2), but not really explicit. Each event or state of affairs has one of three modal status, it is either necessary or impossible or contingent, i.e. neither necessary nor impossible. Obviously, this is a logical truth, whatever the precise meaning of

[9] Cf. also Sosa (1975), p. 2f.

"necessary" might be. Now it is important to note that "under the obtaining circumstances, A is a sufficient condition for B" does not only mean "given the obtaining circumstances and A, B is necessary", but also "given the obtaining circumstances and *non-A*, B is *not* necessary, i.e. contingent or impossible". Likewise, "under the obtaining circumstances, A is a necessary condition for B" does not only mean "given the obtaining circumstances and *non-A*, B is impossible", but also "given the obtaining circumstances and A, B is *not* impossibile, i.e. contingent or necessary". This is reflected in our next reformulation:

(3) An event A is a cause of an event B if and only if:
  (a) both, A and B, happen, and A earlier than B,
  (b) under the obtaining circumustances, A raises the modal status [10] of B. [11]

So far, our discussion went on wholly within a deterministic framework which traditionally is the natural framework for discussing causal matters. Now, within this collection of papers it is not necessary to explain and defend that there is every reason to deal with causality also within a probabilistic setting. However, everybody will agree that there should be as intimate a connection as possible between the treatments of causality within the two frameworks and that treatments that cannot be fitted together in some plausible way are unlikely to be acceptable. This is why I have driven our discussion to the formulation (3). For, there is a very simple and very close probabilistic analogue to (3). We only have to find the probabilistic analogues to the deterministic modal status; and, of course, these are nothing else but the probabilities themselves. Thus, we arrive at the following basic idea for a probabilistic explication of causality, from which we can enter more concretely into our subject:

[10] Where, of course, "necessary" is to be the highest and "impossible" the lowest modal status.

[11] The problem about the "and/or" in (2b) is now transformed into the question which of the various possible raisings of modal status is characteristic of causation.

(4) An event A is a cause of an event B if and only if:
  (a) both, A and B, happen, and A earlier than B,
  (b) under the obtaining circumstances, A raises the probability of B, i.e. $P(B \mid A) > P(B)$. [12]

Of course, this is still not a very precise formulation; the exact force of the clause "under the obtaining circumstances" is as unclear as it was in (2) or (3). Perhaps you also wonder where the unclarity of all the modal expressions in (2) and (3) such as "sufficient condition", "necessary", etc. has its counterpart in (4). It has one, I think. As you will admit, the interpretation of probability is still a very controversial matter, and I think one can observe close parallels between the discussion of the meaning of the modalities in (2) and (3) and that of the interpretation of probability. However, this is much too large a topic; here we have to be content with everybody's applying his own understanding of probability. [13]

3. *Suppes' probabilistic theory of causality*

Let us proceed with the probabilistic theory of causality proposed by Suppes (1970) which is still the major philosophical contribution to this subject. Despite its predecessors Reichenbach (1956) and Good (1961/62), it probably did most towards making reputable the probabilistic framework among philosophers. [14] This section is largely descriptive, though it contains some minor changes. The next section brings forward what I think to be a substantial amendment. As may have been expected, the general tendency is

[12] Within the related context of probabilistic explanation, (something like) (4) has become known as the positive relevance criterion, as Niiniluoto (1972), p. 31, termed it. This criterion seems to be accepted as the most plausible approach to probabilistic explanation.

[13] In the sequel, we shall deal with causation only within a probabilistic framework. In Spohn (1983), I try to show that the analogy between deterministic and probabilistic causality, which I suggested in the last paragraphs, is in fact a general and systematic one. In particular, it is argued there that our considerations in the next two sections also apply to deterministic causality.

[14] Cf. also the excellent discussion given by Salmon (1980).

to drive Suppes' theory as near to (4) as possible.

First, we have to introduce a bit of formal apparatus. We assume a probability space $<\Omega, \mathcal{A}, P>$, where Ω is a sample space, $\mathcal{A}$ is a σ-algebra of events over Ω, and P is a σ-additive probability measure defined on $\mathcal{A}$. Throughout the following we keep this probability space fixed, and all our definitions and considerations are relative to it; this is important. Ω and $\mathcal{A}$ are to have a certain structure, i.e. they shall represent a stochastic process or a time series. However, we shall not need an exact specification of that structure. The only important thing for us is that, among all the possibly very complex events in $\mathcal{A}$, there are events which refer to a definite point of time. To give an example: If Ω and $\mathcal{A}$ describe an infinite series of throws of a certain die, then $\mathcal{A}$ contains events not referring to a definite time such as "in the first ten throws no six turns up", and events referring to a definite time such as "in the fifth throw an even number turns up". Events referring to a definite point of time are denoted by A_t, $B_{t'}$, etc., where t and t', respectively, are the points of time referred to by A_t and $B_{t'}$. Only such events are potential causes and effects, and hence only they will be of interest to us. Of course, we also assume that the points of time $t, t', \ldots$ are ordered by the order relation "earlier than", symbolized by "$<$".

Le me add an important terminological note: There is a big and rather sophisticated discussion going on as to what sort of things causes and effects exactly are, whether they are events, facts, states of affairs, propositions, or still something else.[15] Again, we cannot go more deeply into this question. For us it is only important to observe that there is a double use of the term "event". According to the philosophical as well as the ordinary use, events are always actual events taking place at some point of time or in some time interval. In mathematical probability theory, events are something else; there, an event is only a possible event, it may or may not occur, and it may be scattered over time (though our special events A_t are not). Thus, events in the mathematical sense are rather something like the philosophers' states of affairs or propositions.[16] Here, I use "event" only in the mathematical sense, and I have already done so in section 2.

So, we are well prepared for Suppes' first definition[17]:

(5) A_t is a *prima facie cause* of $B_{t'}$, iff
(a) A_t and $B_{t'}$ occur, and $t < t'$,
(b) $P(B_{t'} \mid A_t) > P(B_{t'})$.

(5) deviates from Suppes' definition 1 in two small respects: Suppes includes "$P(A_t) > 0$" as a further condition, which we assume to be presupposed by (5b). This remark applies as well to our later definitions. On the other hand, Suppes does not require that A_t and $B_{t'}$ occur. On p. 40 he explains his procedure. In all his definitions he explicates, as one may say, potential causal relations which become actual, if the events standing in these relations are actual. I feel it is a bit less confusing to discuss only Suppes' actual causal relations, and that is why I have required in (5a) A_t and $B_{t'}$ to occur. But this is a subordinate point.

Suppes' justification for (5) essentially consists in pointing out that there are a lot of examples to which it seems to fit[18] and that it therefore is a prima facie plausible starting point for the following considerations. These consist in taking into account a number of ways in which (5) goes wrong. As you will have already noted, (5) is rather close to (4). The only difference is that the obtaining circumstances mentioned in (4b) do not appear in (5b), and, as we shall see, it is because of them why (5) may go wrong.

The first way for (5) to go wrong is established by *the argument for spurious causes*: There may have occurred events prior to A_t in the light of which A_t is irrelevant to $B_{t'}$ and, thus, does not raise the probability of $B_{t'}$. In this case, A_t should not be considered to be a cause of $B_{t'}$; and if A_t is a prima facie cause of $B_{t'}$, then it is only spuriously so. The classical example for this is provided by barometers. The fact that my barometer is falling quickly makes it very likely that there will soon be a thunderstorm over Munich. Thus, the barometer's falling is a prima facie cause of the thunderstorm, but certainly not a cause. Now the fact is that there is very

[15] Cf. e.g. Mackie (1974), ch. 10, or Beauchamp, Rosenberg (1981), ch. 7.

[16] I consider the discussion between Martin (1981) and Suppes (1981) to be a nice example of the sort of misunderstanding that may arise if this double use is not clearly recognized.

[17] Cf. Suppes (1970), p. 12, Definition 1.

[18] Cf. Suppes (1970), pp. 12-20.

low pressure rapidly approaching Munich, and given this, the thunderstorm is stochastically independent of the behaviour of my barometer, which therefore is only a spurious cause of the thunderstorm, as desired. Other examples come easily. In fact, the phenomenon of spurious causation is very similar to the phenomenon of spurious correlation which is very familiar in statistics.[19] Thus, we may define:

(6) A_t is *(factually) a spurious cause* of $B_{t'}$ iff
(a) A_t is a prima facie cause of $B_{t'}$,
(b) there are a $t^* < t$ and an event C_{t^*} which has occurred, such that $P(B_{t'} \mid A_t \cap C_{t^*}) = P(B_{t'} \mid C_{t^*})$.

Let me immediately add another definition:

(7) A_t is *necessarily a spurious cause* of $B_{t'}$ iff
(a) A_t is a prima facie cause of $B_{t'}$,
(b) there are a $t^* < t$ and events $C^1_{t^*}, \ldots, C^n_{t^*}$ which are exhaustive and pairwise disjoint, such that for all $i = 1, \ldots, n$ $P(B_{t'} \mid A_t \cap C^i_{t^*}) = P(B_{t'} \mid C^i_{t^*})$.

(6) is almost identical with Suppes' definition of "spurious cause in sense one" in (1970), p. 23, and (7) is completely identical with his definition of "spurious cause in sense two" in (1970), p. 25. Suppes states both definitions because he remains undecided as to which definition to prefer. This indecision may be resolved, I think, by looking at the differences between (6) and his definition of spuriousness in sense one. Suppes' definition contains the further condition that $P(B_{t'} \mid A_t \cap C_{t^*}) \geqslant P(B_{t'} \mid A_t)$ – with so little justification that I have omitted it here. Instead, I have added the requirement that C_{t^*} be an actual event. This is essential. For, almost any event which we would consider to be a cause of something else might be rendered spurious by some possible, unactualized course of the past. But we would not want to conclude from this that that

[19] However, they are not the same, firstly because uncorrelatedness is a weaker notion than statistical independence, and secondly because it is essential for spurious causation that the event responsible for the spuriousness is earlier than the prima facie cause.

event *is* a spurious cause. Rather, we would draw this conclusion only if the circumstances responsible for the spuriousness are *present*.[20] This also makes clear the relation between (7) and (6). If A_t is necessarily a spurious cause of $B_{t'}$, then there must be an event rendering A_t spurious in the sense of (6), simply because one of the possible events $C^1_{t^*}, \ldots, C^n_{t^*}$ must be actual.

With Suppes we may go on to define that A_t is a *genuine cause* of $B_{t'}$ iff A_t is a prima facie, but not a spurious cause of $B_{t'}$.[21] And it is implicit in Suppes that A_t is *no cause* of $B_{t'}$ iff A_t is not a genuine cause of $B_{t'}$.

There is a second way for (5) to be inappropriate. This is shown by *the argument for indirect causes*: If A_t is a prima facie cause of $B_{t'}$, then the suggested causal relation might be mediated by further events, in whose presence A_t again is irrelevant to $B_{t'}$ and does not raise the probability of $B_{t'}$. However, we should not say in this case that A_t is a spurious cause, because A_t has causal influence on these mediating events. We should therefore better say that A_t is shown by these mediating events to be an indirect cause of $B_{t'}$.

As an example consider this sequence of events: the Federal Bank has raised the bank-rate, then the amount of money in circulation decreases, and finally the rate of inflation goes down. It is plausible that the first event raises the probability of the third, but it is also plausible that, given the second event, the first event does not raise the probability of the third. However, this constitutes no reason for denying an influence of the bank-rate on the rate of inflation, it only means that this influence is exerted via the amount of money in circulation.

Suppes tries to capture this idea in the following definition:

(8) A_t is an *indirect cause* of $B_{t'}$ iff
(a) A_t is a prima facie cause of $B_{t'}$,
(b) there are a t^* with $t < t^* < t'$ and events $C^1_{t^*}, \ldots, C^n_{t^*}$ which are exhaustive and pairwise disjoint, such that for all $i = 1, \ldots, n$ $P(B_{t'} \mid A_t \cap C^i_{t^*}) = P(B_{t'} \mid C^i_{t^*})$.

[20] One can only speculate why Suppes seems to have overseen this point. Perhaps he did so because he dealt only with potential causal relations. Or perhaps he slipped from the mathematical to the philosophical use of "event".

[21] Cf. Suppes (1970), p. 24.

Of course, A_t is to be a *direct cause* of $B_{t'}$ iff A_t is a prima facie, but not an indirect cause of $B_{t'}$.[22] The only difference between (7) and (8) is that in (7) t^* is earlier than t, whereas in (8) t^* is between t and t'. In fact, this was the only difference in our arguments and examples leading to (7) and (8). However, Suppes gives no reason at all why he parallels (8) with (7) and not with (6) or with his spuriousness in sense one. But let us not dwell upon this point now, let us take it only as a hint that there still are a number of details which need be worked through thoroughly.

4. *An amendment to Suppes' theory*[23]

What we did in moving from (5) to (6), (7), and (8) was to take into account the circumstances of A_t and $B_{t'}$. However, we have done this only partially so far. The assumption on which Suppes' strategy rests is that only prima facie causes can be causes and that one has therefore only to classify the different sorts of prima facie causes. This assumption is wrong. There is a third way for (5) to be wrong which is pointed out by *the argument for hidden causes*:

Suppose that A_t and $B_{t'}$ both occurred, where $t < t'$, and that A_t is *not* a prima facie cause of $B_{t'}$. This may mean that there is no influence of A_t on $B_{t'}$; but it may also mean that this influence is still hidden and becomes apparent only when further events are considered. That is, there may be events in whose presence A_t does raise the probability of $B_{t'}$. Or to put it formally: even if $P(B_{t'} \mid A_t) = P(B_{t'})$, there may be a time $t^* < t'$ and an event C_{t^*} which has occurred such that $P(B_{t'} \mid A_t \cap C_{t^*}) > P(B_{t'} \mid C_{t^*})$. On second thoughts, A_t then seems to be a cause of $B_{t'}$.

Let me exemplify this by a caricature of an old quarrel about psychoanalysis: Suppose that statistical data show that the spontaneous remission time of people suffering from a certain neurosis has about the same distribution as the remission time of people

[22] Cf. Suppes (1970), p. 28f.

[23] I have presented the following reasoning already in Spohn (1980), though in a formally somewhat pretentious way. Here, I concentrate more on the intuitive essentials.

undergoing psychoanalytic treatment. Thus, psychoanalysis seems to be inefficacious with respect to this neurosis. This is not the full truth, however. If we analyze the data with respect to the income of the psychoanalytically treated patients, we observe that the higher the income of the patients, the more the distribution of the remission time shifts upwards. That is, roughly, for rich people psychoanalysis has a protracting influence on remission time, and for poor people it has an accelerating one.[24] This is certainly a familiar phenomenon in statistics. Just as partial-correlation analysis may reveal spurious correlation, it may also reveal the opposite.[25]

The argument for hidden causes is, so to speak, the counterpart of the arguments for spurious and indirect causes. Note that for the argument for hidden causes it is only essential that the time t^* of the event uncovering the causal relation is earlier than the time t' of the effect. The first two arguments were more specific in this respect.

The consequence of the argument for hidden causes does not consist in adding a further definition of "hidden cause". The essential point is that our three arguments develop a most interesting *interplay*:

Les us assume that A_t is a prima facie cause of $B_{t'}$ (as will be immediately clear, we might just as well start with the assumption that A_t is not a prima facie cause of $B_{t'}$). Now there may have happened an event $C^1_{t_1}$ $(t_1 < t)$ satisfying (6b). Then, secunda facie, A_t is a spurious, i.e. no cause of $B_{t'}$. Still, there may have happened a further event $C^2_{t_2}$ $(t_2 < t)$ such that $P(B_{t'} \mid A_t \cap C^1_{t_1} \cap C^2_{t_2}) > P(B_{t'} \mid C^1_{t_1} \cap C^2_{t_2})$. Then, tertia facie, A_t is a cause of $B_{t'}$ according to the argument for hidden causes. A third event $C^3_{t_3}$ $(t_3 < t)$ might again bring to bear the argument for spurious causes. And so on. When does this to and fro between these two arguments end? In reality probably rather soon; but from a logical point of view only when there are no events left which could support the argument for spurious causes. That is: Let Z_1 be that event of the σ-algebra $\mathcal{A}$ which describes the actual courses of events up to t, and suppose

[24] Here we have assumed, of course, that income and spontaneous remission time are stochastically independent according to the data.

[25] However, footnote 18 applies here as well.

that $P(B_{t'} \mid A_t \cap Z_1) > P(B_{t'} \mid Z_1)$ holds true. Then A_t cannot be a spurious cause of $B_{t'}$, simply because the argument for spurious causes cannot be applied any more. On the other hand, if $P(B_{t'} \mid A_t \cap Z_1) \leqslant P(B_{t'} \mid Z_1)$ the argument for hidden causes can still be applied. In fact, a second to and fro starts. If A_t seems to be a cause of $B_{t'}$ on the basis of Z_1, then A_t may be shown to be only an indirect cause by an actual event $D^1_{s_1}$, where $t < s_1 < t'$. Still, a further actual event $D^2_{s_2}$ $(t < s_2 < t')$ might suggest that A_t is hiddenly a direct cause of $B_{t'}$. And so on again. And again, it is clear that this second to and fro is guaranteed to end only when the actual course Z_2 of events between t and t' is also completely taken into account. Thus, we have finally arrived at our improved definition of direct causes:

(9) Suppose that $Z = Z_1 \cap Z_2$, i.e. that Z is that event of $\mathcal{A}$ which describes the actual courses of events up to t' with the exception of t (and t'). Then A_t is a *direct cause* of $B_{t'}$ iff
(a) A_t and $B_{t'}$ occur, and $t < t'$,
(b) $P(B_{t'} \mid A_t \cap Z) > P(B_{t'} \mid Z)$.

Some remarks are in order: First, I am content here with having stated this explication of the notion of a direct cause. Of course, it can only be the starting point of a probabilistic theory of causality improving upon Suppes' theory. The next step would be to explicate the notion of an indirect cause and then to define a cause as a direct or indirect cause.[26] In principle, this second explication may be obtained in the same way as (9). But it is quite a bit more complicated than (9)[27], and we do not need it for our present purpose, since already (9) lends itself to a smooth transition to

[26] And then, many further steps would be necessary for constructing a theory of which the sciences can really make use.

[27] In Spohn (1980), I have given an explication both of direct and indirect causal relevance. But causal relevance may be positive or negative, whereas to be a cause is to have positive causal relevance. This makes causation more difficult to explicate than causal relevance, and it is one reason why the explication of "indirect cause" cannot be simply paralleled to that explication of indirect causal relevance.

Granger's work, as we shall see in the last section.

Secondly, I find that (9) matches (4) as perfectly as such an imprecise thing as (4) can be matched. The Z of (9), i.e. the whole past of the effect $B_{t'}$ except the cause A_t, plays now the role of the obtaining circumstances in (4). However, have we not mentioned in section 2 two reasons why this whole past contains more than should be admitted as the obtaining circumstances of A_t and $B_{t'}$? Yes, but this does not affect (9). Let me explain why.

The first reason was that this whole past usually contains a lot of events irrelevant to A_t and $B_{t'}$, which should not count as circumstances. This is certainly true. But what does "irrelevant" mean here? It can only mean "causally irrelevant", and, thus, we seem to be caught in another circle.[28] But there is a way out, I believe. As these irrelevant events seem to do no harm, one may at first include them in the Z of (9). Then one should proceed from (9) to an explication of causal irrelevance. And finally, one should *prove* that the definition (9) is equivalent to a definition obtained from (9) by replacing Z by some Z' which is a conjunction of all the events causally relevant to A_t and $B_{t'}$ in the explicated sense.

The second reason was that some events occurring between A_t and $B_{t'}$ must be excluded from the obtaining circumstances, lest the explication (2) becomes inadequate. But why? Let us recall our match example. There, A_t was the striking of the match, $B_{t'}$ was its ignition, and one event to be excluded from the circumstances of A_t and $B_{t'}$ was the event of the match reaching its ignition temperature; let us call it D_s $(t < s < t')$. Intuitively, D_s only shows that A_t is no direct cause, though still a cause, of $B_{t'}$. According to (2), however, A_t would be no cause of $B_{t'}$ at all, if D_s would belong to the obtaining circumstances. This is so, because in (2) "cause" means "direct *or* indirect cause". This consideration shows that such events as D_s must be excluded only from the circumstances of an *indirect* cause and its indirect effect. Hence, as (9) defines only direct causes, the Z of (9) is not incorrect in including all the actual events between A_t and $B_{t'}$.

A last critical question is suggested by the foregoing: Is it really possible, as we have apparently assumed, that a direct cause is tem-

[28] This is not the same circularity as the one pointed out in section 2.

porally separated from its direct effect? Yes, it is possible in a twofold sense. First, as I have emphasized, all our considerations were relativized to the given probability space $<\Omega, \mathcal{A}, P>$. We should not presuppose that the points of time referred to by the events of $\mathcal{A}$ exhaust our physical time continuum; time may as well be discrete in this probability space.[29] Thus, it is possible that physical time elapses between the direct cause and its effect, although $\mathcal{A}$ contains no event temporally between them. But even in terms of the time of the given probability space, it is conceptually possible that a direct effect does not immediately follow its direct cause. Our reluctance to accept this possibility does, I think, not have purely conceptual grounds, but rests on our firm conviction that causal chains are continuous and do not make jumps; and this conviction is not an analytical truth (though it is not purely empirical either). Thus, as explication is first of all a conceptual matter, we should not criticize (9) for allowing this conceptual possibility.

5. *Granger's theory of causality*

Let us finally relate our explication (9) with Granger's work on causality. To be precise, I shall relate (9) only with what he calls his general definition (in (1980a), p. 330), though his main concern is not about this general definition in itself, but about making it operational so that it can be used in concrete econometric work. However, I am content with using this general definition as a bridgehead for the philosophical invasion of econometrics and with occupying this bridgehead. Its hinterland is somewhat impassable for the philosopher.

The different versions of this general definition amount to the same thing.[30] So let me quote the version given in (1980a), p. 330:

"For ease of exposition, a universe is considered in which all variables are measured just at prespecified time points at constant

[29] In fact, for (9) to be mathematically meaningful, time must be discrete. How to extend the envisaged theory of causality to continuous time processes, is a difficult problem.

[30] Cf. also Granger (1969), p. 428, and (1980b), p. 3.

intervals $t = 1, 2, \ldots$ When at time n, let all the knowledge in the universe available at that time be denoted Ω_n and denote by $\Omega_n - Y_n$ this information except the values taken by a variable Y_t up to time n, where $Y_n \in \Omega_n$. ... Ω_n will certainly be multivariate and Y_n could be, and both will be stochastic variables. ...

Suppose that we are interested in the proposition that the variable Y causes the variable X. At time n, the value X_{n+1} will be, in general, a random variable and so can be characterized by probability statements of the form Prob $(X_{n+1} \in A)$ for a set A. This suggests the following:

General Definition. Y_n is said to cause X_{n+1} if

$$\text{Prob}\,(X_{n+1} \in A \mid \Omega_n) \neq \text{Prob}\,(X_{n+1} \in A \mid \Omega_n - Y_n) \text{ for some } A.$$

For causation to occur, the variable Y_n needs to have some unique information about what value X_{n+1} will take in the immediate future."[31]

This is easily translated into our terminology: Let $X_{t'}$ be a random variable defined on our Ω and referring to the time t' [32]; this is to be Granger's X_{n+1}. Let Y_t be a random variable defined on our Ω and referring to the time $t < t'$; this is to be Granger's Y_n. Let y be the actual value of Y_t; then the event $\{Y_t = y\} = \{\omega \in \Omega \mid Y_t(\omega) = y\} \in \mathcal{A}$ corresponds also to Granger's Y_n. Moreover, Granger's $\Omega_n - Y_n$ translates into our Z of (9)[33], and Granger's Ω_n translates into our $\{Y_t = y\} \cap Z$. Thus, we may finally translate Granger's definiens by:

(10) $P(\{X_{t'} \in R\} \mid \{Y_t = y\} \cap Z) > P(\{X_{t'} \in R\} \mid Z)$ for some set R of possible values of $X_{t'}$.[34]

[31] Bluntly, I wonder how statisticians manage to get along with such a sloppy notation. Why, in particular, is it so hard to distinguish between random variables (which are functions) and their values (which usually are numbers)?

[32] That is, $X_{t'}$, is $\mathcal{A}_{t'}$-measurable, where $\mathcal{A}_{t'}$ is the sub-σ-algebra of $\mathcal{A}$ consisting of all events $B_{t'}$ referring to the time t'.

[33] Granger's more restricted J_n (cf. (1980a), p. 336), which he uses in his operational definitions, may as well be translated into our Z of (9). This is so because we can make our probability space $<\Omega, \mathcal{A}, P>$ more or less inclusive.

[34] Obviously, if the $\neq$-assertion holds for some R, so does the $>$-assertion.

If (10) holds true, Granger says that Y_t causes $X_{t'}$, and we would say that the actual value of Y_t is a direct cause (in the sense of (9)) of $X_{t'}$'s taking a value from some set R (if $X_{t'}$ actually takes some value from R). This nicely marks the small differences and the almost perfect correspondence between Granger's general definition and our explication (9).

To list the differences: (i) Granger requires $X_{t'}$ to immediately follow Y_t; I do not – for the reasons given in the last paragraph of section 4. (ii) There was an imprecision in our translation. Granger's Y_n (in the event sense) consists in "the values taken by a variable Y_t *up to time n*", whereas our $\{Y_t = y\}$ describes the value taken by Granger's Y_t *at* our time t. I do not see a reason for following Granger in this point. (iii) I think one should not speak of random variables causing one another. I prefer to say that (10) only defines that Y_t is actually causally relevant for $X_{t'}$ (also because it is not guaranteed that $X_{t'}$ actually takes some value from a set R for which (10) holds).[35] (iv) Granger speaks of causation in general, where we only speak of direct causation. Thus, Granger's definition is still insensible of the distinctions coming forth from our sections 3 and 4.

This list shows that there still are a number of details to be cleared up in order to establish a perfect matching between philosophers' and econometricians' probabilistic theories of causality. But already now the match is so good that I think I have not promised you too much in the introduction.

[35] One might go on to say that Y_t is necessarily causally relevant for $X_{t'}$ iff (10) holds for all possible values y of Y_t.