FRANK ARNTZENIUS

PHYSICS AND COMMON CAUSES*

ABSTRACT. The common cause principle states that common causes produce correlations amongst their effects, but that common effects do not produce correlations amongst their causes. I claim that this principle, as explicated in terms of probabilistic relations, is false in classical statistical mechanics. Indeterminism in the form of stationary Markov processes rather than quantum mechanics is found to be a possible saviour of the principle. In addition I argue that if causation is to be explicated in terms of probabilities, then it should be done in terms of probabilistic relations which are invariant under changes of initial distributions. Such relations can also give rise to an asymmetric cause-effect relationship which always runs forwards in time.

1. INTRODUCTION

One does not cure an infection by sitting in a cold bath so as to lower one's body temperature. It is folk wisdom that one cures a disease by removing some of the causes of the disease and not by removing some of the effects. One of the long-standing problems of philosophy is to provide an account of causation that incorporates such a difference between cause and effect. For instance, an extremely simple-minded account of causation, which I attribute to nobody, is that A causes B precisely when there is a positive correlation between A- and B-type events. Such an account fails to account for folk-wisdom, since it entails that A causes B whenever B causes A.

Recently many philosophers have argued that one can specify a precise probabilistic sense in which causes differ from effects. Their claim, known as the common cause principle, is that common causes lead to correlations among their effects, while common effects do not produce correlations among their causes. It has also been claimed by some that one can derive such an asymmetry from classical statistical mechanics, or at the very least that the asymmetry follows directly from reasoning analogous to reasoning in statistical mechanics. In this paper I shall analyse such claims, and find that, quite to the contrary, statistical mechanics implies the falsity of the common cause principle. In the final section I shall consider what can be salvaged from the wreckage, and I will find that probabilistic accounts of causality can be improved by looking at the properties of Markov processes.

Synthese **82**: 77–96, 1990.

2. THE COMMON CAUSE PRINCIPLE

Let us begin with an example of a common cause explanation, suggested by D. Papineau (Papineau 1985). Smoking causes cancer and smoking causes yellow fingers, in the sense that smokers are more likely to develop cancer than non-smokers, and are more likely to develop yellow fingers than non-smokers. One would therefore expect there to be a correlation between cancer and yellow fingers. For if one has cancer one is more likely to be a smoker than if one did not have cancer, and hence one is more likely to have yellow fingers than if one did not have cancer. Thus, a good explanation of such an observed correlation between yellow fingers and cancer is to point at the (statistical) presence of a common cause of the correlata, which in this case is smoking. Such an explanation is a common cause explanation.

Now consider the case of common effects. Working in asbestos factories raises the probability of contracting cancer, and so does smoking. Smoking and working in asbestos factories can be said to have cancer as a common effect. Nevertheless, one does not at all expect there to be a correlation between smoking and working in asbestos factories. Moreover, if there were such a correlation, one would not regard the presence of the common effect cancer as an explanation of the correlation between working in asbestos factories and smoking. This purported asymmetry is known as the common cause principle.

Let us now express the common cause principle more clearly, without presupposing that one already knows how to distinguish causes from effects, in terms of probabilistic relations between events. This approach was first suggested by H. Reichenbach (Reichenbach 1956), and further elaborated by W. Salmon and others (see, e.g., Salmon 1984).

Let us define a conjunctive fork by the following four probabilistic expressions:

(i) $\mathrm{Prob}(A/C) > \mathrm{Prob}(A/\text{not-}C)$
(ii) $\mathrm{Prob}(B/C) > \mathrm{Prob}(B/\text{not-}C)$
(iii) $\mathrm{Prob}(A\&B/C) = \mathrm{Prob}(A/C).\mathrm{Prob}(B/C)$
(iv) $\mathrm{Prob}(A\&B/\text{not-}C) = \mathrm{Prob}(A/\text{not-}C).\mathrm{Prob}(B/\text{not-}C)$.

Let us also restrict our attention to those cases in which the A-type events and the B-type events occur simultaneously, and the C-type events occur either before or after the A-type and B-type events. The common cause principle then comes in two parts.

(1) Whenever there are correlated simultaneous A- and B-type events, which are not directly causally connected, there are C-type events such that the three types of events form a conjunctive fork.

(2) Whenever a conjunctive fork exists, the C-type events occur *before* the A- and B-type events, and never *after* the A- and B-type events.

In this paper I shall concentrate on the claim of temporal asymmetry made in the second part of the common cause principle. Hence, I shall not discuss here what might be meant by a 'direct causal connection' between events.

One can state the second part more pictorially, as the claim that conjunctive forks can point towards the future, but cannot point towards the past. A common cause explanation of a correlation consists of the finding of earlier events C which raise the probability of each of the correlated events, conditional upon which the correlated events are statistically independent.

Let me immediately amend my formulation of the common cause principle. It is possible that a C-type event *determines* an E-type event to occur at an even later time if and only if C occurred at the earlier time. But if this is the case then E must have exactly the same probabilistic relations with A and B as C does. Thus if A and B form a conjunctive fork with the earlier event C then A and B must also form a conjunctive fork with the later event E. Hence conjunctive forks which point to the past can exist. The amended second clause of the common cause principle therefore states that conjunctive forks which point to the past are always accompanied by conjunctive forks which point to future. Two conjunctive forks connected in such a way are called a closed fork, as opposed to a conjunctive fork occurring solo, which is known as an open fork. The amended second clause of the common cause principle thus reads: conjunctive forks which are open are always open to the future.

3. CLASSICAL PHYSICS AND COMMON CAUSES

It would be nice if one could have more than a few plausible examples to support the common cause principle. T. Horwich (Horwich 1987) and D. Papineau (Papineau 1985) have attempted to provide a more

solid foundation of the principle, by assuming background determinism and some randomness condition on the distribution of initial conditions.

If one is to assume background determinism then the probabilities mentioned above must refer to probabilities of initial conditions and not to the probabilities of transitions of states, as these will be 1 or 0 in deterministic contexts. For instance, in the example concerned with smoking, yellow fingers, and cancer, to assume background determinism is to assume that for any particular person, whether smoker or not, there is a factor c, the presence or absence of which determines whether or not he will later develop cancer, and that for any person there is a factor y, the presence or absence of which determines whether the person will develop yellow fingers or not. The claim that persons who smoke are more likely to develop cancer then translates into the claim that for persons who smoke factor c is more likely to be present than for persons who do not smoke. Similarly factor y is assumed to be more likely for smokers than for non-smokers. (One might wish to call the factor which determines whether a smoker develops cancer factor *sc* and that which determines whether a non-smoker develops cancer factor *nc*. Factor c then is (smoker and *sc*) or (non-smoker and *nc*). Similarly for y.

In order to construct a common cause explanation of the correlation between yellow fingers and cancer, both Horwich and Papineau now make the assumption that the factors y and c, are statistically independent among smokers, and that the factors y and c, are statistically independent among non-smokers. Indeed it then follows immediately that yellow fingers and cancer will be correlated amongst all people at the later time. The assumption that smokers are more likely to have factors y and c, together with the assumption that these factors are statistically independent among smokers and among non-smokers, is enough to imply a later correlation between smoking and yellow fingers among all people.

In order to derive the second clause of the common cause principle they now have to show that a common effect does not similarly lead to a correlation among its causes. To put it more clearly: they have to show that one cannot find conjunctive forks open to the past.

In view of background determinism there must be a factor a which determines for every person whether he previously worked in an asbestos factory or not, and a factor s which determines whether he previously smoked. If one now assumed that people with cancer were more likely to have factors a and s than people without cancer, and that the

factors were statistically independent among people with cancer and among people without cancer, then an earlier correlation between working in asbestos factories and smoking would be entailed. Exactly the same reasoning as in the case of yellow fingers and cancer would entail that cancer would form a conjunctive fork with working in asbestos factories and smoking. This would be a conjunctive fork open to the past.

This argument, though, is blocked by Papineau and Horwich on the grounds that initial conditions can be assumed to satisfy the requirement of statistical independence (a premise needed in the yellow fingers-cancer argument), but that final conditions *cannot* be assumed to satisfy such a requirement of statistical independence (which is a premise needed for the asbestos-smoking argument). Horwich in fact claims that such a statistical independence is implied by the assumption of microscopic chaos in statistical mechanics, whereas Papineau only points out that there are certain similarities between the independence requirement and certain assumptions about randomness in statistical mechanics.

Before using the full power of the assumption of background determinism it is instructive to see that, even in this example, Papineau and Horwich are not simply assuming that the initial conditions are chaotic where the final ones are not. The later correlation between yellow fingers and cancer *among all people* is mirrored by the earlier correlation between factors y and c *among all people*. This has to be the case since one has factor y if and only if one develops yellow fingers and c if and only if one develops cancer. Whatever correlation exists between yellow fingers and cancer at the later time must equal the correlation between y and c, at the earlier time.

Papineau and Horwich have simply run into Simpson's paradox: two factors can be uncorrelated over each of several parts of a space, and yet be correlated over the whole space. But once one realizes that (microscopic) factors y and c are initially correlated amongst all people in the Horwich-Papineau-scenario, it becomes unclear why initial microscopic chaos, or anything similar, should imply the uncorrelatedness of y and c, among smokers and among non-smokers. Why should it not entail the uncorrelatedness of y and c, among all people? Generalizing this problem, we shall see that one cannot only argue against Horwich's purported derivation of the common cause principle, but also show that the common cause principle is false in classical deterministic contexts.

At this point let me apologize for the fact that I will sometimes speak of events occurring at certain times, and sometimes of systems having certain properties at certain times. I am assuming that a system having a certain property p at time t corresponds to the occurrence of an event p at time t. I shall use these locutions as being interchangeable. In any case I am not, for the present, concerned whether causation is a relation between events or a relation between properties, or otherwise. I am interested to assess whether the common cause principle, explicated as a claim about the probabilistic relations between events, properties or otherwise, can be upheld in the face of a deterministic development of such events, properties or otherwise.

In a deterministic theory the state of an isolated system at any time t_0 determines uniquely the state of the system at any other time t. Let us call the set of all possible states of such a system the statespace of the system. Any property of such a system which is expressible in the deterministic theory will correspond to a subset of the statespace. The set of states corresponding to a property is just the set of all states in which the system has that property. (The set of states corresponding to an event is just the collection of all the states in which that event occurs.)

In order to introduce the probabilities mentioned in the common cause principle one has to consider probability distributions over a statespace. For classically deterministic systems, i.e , Hamiltonian systems, the development of such probability distributions in time is that of an incompressible fluid, a so-called Liouville flow. A Hamiltonian development together with two times t_0 and t_1 induces a 1–1 mapping $F(p)$ of properties p onto themselves such that an isolated system has property $F(p)$ at time t_1 if and only if the system has property p at time t_0. This means that for any two times t_0 and t_1 and any property A there is a unique property B which the system will have at time t_1 if and only if the system has property A at time t_o. For the development of probability distributions of states according to the classical laws this means that for any two times t_o and t_1 and for any two properties A and B there are unique properties C, and D which must have exactly the same correlation at time t_1 as properties A and B have at time t_o, whatever initial probability distribution one assumes.

The preceding can be summarized in the slogan 'Correlations are not born and do not die, they merely change variables'. In view of this it is simply false to claim that all initial properties are uncorrelated and some final ones are not. At any time there are exactly as many corre-

lated properties as at any other time, in the sense that for any two times one has a 1–1 mapping between properties which preserves all correlations.

As an illustration, consider again the case where one has as initial events smoking at time t_0, and factors y and c, which are statistically independent among smokers at time t_0 and are statistically independent among non-smokers at time t_0. If being a smoker is correlated to having factors y and c, then this entails a correlation between yellow fingers and cancer at the later time t_1. However, assuming determinism, there must be factors s', y' and $c,'$ which are present at an even later time t_2, if and only if smoking, y and c, respectively were present at t_0. But then, the statistical independence of y' and c', among s' and among not-s' at t_2 is guaranteed, and y' and s' will be correlated at time t_2, as will c' and s'. But these facts together imply the correlation of yellow fingers and cancer at the earlier time t_1. Just as smoking at t_0 forms a conjunctive fork with yellow fingers and cancer at t_1, s' at the later time t_2 forms a conjunctive fork with yellow fingers and cancer at the earlier time t_1.

Whenever there is a conjunctive fork open to the future there is a conjunctive fork with the same endpoints open to the past, which thus closes the fork. *Every fork is a closed fork.* One constructs this fork by finding the event at an even later time which occurs if and only if the common cause occurs at the earlier time. By determinism one knows that such a property exists, albeit perhaps not a property which is easy to observe. The second clause of the common cause principle is therefore false for classically deterministic systems: there is no such asymmetry.

Before seeing what can be salvaged from this wreckage, let us note that D. Lewis's account of the asymmetry of causation is also false in classical deterministic contexts. In D. Lewis's paper on counterfactual dependence and time's arrow (Lewis 1979), he claims that the asymmetry of causation (and certain other asymmetries) depend on what he calls the asymmetry of overdetermination. Lewis claims that in our world, it is de facto (contingently) true that past facts (states of affairs) are overdetermined by future facts (states of affairs) while future facts are not so overdetermined by past facts. He claims that therefore a counterfactually different past normally would result in a different future, whereas a counterfactually different future would not normally result in a different past, which in turn means that causation should ordinarily run in the forwards direction of time. However, given his

own characterization of what it is for a fact to be overdetermined, his claim is simply false in classical deterministic contexts. Lewis writes:

> Any particular fact about a deterministic world is predetermined throughout the past and postdetermined throughout the future. At any time, past or future, it has at least one *determinant*: a minimal set of conditions jointly sufficient, given the laws of nature, for the fact in question. (Members of such a set may be causes of the fact, or traces of it, or neither.) The fact may have only one determinant at a given time, disregarding essential differences in a way I shall not try to make precise. Or it may have two or more essentially different determinants at a given time, each sufficient by itself. If so it is *overdetermined* at that time. (Lewis 1979, pp. 473–74)

However, as we have seen above, in classically deterministic contexts there is for any fact at any time a unique determinant at any other time. The determinant at any time t for a given fact at any other time t_0 is just the largest set in statespace (the minimal set of conditions) which determines the fact to occur at the later time (which is sufficient for the fact) according to the laws of nature. Further down the page Lewis gives a supposed example of such overdetermination: a spherical wave spreading out from a point source. He claims that countless tiny samples of the wave each determine what happened at the space-time point where the wave is emitted. This is just false. Given the deterministic wave laws of nature, such samples by themselves do not determine what happened at the source.[1] There is, in any case, no need to consider the plausibility of certain examples. Classical determinism allows only one determinant for each fact. Thus we see that two at first glance plausible attempts to find a solid foundation for the asymmetry of causation fail when faced with certain simple properties of classical systems.

4. EVASIVE MANOEUVRES

At this point two tactics are open to us, if we are to save the common cause principle. The first option is to claim that the principle relies on the assumption that one does not have background classical determinism. The second option is to amend the principle. For instance, one could deny that the common cause principle is true about *all properties in physical state space*. One could for instance claim that it only holds for a subset of all the properties in state space, or that it holds for properties that are outside the physical state space.

The former approach appears implausible to me. The only serious

candidate for a fundamental theory of physics that is indeterministic is quantum mechanics. It seems unlikely to me that the common cause principle would be valid in a quantum world and invalid in a classical world. Thankfully, we do not have to speculate on this matter, as there are compelling arguments to the effect that the common cause principle is in even worse trouble if the laws of physics are quantum mechanical rather than classical.

In the first place (and this has been noted by several authors), certain quantum mechanical phenomena known as EPR-phenomena present us with correlations for which there cannot be a common cause explanation: the first clause of the common cause principle fails in quantum mechanics. For details see, e.g., Suppes and Zanotti 1981, Fine 1981, or van Fraassen 1982. In the second place, even without such examples quantum mechanics runs into the same problems I have indicated above for classically deterministic systems: the second clause of the common cause principle fails in quantum mechanics. This is so because the Hamiltonian development of classical physics is mirrored by a unitary development in quantum mechanics. This unitary development entails that for any two times t_0 and t_1 and any observable A, there is an observable

$$A' = \exp(-iH(t_1-t_0)) \,.\, A \,.\, \exp(iH(t_1-_0),$$

which must have exactly the same probability distribution of values at time t_1 as A has at time t_0. Thus, for instance, if the probability distribution of the values of a bivalent observable A at some time t_0 provide a common cause explanation for the correlation between the values of bivalent commuting observables B and C, at a later time t_1, then there will be at any other time t_2 (earlier or later) an observable A', the values of which at time t_2 will have the same probabilistic relations with the values of B and C, at time t_1 as the values of A at time t_0 have with the values of B and C, at time t_1. Thus, in addition to the fact that certain correlations in quantum mechanics cannot be part of conjunctive forks, the conjunctive forks that there are, are all part of a closed fork. The first approach does not appear attractive.

Perhaps then, one ought to have been more careful about which properties one is talking about in the common cause principle. One could hypothesize that macroscopic common causes produce correlations among macroscopic effects, but that macroscopic common effects do not produce correlations among macroscopic causes. Since one

normally only observes macroscopic properties, this might explain the sense in which the common cause is observed to be valid. One observes yellow fingers and cancer and the macroscopic common cause smoking. There will be some effect at a much later time which occurs if and only if one smokes at the earlier time, and hence has the same probabilistic relations with yellow fingers and cancer as smoking does, but it is very unlikely to be an easily observable effect. Thus the observed conjunctive forks might never be open to the past.

Before accepting such an amended second clause of the common cause principle I would like to see a clear distinction made between macroscopic observables and microscopic ones. But even assuming that one can make such a distinction, it is not at all obvious why one could not similarly claim that whenever one has a macroscopic conjunctive fork open to the past it is not going to be closed by a macroscopic earlier common cause. For the backwards development of the later macroscopic event of the conjunctive fork which is open to the past, is similarly very unlikely to be an easily observable property. Perhaps, on the basis of the assumed complexity of the development of most Liouville flows, one can only conclude that each macroscopic conjunctive fork will be open in one direction of time, but not that they will all be open in the same (future) direction of time. Let me illustrate such a suspicion with an example.

Cleopatra is throwing a big party, and wants to sacrifice about fifty slaves to appease the gods. She is having a hard time convincing the slaves that this is a good idea, and decides that she ought to give them a chance at least. She has obtained a very strong poison, so strong that one molecule of it will kill a person. She puts one molecule of the poison in each of a hundred goblets of wine, which she presents to one hundred slaves. Having let the molecules of poison move around in Brownian motion for a while, she then orders the slaves to drink half a goblet of wine each. Let us now assume that if one consumes the poison then in many cases death is preceded by an ominous reddening of the left hand, or by a reddening of the right hand or by both. Let us also assume that, given that one has swallowed the poison, the reddening of the left hand and that of the right hand are statistically independent. This being the case, one will observe a macroscopic conjunctive fork open to the past. Death will form a conjunctive fork with the reddening of left hands and the reddening of right hands, and left hand reddening will be correlated with right hand reddening among all

slaves. Of course, the molecule of poison being in the swallowed half will also form a conjunctive fork with left hand reddening and right hand reddening. However, the position of the poisonous molecule should hardly count as a macroscopic property. In this case it is precisely the assumption of microscopic chaos (Brownian motion) which guarantees that the macroscopic conjunctive fork will be open to the past.

Perhaps one must conclude that one prefers a common cause explanation to a common effect explanation because one is inclined in favour of explanation in the forwards direction of time. Perhaps whenever one runs into a correlation with an obvious common effect explanation one hypothesizes as a knee-jerk reaction some unobserved common cause, such as positions of molecules. One knows by determinism that such is a safe strategy, since there must be such a common cause explanation if there is a common effect explanation. In the next section I will indicate why I believe that there is still hope for the common cause principle.

5. PROBABILISTIC CAUSALITY AND MARKOV PROCESSES

Let us assume that the development of states in time in some state space is not deterministic, but governed by fixed probabilities of transition p_{ij} from state s_i at time t to state s_j at time t' where $(t'-t)>0$, and the p_{ij} depend only on the time difference $(t'-t)$. (For uncountable state spaces one has to replace the transition probabilities by a transition function $p(x,A,t'-t)$. For details, which are not relevant for this paper, see, e.g., Doob 1953.) For probability distributions of such states developing according to such transition probabilities, i.e., for stationary Markov processes, one can in certain cases prove that the entropy (a measure of the spread of such distributions over state space) must increase in time (See for instance Doob 1953, Kelly 1979). If one could prove that under a Hamiltonian development of the microscopic states certain coarse-grained states would develop according to such fixed transition probabilities then one would have gone a long way towards a reconciliation of the time irreversibility of thermodynamics with the reversibility of the underlying microscopic laws of physics. (For attempts to found the irreversibility of thermodynamics on the properties of Markov processes see, e.g., O. Penrose 1970, and P. and T. Ehrenfest 1959; for attempts to derive Markov properties from Hamiltonian

physics see, e.g., Misra, Prigogine, and Courbage 1979, Goldstein, Misra, and Courbage 1981, or Prigogine 1980.)

Let us assume for the purposes of this section that one can indeed prove that the Hamiltonian development of microscopic states entails a development in some coarse-grained statespace according to given (forwards) transition probabilities (not always equalling 0 or 1), which are fixed by the Hamiltonian in question. What does this entail about the common cause principle, and what does it mean for probabilistic accounts of causality?

Consider defining causation in the following manner for a set of given (forwards) transition probabilities in statespace. An event A which occurs at time t_0 is the cause of an event B which occurs at time t_1 precisely when A is the 'largest' event (the largest set in statespace) at time t_0 which fixes the probability of event B at time t_1. An event A at t_0 is said to 'fix' the probability of B at t_1, given a set of transition probabilities $p_{ij}(t'-t)$, if for any points i and i' in the set A one has Prob(B at t_1/i at t_0) = Prob(B at t_1/i' at t_o), for any initial probability distribution of states. (The qualification 'largest' is inserted into the definition so as to rule out events which are redundant regarding an effect as being part of the cause of that effect. If smoking is the cause of lung cancer then I would not wish to imply that 'smoking and having blue eyes' was the cause of lung cancer. This qualification is not essential for most of the points I make below, but it is essential if one wishes to have one cause rather than many causes of a particular event.)

The above definition yields an asymmetric cause-effect relationship! Given any particular set of transition probabilities and any event B at any time t_1 there will be at any earlier time t_0 a unique cause A of the event B at the later time t_1. However, there will not be a cause of the event B at t_1 at a time t_2 later than t_1 (unless there is an event C which occurs at time t_2 if and only if B occurs at time t_1, i.e., unless the transition probability from B to C, in time interval t_2–t_1 equals 1, and the transition probability from anywhere outside B to C equals zero).

This is so because the transition probabilities of the reversed Markov process, given the forwards transition probabilities, depend on the initial distribution (unless the transition probabilities are equal to 0 or 1). Thus causation on this definition always runs forwards in time, (except between pairs of events between which the transition probabilities are equal to 1).

Let me give an example to illustrate this asymmetry. Suppose that

the microscopic laws of physics somehow entail that the probability of a transition from smoking to having lung cancer after some given time interval t_1–t_0 is 1/2 (irrespective of the absence or presence of other coarse-grained factors). Let us also assume that the probability for a non-smoker to develop lung cancer after the time interval is 1/10 (irrespective of other coarse-grained factors). Now consider the development of two distinct initial probability distributions.

CASE 1. Initially Prob(smoker at $t_0 = 0.9$). So, Prob(non-smoker at t_0) = 0.1.

Then, according to the assumed (forwards) transition probabilities

Prob(smoker at t_0 and cancer at t_1) = 0.45,
Prob(smoker at t_0 and not cancer at t_1) = 0.45,
Prob(not smoker at t_0 and cancer at t_1) = 0.01,
Prob(not smoker at t_0 and not cancer at t_1) = 0.09.

Therefore, Prob(smoker at t_0/cancer at t_1) = 0.45/0.46.

CASE 2. Initially Prob(smoker at t_0) = 0.1. Then, according to the assumed (forwards) transition probabilities

Prob(smoker at t_0 and cancer at t_1) = 0.05,
Prob(smoker at t_0 and not cancer at t_1) = 0.05,
Prob(not smoker at t_0 and cancer at t_1) = 0.09,
Prob(not smoker at t_0 and not smoker at t_1) = 0.81.

Therefore, Prob(smoker at t_0/cancer at t_1) = 0.05/0.14.

For the given (forwards) transition probabilities, smoking fixes the probability of having cancer after time interval t_1–t_0. Prob(cancer at t_1/smoking at t_0) is independent of the initial Prob(smoking at t_0). However, the reverse conditional probability Prob(smoking at t_0/cancer at t_1) *does* depend on the initial Prob(smoking at t_0). Having cancer at time t_1 does *not* fix the probability of smoking at time t_0. Thus, given such forwards transition probabilities, and given the above definition of causation, smoking causes cancer, but cancer does not cause smoking.

One might feel that I have cheated myself into an asymmetry by the assumption that the microscopic laws of physics fix the forwards transition probabilities rather than the backwards transition probabilities. I

will admit that I believe that one will have to inject some de facto asymmetry into statistical mechanics if one is to derive such an asymmetry, and I will admit that the project of deriving the Markovian development for some coarse-grained state space is far from finished. It is interesting, however, to see that such an asymmetry in time can be made to imply a causal asymmetry given a very simple definition of causation. Moreover, although it might be difficult to derive the asymmetry from fundamental physics, the result is certainly in accord with common sense. It certainly seems to be the case that if one smokes one has fixed the probability that one will develop lung cancer irrespective of the amount of other people that smoke or do not smoke, whereas the occurrence of lung cancer does not at all fix the probability that one smoked, irrespective of the initial numbers concerned. Before smoking occurred on a large scale in Europe, the probability that one was a smoker given that one had lung cancer was very small, whereas nowadays it is a reasonable bet that a European with lung cancer has been a smoker. In contrast, the proportion of smokers that developed lung cancer has presumably remained fairly constant. If this proportion has not remained the same, then this presumably is due to the fact that other coarse-grained factors influence the probability of the development of lung cancer, and that the distribution of these factors among smokers has changed. It is presumably not the case that the forwards transition probabilities in coarse-grained statespace have altered.

According to the above definition of causation in terms of Markov processes, whether some event is the cause of another event does not depend merely on the statistics of the occurrences of the respective events in the world. Whether an event A is the cause of an event B does not just depend on the proportion of A-type events that have been followed or preceded by B-type events, but on the invariance of the probabilistic relations between A- and B-type events under changes of initial distributions of states. This is the analogue of the assumption in the deterministic case that whether an event A causes an event B does not merely depend on the actual occurrence of events A and B, but on the lawlike relations between the occurrence of A and B. If one wishes to define causality in probabilistic contexts, it should be sought in the lawlike transition probabilities, and not in quantities dependent on initial distributions (initial states).

Let me illustrate this claim by an example. Consider the following transition probabilities.

(i) $\text{Prob}(E(t_1)/A(t_0)\&B(t_0)) = 4/5$,
(ii) $\text{Prob}(E(t_1)/\text{not-}A(t_0)\&B(t_0)) = 0$,
(iii) $\text{Prob}(E(t_1)/A(t_0)\&\text{not-}B(t_0) = 0$,
(iv) $\text{Prob}(E(t_1)/\text{not-}A(t_0)\&\text{not-}B(t_0) = 0$,

for some fixed time interval $t_1 - t_0 > 0$.

Assume the following initial amounts:

$$5\ A\&B,\ 1\ A\&\text{not-}B,\ 1\ \text{not-}A\&B,\ 1\ \text{not-}A\&\text{not-}B.$$

Of the 5 $A\&B$, 4 are likely to transit to event E, and 1 to not-E. The others will all transit to not-E. The observed frequencies will satisfy the following relations:

(i) $\text{Freq}(A(t_0)/E(t_1)) = 1$
(ii) $\text{Freq}(B(t_0)/E(t_1)) = 1$
(iii) $\text{Freq}(A(t_0)\&B(t_0)/E(t_1)) = 1 = 1.1$
(iv) $\text{Freq}(A(t_0)/\text{not-}E(t_1)) = 2/4 = 1/2$
(v) $\text{Freq}(B(t_0)/\text{not-}E(t_1)) = 2/4 = 1/2$
(vi) $\text{Freq}(A(t_0)\&B(t_0)/\text{not-}E(t_1)) = 1/4 = 1/2.1/2!$

According to the observed frequencies the occurrences of the event E at the later time t_1 form a conjunctive fork with the occurrences of the events A and B at the earlier time t_0. Thus, if common causes were simply spelled out in terms of observed frequencies one might have to classify later events as the common cause of correlations between earlier events. One does not feel that this is right. My claim is that at least one of the reasons that one feels that this is not right is that the above conjunctive fork formed by the observed frequencies is extremely sensitive to changes in the initial numbers. Any change in the numbers of initial A's and B's will destroy the backwards conjunctive fork, if the transition probabilities are kept fixed. On the other hand, if one has forward transition probabilities which form a conjunctive fork, then the observed frequencies will (be likely to) form a conjunctive fork open to the future no matter what the initial distribution of states is. It is this lawlike regularity which manifests itself in different circumstances (given different initial distributions) which is the obvious candidate for a causal relationship, if one wishes to define causality in probabilistic contexts.

Let me illustrate this claim yet again. It has been argued by E. Sober (1987) that certain correlations might not need a common cause

explanation. For instance the fact that the price of bread in Britain has been correlated to the level of water in Venice appears in no immediate need of a common cause explanation. My claim would be that one believes that there is no need for a common cause (or other) explanation in so far as one believes that the correlation is not invariant under changes of initial distributions of states. One presumably feels that for a different distribution of coarse-grained states the bread price need not at all be correlated to the water level, but that the presence of any (not too small) amount of smokers would always be likely to lead to a correlation between cancer and yellow fingers. The ultimate justification of such a belief might be the exact properties of the Liouville flow which undergirds the Markovian development of the coarse-grained properties, but in the meantime we can guess which the events are that fix the probabilities of later events. On the above definition of causality for Markov processes we would then be making conjectures about the causal structure of coarse-grained statespace.

Before considering what would happen to the common cause principle for Markovian developments of probability distributions, let me remark on the connection between the above definition of causation for Markov processes and certain well-known probabilistic accounts of causation. The condition that the cause at time t_0 of an event at time t_1 must fix the probability of the event at time t_1 irrespective of the initial distribution assumed at time t_0, has as an immediate consequence that the cause of an event will make any other event (expressible in the statespace) occurring at the same time as the cause, statistically irrelevant. The cause of an effect 'screens off' any other event from the effect. Given that the cause has occurred it is irrelevant which other events occurred at the same time as the cause, regarding the probability of the occurrence of the effect at the later time. If according to the Markovian development in coarse-grained statespace smoking fixes the probability of a later development of lung cancer, then it becomes irrelevant regarding the likelihood of lung cancer what other things were happening when a particular person was smoking. This screening-off relationship has, by some, been taken to indicate a causal relationship (see, for instance, Salmon 1984, Suppes 1984). On the above definition the cause of an effect screens off all events simultaneous with the cause from the effect.

On my account the cause of an event B at time t_1 for a deterministic theory will be the event A at time t_0 which occurs at time t_0 if and only

if B occurs at time t_1. A deterministic theory is a trivial Markov process with all transition probabilities equalling 0 or 1. The largest event in statespace whose occurrence at time t_0 fixes the probability of the occurrence of some other event at time t_1 is simply the event which occurs at t_0 if and only if the other event occurs at t_1 for such trivial, deterministic, Markov processes.

Let us call an event A which screens off all other events occurring simultaneous with A regarding some event B which occurs some time after A 'causally homogeneous' with respect to B (following Suppes 1984). It follows from this definition that for deterministic theories an event which is causally homogeneous with respect to an effect must be an event which determines the effect to occur, or determines it not to occur. A set in statespace which contains some states which will develop according to the deterministic laws into states for which the effect does occur, and some states which will develop into states for which the effect in question does not occur, is not causally homogeneous with respect to that effect.

However, for a Markov process with non-trivial transition probabilities, i.e., for a genuine Markov process, events which are causally homogeneous with respect to some later event will not in general determine the later event to occur with probability 1 or 0. One can think of causal homogeneity in the following manner. A property A (an event A) is causally homogeneous with respect to the occurrence of some other property B (event B) after a particular time interval if the transition probability from any state with property A to a state with property B after the time interval equals the probability that any other state with property A will transit to a state with property B. A property A is causally homogeneous with respect to the later occurrence of a property B if no subselection of states with property A affects the probability of the occurrence of B at the later time.

The collection of 'largest' properties $[A_i]$ (the largest sets in state-space) at time t_0, which are causally homogeneous with respect to the occurrence of B at t_1 form a partition of statespace, such that each of the cells of the partition have a fixed probability of transition to property B at t_1. Such a partition could consist of just the one set, namely the entire state-space. This would occur if the conditional probability of the occurrence of B is uniform over the entire statespace. For instance, for consecutive tosses of a coin the probability of any particular number is uniformly 1/6 conditional upon any previous result. Alternatively,

the partition could consist of two sets A and not-A. For instance, if indeed the probability of lung cancer were solely determined by the fact whether one is a smoker or not. It may very well be the case, however, that a property A is causally homogeneous with respect to the later occurrence of a property B, but that not-A is not causally homogeneous with respect to the later occurrence of B, so that a further partition is necessary to obtain causally homogeneous sets. In fact, the transition probabilities may be such that the only causally homogeneous sets are just each of the states themselves (the singleton sets in state-space). However for Markov processes for any property B causally homogeneous properties at earlier times must exist, and which ones they are is determined by the transition probabilities. The associated problem of the characterization of what it is to be a homogeneous reference class does not occur in this setting: the statespace properties together with the transition probabilities *determine* which are the homogeneous reference classes for which properties occurring at later times. Hence some of the nagging problems of well-known probabilistic accounts of causality disappear for the suggested definition of causality in Markovian contexts.

Finally, let me consider the common cause principle against such a backdrop. As noted above, it may be the case for a particular property B that there is no property A such that both A and not-A are causally homogeneous with respect to the occurrence of B after some specified time interval. Thus it is certainly not guaranteed that for a particular Markov process and two correlated properties A and B there will be a property C such that C will be causally homogeneous with respect to the later occurrence of each of the properties $A\&B$, $A\&$not-B, not-$A\&B$, and of not-$A\&$not-B. However if C is not causally homogeneous with respect to one of these properties then the Markov process does not yield a transition probability from C to that property *independent of the initial distribution*. In view of this let us rephrase the common cause principle for Markov processes in the following manner.

A set of transition probabilities in some statespace, a partition of statespace $[C_i]$ and an initial distribution $P(C_i)$ at time t_0 form a common cause explanation of a joint distribution $P^1(A, B)$ at time t_1 of bivalent properties A and B if and only if:

(i) each of the C_i at t_0 are causally homogeneous with respect

to the later occurrence of $A\&B$, $A\&$not-B, not-$A\&B$, not-$A\&$not-B at t_1,

(ii) the distribution $P(C_i)$ at t_0 to develops into the distribution $P^1(A, B)$ at t_1 according to the transition probabilities, and

(iii) for each C_i, Prob($A\&B$ at t_1/C_i at t_0) = Prob(A at t_1/C_i at t_0). Prob(B at t_1/C_i at t_0), i.e., A and B must be statistically independent conditional upon each of the C_i.

For distributions developing according to fixed transition probabilities, we have seen that for any two properties A and B with distribution $P^1(A, B)$ at time t_1 one can always find at any earlier time t_0 a partition $[C_i]$ which has distribution $P(Ci)$ at time t_0 and which will satisfy conditions (i) and (ii). However such a partition may not satisfy condition (iii). Whether there is a partition satisfying condition (iii) depends on the transition probabilities of the Markov process in question. If such transition probabilities in some coarse-grained statespace have their grounding in the Liouville flow of the underlying fine-grained statespace, then the properties of the Liouville flow will ultimately determine whether the common cause principle holds in the coarse-grained statespace.

Let me summarize. I have suggested that if one is to have a probabilistic account of causation, one ought to relate causation to the existence of transition probabilities which are invariant under changes of initial distributions. I have also suggested that the origin of the transition probabilities, which common sense continuously attributes to observed events, does not usually lie in quantum mechanics, but in the consequences of complex developments in fine-grained statespace for the developments of coarse-grained states. Although we have found that on our suggested definition of causation for processes developing according to such fixed transition probabilities, any event always has a unique earlier cause (and never a later cause), it may be the case that even outside quantum mechanics there are correlations for which there are no common cause explanations. If so, so be it.

NOTES

*This paper was written while I was on an Andrew Mellon postdoctoral fellowship at the University of Pittsburgh, for which I am grateful. I am also grateful for comments from John Norton and an anonymous referee.

[1]They do, if one makes the additional assumptions that there are no source free waves, that there is only one source, that there are no obstacles, mirrors etc., and, more pedantically, that space does not have any 'funny' topological connections. But, in that case, one has just as much forwards 'overdetermination'. A sample wavelet at some point x at a distance r from the source at time t is 'overdetermined' at the earlier time t-dt by each sample at every point on a circle of radius r-dtv, where v is the speed of the wave (I assume a dispersion free space).

REFERENCES

Doob, J. L.: 1953, *Stochastic Processes*, A. I. Wylie, New York.

Ehrenfest, P. and T. Ehrenfest: 1959, *The Conceptual Foundation of the Statistical Approach to Mechanics*, Cornell University Press, Ithaca, New York.

Fine, A.: 1981, 'Joint Distributions, Quantum Correlations and Commuting Observables', *J. Math. Phys.* **23,** 1306–10.

van Fraassen, B.: 1982, 'The Charybdis of Realism: Epistemological Implications of Bell's Inequality', *Synthese* **52,** 25–38.

Goldstein, S., B. Mistra, and M. Courbage: 1979, 'On Intrinsic Randomness of Statistical Systems'. *Journal of Statistical Physics* **25,** 111–26.

Horwich, P.: 1987, *Asymmetries in Time*, MIT Press, Cambridge, Massachusetts.

Kelly, F. P.: 1979, *Reversibility and Stochastic Networks*, A. I. Wylie, New York.

Lewis, D.: 1979, 'Counterfactual Dependence and Time's Arrow', *Nous* **13,** 455–76.

Misra, B., I. Prigogine, and M. Courbage: 1981, 'From Deterministic Dynamics to Probabilistic Descriptions', *Physica* **98A,** 1–26.

Papineau, D.: 1985, 'Causal Asymmetry', *British Journal for the Philosophy of Science* **36,** 273–89.

Penrose, O.: 1970, *Foundations of Statistical Mechanics*, Pergamon Press, Oxford.

Prigogine, I.: 1980, *From Being to Becoming*, W. H. Freeman, San Fransisco.

Reichenbach, H.: 1956, *The Direction of Time*, University of California Press, Berkeley.

Salmon, W.: 1984, *Scientific Explanation and the Causal Structure of the World*, Princeton University Press, Princeton, New Jersey.

Sober, E.: 1987, 'The Principle of the Common Cause', preprint.

Suppes, P.: 1984, *Probabilistic Metaphysics*, Basil Blackwell, Oxford.

Suppes, P. and M. Zanotti: 1981, 'When Are Probabilistic Explanations Possible?', *Synthese* **52,** 25–38.

School of Philosophy
University of Southern California
Los Angeles, CA 90089-0451
U.S.A.