CAUSAL FACTORS, CAUSAL INFERENCE, CAUSAL EXPLANATION

Elliott Sober and David Papineau

I—Elliott Sober

I

Two Concepts of Cause

What is it for smoking to be a positive causal factor in the production of heart attacks among U.S. adults? The probabilistic theory of causality answers that smoking must raise each individual's chance of a coronary or, more modestly, that smoking must raise at least one individual's chance, but not lower anyone else's.[1] This theory provides an account of the concept of *property causation*.

Smoking may increase different individuals' chances of coronaries by different amounts, since individuals may differ in relevant physical ways. Some individuals exercise while others do not; some have high cholesterol diets while others do not; and so on. Let each B_i be a maximal conjunction of all such relevant background factors so that the set of all the B_i's encompasses all possible combinations of factors (including the presence or absence of each). The probabilistic theory of causality asserts that smoking (C) is a positive causal factor in producing heart attacks (E) in the population if and only if

$$\Pr(E/C \ \& \ B_i) > \Pr(E/\text{not-}C \ \& \ B_i), \text{ for each } B_i.$$

This theory, defended in various forms by Good [1961-2], Suppes [1970], Cartwright [1978], Skyrms [1980], Eells and Sober [1983], and Sober [1984b, 1985], has been criticized by Hesslow [1976], Salmon [1980], and Otte [1981]. Some of these criticisms can be reinterpreted as identifying contingent assumptions that the theory must impose. First, the probabilities linking causal factors to their effects must be *intermediate*, since deterministic relationships render relevant probabilities undefined. Second, correlated effects must possess a common cause

[1] I will ignore this latter, more plausible and weaker, formulation, due to Skyrms [1980], since my arguments apply equally to both.

that *screens off* each from the other; without this principle of the common cause, on which see Reichenbach [1956], Van Fraassen [1980], and Salmon [1984], the probabilistic theory of causality can erroneously judge an event merely correlated with an effect to be a cause of it. These two restrictions, plus some care in specifying what a 'relevant background context' is, on which see Eells and Sober [1983] and Eells [1985], suffice to insure the correctness of the probabilistic theory.

What is it for Harry's smoking to cause him to have a heart attack? It is neither necessary nor sufficient that smoking should be a positive causal factor for coronaries in the population containing Harry—among U.S. adults, say. It is not necessary, since if smoking increased some individuals' chances and reduced those of others, this would mean that smoking is not a positive causal factor in the population as a whole; there would be no such thing as *the* causal role it plays with respect to heart attacks in the entire population. But this need not prevent smoking from producing an infarction in poor Harry.

It is not sufficient, since smoking could be a positive causal factor, even if no one smoked, and even if no one had a heart attack. Smoking counts as a positive causal factor in virtue of a relationship between various conditional probabilities; these probabilities can be well defined even if the conditioned and conditioning properties are not in fact exemplified.[2] The singular fact about Harry involves the relation of *token causality*; Harry's smoking can't cause him to have a heart attack unless he actually smokes and actually has a heart attack. Token causality, unlike property causality, requires that the *relata* actually obtain.

The two concepts of cause differ in further ways. Even if Harry smoked and then had a heart attack, the fact about property causality does not suffice to ensure that his smoking actually caused his heart attack. The smoking may have placed him at greater risk, but some quite different cause may have produced the coronary. Furthermore, it is provable that property causality is not in general transitive (Suppes 1970, Eells and Sober 1983), but no plausible case has been made for

[2] This is a consequence of avoiding an actual relative frequency interpretation of probability, which there is ample reason in this context to do anyway.

thinking that token causality is not transitive (Sober 1984b, 1985).

In addition, examples that have identical probabilistic structure can differ over corresponding facts about token causation. Here I refer to the comparison of the kicked golf ball and the sprayed plant, discussed in Eells and Sober [1983] and Sober [1985]. A golf ball rolling towards the cup is kicked by a squirrel, which thereby lowers the ball's probability of going into the cup. Improbably enough, though, the ball ricochets here and there and then drops in. Here is a case in which the kick token-caused the ball to drop in, even though a kick of just that kind is not a positive causal factor for producing balls in cups.

Contrast this example with Cartwright's [1979] case of a healthy plant that is sprayed by a defoliant, which thereby lowers the plant's probability of surviving. Improbably enough, though, the plant survives. The spraying is not a positive causal factor for producing survival. However, in striking contrast to the squirrel's kick, we do not say, in this case, that the spraying token-caused the surviving. This pair of examples suggests, not only that token causes need not raise the probability of their effects, but also that token causation is not definable from probabilities alone.[3]

It is sometimes suggested that token causality involves energy transfer. This may be so, but the problem of the sprayed plant versus the kicked golf ball is not resolved by this observation. The squirrel transferred energy to the golf ball, but so too did the defoliant to the plant. A theory of token causality that appeals to energy transfer will have to give more details, if it is to resolve this puzzle.

To all these differences between property causation and token causation, I add another: Property causation is pretty well understood, at least if the assumptions made by the probabilistic theory are not too restrictive. But token causality is fascinating in its opacity. We know what it is not, but very little about what it is.

This last contrast would not matter, if token causality did not matter. But it does. In causally explaining a token event, we try

[3] I say 'suggests' since it doesn't necessarily follow that a more fine-grained look at these examples won't uncover probabilistic differences of a relevant sort. I myself am skeptical that this can be done, though.

to trace it back to the causes that produced it. Causal explanation is preeminently the search for token causes. Yet, we frequently try to fulfill requests for causal explanation by amassing facts about property causality. If our question is why Harry had a heart attack, we assemble information about his characteristics and identify which were positive causal factors (and which negative) for having a coronary. But this standard strategy raises a question: How can facts about property causality be relevant in our search for causal explanation, if token causality and property causality are two concepts, not one?

Hempel [1965] declined to require that an explanation cite a causal generalization, though he did demand that a law connect the initial conditions with the *explanandum* event. Some critics (e.g., Salmon [1971, 1984]) have required causal generalizations, while others (e.g., Scriven [1959]) have insisted that the law isn't part of the explanation proper, but merely serves to justify the claim that the initial conditions explain the *explanandum*. These two reactions to Hempel derived from very different considerations, but the distinction drawn here between property causality and token causality leads them to converge. For Harry's smoking to causally explain Harry's heart attack, it isn't enough that smoking causes heart attacks—what is essential is that *Harry's* smoking caused his heart attack.[4] Setting to one side the question of whether all explanation is causal explanation,[5] we may yet recognize that a causal explanation must cite a token causal relationship, not a property causal generalization. The role of the causal generalization is epistemological, as we shall now see.

II

The Connecting Principle

The principle I want to discuss describes an epistemological connection between the two concepts of cause. The rough idea is that if a token event of type *C* is followed by a token event of type

[4] I therefore am in agreement with Anscombe's [1971] contention that (token) causality does not consist in the existence of a lawful regularity. See Sober [1985] for further discussion.

[5] I argue that equilibrium explanations are a counterexample to the claim that all explanation is causal explanation in Sober [1983a] and [1984b].

E,[6] then the support of the hypothesis that the first event token-caused the second increases as the strength of the *property causal* relation of C to E does. So, in our coronary example, if Harry's smoking is followed by his having a coronary, then the hypothesis that the first token-caused the second is better supported to the degree that smoking is a strong property-cause of coronaries.

I will call this the Connecting Principle. The question immediately arises of how the idea of 'support' is to be represented. One could view this as a primitive concept or try to clarify it in terms of the notion of subjective probability. In the next section, I will follow a third approach: the likelihood of a hypothesis, relative to an observation, indicates how well the observation supports the hypothesis (Hacking [1965], Edwards [1972]). I must emphasize that the likelihood of a hypothesis is not the probability it receives from an observation, but is the probability it confers on the observation.

For the present, however, I will talk about a primitive support metric, expressed by '$S(H/E)$', which measures the support of H, given E. It can assume values between -1 and +1. I will use 't_i' to denote a token event that occurs at place-time i, and 'C' and 'E' to represent properties. '$C(t_i)$' also denotes an event—the event of place-time t_i's having the property C. The Connecting Principle can be stated as follows:

> If C is a causal factor for producing E in population P of magnitude m, then $S\{C(t_1)$ token caused $E(t_2)/C(t_1)$ and $E(t_2)$ occurred in population $P\} = m$.

This principle, if correct, shows how facts of property causality bear on claims of token causality.

How might one measure the strength of a causal factor in a population? We know that smoking may be more risky for some individuals than for others, depending on which suite of background factors they exemplify. Suppose a few individuals have their chances greatly augmented, most face only an intermediate increase in probability, while a few others have their chances increased little or not at all. How are these various

[6] More precisely, they are causally connect*able* in the sense allowed by relativity theory.

facts to be assembled into a single number representing *the* magnitude of the impact of smoking on heart attacks in the population as a whole?

A natural strategy is averaging, where the impact within a background context is weighted by that background context's probability of occurrence in the population. In the above example, the magnitude in the population should be intermediate, because the vulnerabilities weighted by their frequencies are symmetrically distributed around this mean. The proposal is this:

> If C is a causal factor for producing E in population P, then the magnitude of the causal factor is
> $\sum_{i=1} \{(\Pr(E/C \ \& \ B_i) - \Pr(E/\text{not-}C \ \& \ B_i)\} \times \Pr(B_i)$.

I adopt the average *difference* the causal factor makes in the probability of the effect, rather than the average *ratio*, for reasons that Salmon [1980] raised against Good's [1961-2] proposal. A factor that on average raises the probability from 0.25 to 0.75 should count as more powerful than one that on average increases the probability from 0.001 to 0.003. Notice that this measure will fall between 0 and 1 if the factor is positive and between -1 and 0 if the factor is negative.

An important feature of the probabilistic theory of property causality is that it is a three-place relation; the population is one of the *relata*. C may count as a positive causal factor for producing E in one population, but not in another; for example, C may be a positive causal factor in a subpopulation, but not in a more inclusive superpopulation. This raises an important question for the Connecting Princicple. If we are interested in the relationship between two token events, what population ought we to consider? If the question is whether Harry's smoking caused his heart attack, should we examine the causal role of smoking in his age group, his neighbourhood, among all U.S. adults, or what?

A principle of total evidence is appropriate here. One should circumscribe the population by using all the factors one knows to be true of Harry. If you know his age, weight, and dietary habits, then the relevant question is the degree to which smoking affects the probability of heart attacks among individuals of that sort. If, on the other hand, you know only that he is a U.S. adult, then that will be the relevant population.

A problem nevertheless remains. Suppose, of the traits one knows to be causally relevant to having a heart attack, one only knows Harry's situation with respect to some of them. For example, suppose exercise level is relevant, but we don't know Harry's. In assessing the magnitude of the impact of smoking on heart attacks, we must see how smoking affects the probability of coronaries within background contexts that mention different levels of exercise. What weight are we to give these different contexts—how are their probabilities to be ascertained (Sklar [1971])? An expedient would be to take the actual frequency of a background context as the best estimate of its probability. Whether this strategy is defensible depends on the connection between the population frequencies and Harry's chances of having the various traits in question.

What will this proposal imply about Harry's heart attack? Suppose we are interested in deciding whether the coronary was due to Harry's smoking or to Harry's high level of cholesterol consumption. It is known that Harry has both these traits; the question is whether we should trace back the coronary to one of them or to the other. Suppose, for simplicity, that these are the only two causal factors for coronaries. We need to compare the difference between the impact of smoking (S) on heart attacks (H), holding fixed the fact that Harry consumes lots of cholesterol (C), and the impact of cholesterol, holding fixed the fact that Harry is a smoker. That is, we need to discover whether this inequality is true:

$$\Pr(H/S \ \& \ C) - \Pr(H/\text{not-}S \ \& \ C) > \Pr(H/S \ \& \ C) - \Pr(H/S \ \& \ \text{not-}C)$$

This simplifies to the constraint that smoking is the more powerful causal factor if and only if

$$\Pr(H/S \ \& \ \text{not-}C) > \Pr(H/\text{not-}S \ \& \ C).$$

If this inequality is true, the Connecting Principle will imply that it is more plausible to think that Harry's coronary was token-caused by his smoking than that it was caused by his cholesterol consumption.

The Connecting Principle singles out as the most plausible token cause the causal factor that makes the largest (positive) difference in the probability of the effect; indeed, if all the

candidates for token causation are positive factors, they each will have the same absolute maximum; each raises the probability to the same maximum when it acts in conjunction with all the rest. The relevant question is what difference each characteristic makes—i.e., how much difference in the probability of the effect is implied by *not* having the trait present.

Of course, there is a third candidate that might be considered. Suppose one wishes to consider whether the conjunctive property of smoking *and* cholesterol token-caused Harry's heart attack. The conjunction of smoking and cholesterol may be a more powerful factor than is either smoking or cholesterol, when each is taken alone. The Connecting Principle then implies that the best supported hypothesis about token-causality is the conjunctive one that Harry's smoking and cholesterol intake jointly caused his coronary.

This is hardly surprising and may even be some sign that the Connecting Principle is on the right track. The problem of assessing the evidence for hypotheses of token causality is like any evidential problem; whether a hypothesis is the one best supported by the evidence depends on what the field of competing hypotheses is. It is not especially counterintuitive in the case in which conjunctive causes are among the alternatives that the best supported hypothesis may be that all the causal factors played a role. This is perfectly consistent with there being investigative contexts in which such conjunctions are not taken to be in the set of available alternatives. Examples of this latter type will be described in the next section.

Not only does applying the Connecting Principle depend on what the alternative hypotheses are taken to be; in addition, our formulation of the set of competitors involves presuppositions about the underlying causal processes. It is hardly inevitable that Harry's heart attack must trace back to either his smoking or his cholesterol level. If these two factors bring about heart attacks in quite different ways, then the discrimination problem may make sense. But if they work in basically the same way, the question may fail to be well formed. It may be like the famous nature/nurture confusion of asking whether someone's height is due to his genes or his environment, or of asking of a lit match whether its ignition was due to the presence of oxygen or to its dryness.

III

Phylogenies, Infections, and Rumours

Systematists attempt to reconstruct genealogical relationships among species from data concerning their characteristics. A set of taxa might be taken to be the terminal nodes of a branching phylogenetic tree; the question is how these taxa are to be grouped by postulating common ancestors. In its simplest form the question is posed by three taxa: Do any two share a common ancestor that neither shares with the third?

One aspect of this problem involves deciding whether a given species with a particular characteristic ought to be assigned an ancestor in the tree with that or another characteristic. In its simplest form, a character will come in two states, *P* and not-*P*. If a species has one of these, should it be assigned an ancestor that is the same or different?

Regardless of whether ancestor and descendant had the same or different characteristics, it will be true that the ancestor *produced*—was the token cause of—the descendent. And not only will it be true that the ancestor in some sense caused the descendent to exist; in addition, the descendent's character states trace back on a causal chain to the ancestor's even when ancestor and descendent exhibit different character states.

If ancestor and descendent are in the same character state, the descendent might have obtained its character state by unmodified descent. But there is no guarantee that this is true, since any even number of changes can lead ancestor and descendent to match. On the other hand, if ancestor and descendent do not match, then an odd number of changes must have occurred in the lineage linking them. Hence, assigning a *P* ancestor to a *P* descendent requires no evolutionary changes, but is consistent with the existence of any even number of such; on the other hand, assigning a not-*P* ancestor to a *P* descendent requires at least one evolutionary change, but is consistent with the existence of any odd number of such. The principle of cladistic parsimony asserts that one should minimize the number of required evolutionary transitions (Farris [1982], (Sober [1983b]) and so it favors assigning a *P* ancestor to a *P* descendent.

I have argued elsewhere that broad structural features of evolutionary processes imply that species character states are

positive causal factors for themselves, so to speak; whatever evolutionary forces impinge in a lineage, the probability of its ending in the *P* state is greater if it begins in that state than it would be if it began in the not-*P* state (Sober [1983b, 1984a, 1985]). This, I have argued, is part of the underlying rationale of cladistic parsimony. My point here is not to argue that evolutionary theory implies this claim about causal factors, but to show how its use can justify linking a descendent to one ancestor rather than to another.

The inference proceeds in accordance with the Connecting Principle. If one possible ancestor has character states that are positive with respect to a species' states, while another possible ancestor has character states that are negative with respect to the species' states, the evidence supports the claim that the species traces back to the first ancestor rather than to the second.

The spread of an infection through a population is a diffusion process that is structurally similar to the evolutionary spread of a novel characteristic through a population. I now sketch an epidemiological example that illustrates the Connecting Principle.

Introducing an infected individual into a population that has never before been exposed to the disease is a positive causal factor for the occurrence of that disease in others at a later date. Suppose two such infected individuals are introduced simultaneously. It is found that a third individual has come down with the disease a while later. The question is: to which of the two initially infected individuals does the third individual's disease trace back?

Assume that the presence of the two initial individuals is the only relevant factor to consider, so that we do not have to bring in further background contexts. Suppose further that the first individual was much more *contagious* than the second—i.e., an individual's probability of coming down with the disease, given contact with the first, exceeds an individual's probability of getting sick, given contact with the second. This means that the first individual's diseased state is a more powerful causal factor than the second's. If so, the Connecting Principle licences the conclusion that it is more plausible that the third individual contracted the disease from the first than that he did so from the second.

Rumours spread somewhat the way infections do (Cavalli-Sforza and Feldman [1981]). Reformulate the above example so that two individuals simultaneously discover or invent a new idea, which is found in others in the population some time later. From which of the inventors did the later individuals acquire their ideas? If the first is more apt to impart the discovery, given contact with him, then his ideas are more contagious than the second individual's. It is plausible to think that historians of cultural evolution use the Connecting Principle when they reconstruct lineages of intellectual influence.

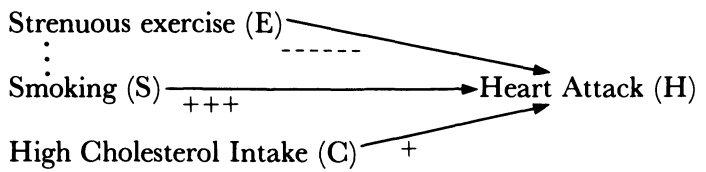
IV

Inferring Causal Factors Versus Inferring Causal Connections

The logic underlying the Connecting Principle implies that there is an important difference between the following two questions: Given that one or more individuals have traits that are causal factors in the production of an effect, which of them actually produced the effect? Given that an effect occurs, which positive causal factor is most plausibly thought to be present in its cause? In the first case, you infer a causal connection, given knowledge that various causal factors are present. In the second, you infer the presence of a causal factor, given knowledge that the effect occurred. All the earlier examples were of the first sort; the problem has consistently been one of postulating causal connections, not causally relevant events.

Let's go back to Harry to construct an example that illustrates this difference. Harry smoked and had a high cholesterol diet. I ask you to say which of these in fact produced his heart attack. This is a question of the first kind.

If smoking is a more powerful causal factor than cholesterol, the Connecting Principle instructs us to trace the coronary back to the smoking. But suppose almost everyone who smokes also exercises strenuously and that exercise is a powerful preventer of coronaries. As a result, the incidence of heart attacks among smokers is actually quite low—in fact even lower than the incidence among nonsmokers. Assume further that high cholesterol intake is not correlated with any preventer of coronaries. The facts about property causality are shown below:



Solid arrows indicate causal relevance; the number of plus or minus signs represents the magnitude of causal influence. A dotted line denotes strong positive correlation. Notice that it will be true in this circumstance that the incidence of heart attacks among those with high cholesterol intake exceeds the incidence of heart attacks among those who smoke.

If we know that Harry smokes and consumes lots of cholesterol, we can infer that he exercises. The Connecting Principle is not deterred by the correlation of smoking, but not cholesterol intake, with exercise. It quite sensibly reaches the conclusion that the smoking produced the coronary.

However, matters change if we do not know whether Harry exercised, and we wish to infer whether he smoked or consumed lots of cholesterol, based on the observation that he had a heart attack. In this case, the fact that smoking but not cholesterol intake is correlated with exercise would be relevant. We then would infer via an appeal to likelihood that the better supported hypothesis is that Harry consumed cholesterol, since this makes the observation (the heart attack) more probable.

We have here an instance of Simpson's paradox (Cartwright [1979], Sober [1984]). The probabilistic inequalities in the example are as follows:

(1) $\Pr(H/S \ \& \ C \ \& \ E) >>> \Pr(H/\text{not-}S \ \& \ C \ \& \ E)$.
(2) $\Pr(H/S \ \& \ C \ \& \ E) > \Pr(H/S \ \& \ \text{not-}C \ \& \ E)$.
(3) $\Pr(H/S) < \Pr(H/\text{not-}S)$.
(3) $\Pr(H/C) > \Pr(H/\text{not-}C)$.

The idea that smoking is a much more powerful cause of coronaries than is high cholesterol intake (among individuals who, like Harry, smoke, consume lots of cholesterol, and exercise strenuously) is reflected in the size of the inequalities stated in (1) and (2). However, if we wish to infer whether Harry was a smoker or a consumer of lots of cholesterol, we would

attend to (3) and (4), not to (1) and (2).[7] Although smoking is a more powerful cause of coronaries than is cholesterol intake, a coronary is better evidence for high cholesterol intake than it is for smoking.

V

Connecting Processes

Although the Connecting Principle was intially stated in terms of a primitive notion of support, its application to phylogenies, infections, and rumours involved understanding support in terms of likelihood. In each case, one hypothesis of token causality was favoured over another because the first conferred a higher probability on the observed effect.

In the phylogenetic example, a descendent (d) with character state P is traced back to an ancestor (a_1) that also has that character state, rather than to an ancestor (a_2) that lacks that character state. The likelihood rationale rested on the following inequality:

> Pr(d is P/a_1 is P and a_1 token-caused d) >
> Pr(d is P/a_2 is not-P and a_2 token-caused d).

Some features of the token causal relation involved here bear mentioning.

First, notice that the relationship described obtains between the objects (species) d, a_1, and a_2. This is somewhat contrary to the idea that token-causal relationships canonically obtain among events. Would it not be more apposite to formulate this relationship as 'a_1's having P token-caused d's having P?' Indeed, this is precisely what we did when we spoke of Harry's smoking as the token cause of his heart attack.

The reason this formulation must be avoided is that it collapses the likelihood differences. Recall that 'X token-caused Y' implies that X and Y both occurred. For this reason,

> Pr(d is P/a_1 is P and a_1's being P token-caused d's being P)
> =
> Pr(d is P/a_2 is not-P and a_2's being not-P token-caused d's being P) = 1.

[7] In particular, the question of whether Pr(H/S) > Pr(H/C) would be decisive.

The token causal connections under study must be conceptually distinct from the cause and effect events they link, at least if likelihood is to measure evidential support.

In the phylogenetic case, we already have a vocabulary that obeys this requirement. We say that one species *gave rise to* or *begat* another, and this leaves the characteristics of parent and offspring entirely unspecified. In the infection example, we may say that one individual *came into contact with* another (where different diseases require different modes of physical contact). In the case of rumour and intellectual influence, we speak of one person's *listening to* or *being a student of* another.

Likelihood is an epistemological principle. Yet, it enforces a requirement on how we formulate token causal hypotheses that has ontological significance. Connecting processes exist independently of the events they connect. Such processes are like channels in which information flows; the existence of the channel does not imply what information (if any) actually flows over it.

When one realizes that claims of property causality like 'smoking causes heart attacks' do not imply that anyone smokes or that anyone has a heart attack, it is natural to interpret them as meaning that smoking *can* cause heart attacks; claims of property causality describe the *potential* causal efficacy of properties in populations. However, this does not discriminate between the two concepts of cause, in that a similar conclusion can be drawn about claims of token causality, at least when they are formulated in the way just described. To say that one species gave rise to another implies that the first *can* transmit its characteristics to the second. Whether it does so is a further matter.

We can describe the potential causal bearing of one property on another, leaving open whether tokens of those types are actually connected by a process. Or we can describe the actual physical connection of two events, leaving it open which properties of the first were causally efficacious. Or we can, as it were, fuse these two descriptions together so that connecting process and connected properties are both implied. This is what we do when we say that Harry's smoking token-caused his heart attack.

For phylogenies, infections, and rumours, we isolated an

autonomous description of the connecting process, one which does not imply the states of the connected events. In Harry's case, however, it is not so clear how to do this. What this reveals, I would suggest, is ignorance on our part, not a special feature that distinguishes heart attacks from phylogenies, infections, and rumours. Token causality is a supervenient relationship, instantiated by different physical processes in different physical contexts. One species' generating another is not much like one person's talking to another, except that both are possible instantiations of the relationship of token causality.

Supervenient properties can be pressed into service when the underlying physical instantiation is unknown. This is what we may do in Harry's case. The state of Harry's circulatory system in some way traces back to his smoking (among other things). We sometimes use the word 'cause' when we are at a loss for further details.

Russell [1913] said that causality is a concept that a maturing science does without. In a not uncharacteristic use/mention elision, he also said that the word 'cause' never occurs in 'advanced sciences'. Unlike Russell, I think that causality is doubly central to science; property causality and token causality are both fundamental. But causality's supervenience explains why Russell may have been partly right about the word, if not the thing. The word *does* tend to vanish, as underlying physical processes are identified and given names of their own.

VI

Conclusion

Those of us who are skeptical of Hume's idea that there is nothing objective in causality beyond certain abstract relationships obtaining between the cause and its effect may nonetheless wonder about the epistemology of this recondite entity—the causal connection. Do we have independent epistemological access to causal connections? Are inferences about them separable from inferences about the events such connections are said to link? I think the above considerations suggest affirmative answers to these questions. The probabilistic theory of causality describes probability relations that obtain between causal factors and their effects. Such relationships provide evidence

for the existence of token causal connections, but neither of these concepts of cause can be reduced to the other. In addition, we have seen how inference to the existence of causal connections, mediated by the Connecting Principle, can be unimpeded by a phenomenon (Simpson's paradox) that has important consequences for how one infers the properties of a cause, given knowledge of the effect. This conceptual distinctness of causal processes from the events they connect is reinforced by the use of likelihood to confirm hypotheses of token causality. Although I am by no means confident that the Connecting Principle is ultimately correct (it may need to be hedged by more assumptions), it seems to be a useful point of departure into an important aspect of the epistemology of causality.[8]

[8] My thanks to Ellery Eells for useful discussion and to the National Science Foundation for financial support.

REFERENCES

Anscombe, G. [1971]: 'Causality and Determination', in E. Sosa (ed.), *Causation and Conditionals*, Oxford: Oxford University Press, 1975, 63-81.

Cartwright, N. [1979]: 'Causal Laws and Effective Strategies', *Nous* 13: 419-37.

Cavalli-Sforza, L. and Feldman, M. [1981]: *Cultrual Transmission and Evolution: A Quantitative Approach*, Princeton: Princeton University Press.

Edwards, A. W. F. [1972]: *Likelihood*, Cambridge: Cambridge University Press.

Eells, E. [1985]: 'Probabilistic Causal Interaction', *Philosophy of Science*, forthcoming.

Eells, E. and Sober, E. [1983]: 'Probabilistic Causality and the Question of Transitivity', *Philosophy of Science* 50: 35-57.

Farris, J. [1982]: 'The Logical Basis of Phylogenetic Analysis', in N. Platnick and V. Funk (eds.) *Advances in Cladistics*, vol. 2. New York: Columbia University Press. 7-36.

Good, I. J. [1961-2]: 'A Causal Calculus I and II', *British Journal for the Philosophy of Science* 11: 305-18; 12: 43-51; 13: 88.

Hacking, I. [1965]; *The Logic of Statistical Inference*, Cambridge: Cambridge University Press.

Hempel, C. G. [1965]: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.

Hesslow, G. [1976]: 'Two Notes on the Probabilistic Approach to Causality', *Philosophy of Science* 43: 290-92.

Otte, R. [1981]: 'A Critique of Suppes' Theory of Probabilistic Causality', *Synthese* 48: 167-81.

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

Russell, B. [1913]: 'On the Notion of Cause', *Proceedings of the Aristotelian Society* 13: 1-26.

Salmon, W. [1971]: 'Statistical Explanation', In W. Salmon (ed), *Statistical Explanation and Statistical Relevance*, Pittsburgh: University of Pittsburgh Press, 29-88.
Salmon, W. [1981]: 'Probabilistic Causality', *Pacific Philosophical Quarterly* 61: 59-74.
Salmon, W. [1984]: *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.
Scriven, M. [1959]: 'Explanation and Prediction in Evolutionary Theory', *Science* 130: 477-82.
Sklar, L. [1970]: 'Is Probability a Dispositional Property?', *Journal of Philosophy* 67: 355-66.
Skyrms, B. [1980]: *Causal Necessity*, New Haven: Yale University Press.
Sober, E. [1983a]: 'Equilibrium Explanation', *Philosophical Studies* 43: 201-210.
Sober, E. [1983b]: 'Parsimony in Systematics: Philosophical Issues', *Annual Review of Ecology and Systematics* 14: 335-58.
Sober, E. [1984a]: 'Common Cause Explanation', *Philosophy of Science* 51: 212-42.
Sober, E. [1984b]: *The Nature of Selection*, Cambridge: Bradford/MIT Press.
Sober, E. [1985]: 'Two Concepts of Cause', in. P. Asquith and P. Kitcher (eds.), *PSA 1984*. vol. 2, East Lansing, Michigan: The Philosophy of Science Association, forthcoming.
Suppes, P. [1970]: *A Probabilistic Theory of Causality*, Amsterdam: North Holland Publishing Co.
Van Fraassen, B. V. [1980]: *The Scientific Image*, Oxford: Oxford University Press.

CAUSAL FACTORS, CAUSAL INFERENCE, CAUSAL EXPLANATION

Elliott Sober and David Papineau

II—David Papineau

I

Two Types of Causation

Elliott Sober distinguishes two types of causation, which he calls property causation and token causation. Property causation is straightforwardly definable in terms of probabilities. But token causation is more problematic, not definable in terms of probabilities, though probabilities can be epistemologically relevant to judgements about token causation. I agree entirely that some such distinction is needed in the discussion of probabilistic causation. But I am not happy with the way Sober makes the distinction.

I shall not be able to deal with all the important issues Sober raises in his interesting paper. I shall concentrate on the distinction between two types of causation. In particular, I want to argue, against Sober, that both kinds of causation should be defined straightforwardly in terms of probabilities.

As it happens, I myself am suspicious of causes that only make their effects probable, and prefer the old-fashioned view that all causation is deterministic. But, apart from a few remarks at the end, I shall supress these suspicions here, and simply aim to show those of you who do believe in non-determining causes that Sober's path is unnecessarily circuitous.

At its simplest, my thesis is that token causation, as well as property causation, can be defined in terms of probabilities. But in fact my story is rather more complicated. The difficulty is that I don't really have a clear idea of what Sober means by 'token causation', and indeed am going to suggest that he is running some rather different ideas together. But, still, I want to argue that *however* we read the notion of token causation, there is no reason to think of token causes as irreducible to probabilities.

There is nothing unclear about Sober's notion of property causation. Within a given population, one factor S (smoking,

say) is a property cause of another factor H (heart attacks, say) if and only if S increases the chance of H for every other combination of relevant background factors.

But then, as Sober points out (p. 98), facts about property causation leave open the probabilistic structure of the single case. *Harry's* smoking can make *Harry's* heart attack highly probable, even if smoking is *not* a property cause of heart attacks: suppose that smoking increases the probability of heart attacks for people like Harry, but decreases it for other kinds of people.

I agree with Sober that in such a case it is natural to say that Harry's smoking *caused* Harry's heart attack, even though smoking is not a property cause of heart attacks. But this doesn't seem to me to be any argument at all for a non-probabilistic notion of 'token' causation. Why not simply insist that when we are interested in *single-case* causation, rather than in property causation in the overall population, it is the single-case probabilites alone that are relevant? That is, Harry's smoking caused Harry's heart attack just in case his smoking increased the chance of the heart attack given all the other relevant factors *then* present. Why ever should it matter to the efficacy of Harry's smoking that other sets of background conditions, present in other cases, but not in Harry's, allow smoking to reduce the chance of cancer? From now on I shall say that an instantiation of A is a *single-case cause* of an instantiation of B if the chance of B, given A and all the other relevant circumstances then present, was greater than the chance B would have had given those other circumstances alone.

Sober seems to feel (p. 103, especially) that even when we know all the probabilistic facts about the single case, there can still be a question as to whether it was Harry's smoking or, say, his high cholesterol level (C) that caused his heart attack. This I find puzzling. Suppose for simplicity, as Sober does on p. 103, that S and C are the only relevant factors, and that they both increase the chance of H in all circumstances. Then surely the situation is straightforward. If Harry smokes and has a heart attack, then his smoking caused the heart attack, for his smoking increased the chance of a heart attack in the circumstances. Similarly, if he has high cholesterol and suffers a heart attack, we should say the high cholesterol caused his heart attack. And if he *both* smokes and has high cholesterol, then we ought to say that *both* factors

(are part of what) caused H. Both factors were required for H to have the chance it did in the circumstances; if either had been absent H would have had a lesser chance.

Of course there may be pragmatic reasons for speaking of Harry's smoking, say, rather than his high cholesterol level, as 'the' cause of his heart attack. Perhaps your audience already knows about his high cholesterol level; perhaps you are particularly concerned to draw attention to the dangers of smoking. But this is a familiar point. There are pragmatic reasons for saying it was the short circuit, rather than the presence of oxygen, that was 'the' cause of the fire. But since Mill it has been clear that such pragmatic differences oughtn't to be elevated into a metaphysical distinction.

Here is a slightly awkward case. Suppose smoking on its own increases the chance of heart attack; that cholesterol on its own increases it to the same degree; and that smoking and cholesterol in combination still give it that same chance. (Prob(H/S & -C) = Prob(H/-S & C) = Prob(H/S & C) > Prob(H/-S & -C).) Suppose Harry smokes, and has a high cholesterol level, and suffers a heart attack. Given what I've said so far, his smoking wouldn't have (been part of what) caused his heart attack, because it didn't increase the chance, given the other factors then present. But the same goes for the cholesterol level. So we seem threatened with the undesirable conclusion that neither (was part of what) caused his heart attack.

But this isn't really a problem about single-case probabilistic causation as such. It's simply a particular instance of a more general problem: what do we say when something is overdetermined by two prior factors, each of which would have been sufficient by itself? The only difference here is that it's the chance of H that is overdetermined, not H itself. In general, overdetermination needs somehow to be dealt with as a special case, if we are to avoid the unwanted conclusion that nothing causes an overdetermined result. I don't see any reason why we shouldn't adopt the same strategy when thinking about probabilistic causes.

II

Mixers and Screeners-Off

Single-case probabilistic causation seems fairly unproblematic.

Yet Sober says that 'token causality is fascinating in its opacity' (p. 99); he prefers to avoid any explicit analysis of token causality, and restricts himself to suggestions about the epistemological connection between property causation and token causation. Why does Sober find token causality so opaque?

One reason is no doubt that his token causation isn't the same as my single-case causation: there are remarks at the end of his paper (pp. 110–11) which suggest that he would count my 'single-case' causation as a version of *property* causation, with 'token' causation as something else again. But I shall come back to this. For the moment I want to continue focussing on single-case causation. After all, whatever further rationale lies behind Sober's 'token causality', single-case causality seems perfectly adequate to answer the question with which Sober himself introduces the topic: 'What is it for Harry's smoking to cause him to have a heart attack?' (p. 98).

Moreover, much of what Sober says makes good sense if we read 'token causation' as 'single-case causation'. Indeed I would like to suggest that much of the plausibility of his 'Connecting Principle' derives from this reading of 'token' causation. For there is an important, and intuitively familiar, sense in which facts about *single-case* causation are indeed evidenced by, though conceptually independent of, facts about property causation (or, better, facts about 'population' causation—on which more below). This is to do with the familiar situation where we are ignorant of some of the background factors which are probabilistically relevant to some result. I want to spend some time on this, for it is of some independent interest, quite apart from its relevance to Sober's argument.

Suppose the overall statistics tell us that amongst humans heart attacks are more likely if you smoke: $\text{Prob}(H/S) > \text{Prob}(H)$. In such a case $\text{Prob}(H/S)$ will almost certainly be a 'mixed' probability, in the sense that various further factors X (metabolic factors, say) will further alter the probability of heart attacks: that is, $\text{Prob}(H/S \,\&\, X) \neq \text{Prob}(H/S \,\&\, -X)$, with $\text{Prob}(H/S)$ therefore being a weighted average of these further two probabilities. No doubt some of these further factors, some of these further ways of dividing up our reference class, will be known to us. But, still, even after we have divided up the reference class in all the ways we know to be relevant, the

conditional probabilities we are left with will still usually be mixed: even after we have taken into account the known X's, there will often be still further *un*known factors that make a difference to the difference that S makes to H. This is surely the situation with smoking and heart attacks: surely, in addition to the factors the medical researchers already know about, there are further metabolic, environmental, etc., factors they don't yet know about. (I'm not here making the false deterministic assumption that the only 'pure', unmixed probabilities are zero and one. I'm quite happy to believe that heart attacks are undetermined. My point is only that there are surely still some factors relevant to their indeterministic chances that we don't yet know about.)

However, even if medical (and agricultural, and sociological, and psychological) research generally leaves us with mixed probabilities, this doesn't mean that such research is necessarily unreliable as a guide to causes. For the inequality Prob(H/S) > Prob(H) is a good indication that S *sometimes* acts as *single-case* cause of H, even if Prob(H/S) is mixed. Think of it this way round. If S is ever a single-case cause of H, then there is a set of (possibly unknown) background factors in the presence of which S increases the chance of H. But Prob(H/S) is a weighted average of the difference S makes to H in the presence of all the different possible combinations of other relevant factors. So if Prob(H/S) > Prob(H), there must be at least one combination of other relevant factors where S really increases the chance of H.

There is of course a well-known flaw in this inference. The inequality Prob(H/S) > Prob(H) can be a 'spurious correlation': S never really increases the chance of H, but only seems to do so because it is itself correlated with the things that do. Smoking doesn't really affect your chance of a heart attack, it just seems to because anxiety, which really does cause heart attacks, makes you more likely to smoke.

A few sums will help here. Let X stand for some unknown further cause of H (like anxiety, for example). Now

(1) Prob(H/S) = Prob(X/S)Prob(H/S & X) + Prob(-X/S)Prob(H/S & -X).

That's the sense in which Prob(H/S) is a weighted average. And similarly,

(2) $\text{Prob}(H/-S) = \text{Prob}(X/-S)\text{Prob}(H/-S \,\&\, X) + \text{Prob}(-X/-S)\text{Prob}(H/-S \,\&\, -X)$.

But a comparison between (1) and (2) now makes it clear that Prob(H/S) can exceed Prob(H/-S) even though S itself never makes a difference to the chance of H: even though, as we say, X *screens* H *off* from S:

(3) $\text{Prob}(H/S \,\&\, X) = \text{Prob}(H/-S \,\&\, X) = \text{Prob}(H/X)$, and $\text{Prob}(H/S \,\&\, -X) = \text{Prob}(H/-S \,\&\, -X) = \text{Prob}(H/-X)$.

For even if these last two equalities hold, Prob(H/S) will exceed Prob(H/-S) if S is itself positively correlated with X, that is, if $\text{Prob}(X/S) > \text{Prob}(X/-S)$.

Some further explanation of these sums might help. It's important in this analysis that X is itself *positively* relevant to H. Compare (1) and (2) again. If X and S are positively correlated, then -X and S are negatively correlated, and indeed to just the same degree: $\text{Prob}(X/S) - \text{Prob}(X/-S) = \text{Prob}(-X/-S) - \text{Prob}(-X/S)$. But what now guarantees that Prob(H/S) will *exceed* Prob(H/-S) when we have the screening-off equalities (3)? Why can't the negative difference between the second pair of weighting factors, in (1) and (2) respectively, cancel out the positive difference between the first pair? But this is where the *positive* relevance of X to H comes in: it means that Prob(H/X) is itself bigger than Prob(H/-X), and thus (given (3)) that the figure which gets weighted by the first pair of weighting factors exceeds that weighted by the second pair.

Conversely, if X is itself *negatively* relevant to H, then (still assuming the screening-off equalities (3)) a positive association between S and X will mean that we end up with a 'spurious' *negative* correlation between S and H.

The point of all this has been to make it clear that not all mixed probabilities are spurious. In order for a mixed probability to be spurious, the X that does the mixing has to do something quite special. Not only must it somehow change the probability of H given S, but it has to change it so as to satisfy the equations in (3). By no means all 'mixers' do this: presumably it is in fact true that anxiety changes the probability of heart attacks given smoking, but false that amongst people who are anxious (and

amongst those who aren't) smoking makes no further difference to the chance of heart attacks.

Moreover, equations (1) and (2) make it clear that the only way a mixed probability can be spurious is for there to be a correlation between the putative cause S and some relevant mixer X. This point is of fundamental importance for all empirical research using statistics. As I suggested above, the probabilities uncovered in such research (medical, agricultural, etc.) are always likely to be mixed by as yet unknown factors. If mixing always carried with it the possibility of screening-off, then these research probabilities would be worthless as a basis for causal conclusions. But only mixers which are themselves associated with S can be screeners-off. So provided we can be sure that any unknown mixers aren't themselves associated with S, then the mixedness won't matter, for we can still securely infer that S does on occasion make a real difference to the chance of H.

This is the point of *randomized* experiments. If we are in a position to ensure that the subjects in some experiment (fields of wheat, say) are assigned to the putative cause (a fertilizer) at random, then we can be sure that the other factors (amount of sunlight, rainfall, and, no doubt, many other unknown causes) relevant to the result (crop yield) are themselves probabilistically independent of the putative cause: and this then means that any probabilistic connection between fertilizer and crop yield must be non-spurious, however mixed it might be. (Of course we might get a *sample* association which was misleading about real population probabilities. But that's a quite different point. Spuriousness, in the sense I have been using it, is nothing to do with misleading samples. Spurious probabilities are just as much characteristics of the underlying population as pure probabilities, or, for that matter, mixed probabilities. All that's spurious about them is their causal significance, not their probabilistic status.)

Randomization isn't always possible. We can't always use brute experimental force to ensure that any unknown mixers are probabilistically independent of the putative cause. In many cases it would be morally, or even legally, wrong for experimenters to give their 'subjects' the 'relevant treatment' at random. We would quite rightly object to a medical experiment which

divided a sample of people into two groups at random, and then constrained the first group to smoke and the second group not to. But even when we can't experiment, we can still *survey* in a way designed to ensure that the resulting statistics aren't spurious. What we need to do is consider what unknown mixers could possibly be associated with the putative cause, and bring them explicitly into the analysis. Thus in conducting a survey designed to discover whether smoking is a cause of heart attacks, we ought to check which people are anxious and which aren't, and see whether smoking still makes a difference within these subdivisions; but we needn't worry too much about factors like congenital heart abnormalities, say, for, although these are certainly relevant to heart attacks, it is difficult to see how they could be statistically associated with smoking. Provided we have taken explicit account of all the further causes that are associated with smoking, it doesn't matter if we ignore the rest, for, once more, further causes that are independent of smoking can't be responsible for a spurious correlation between smoking and heart attacks.

True, causal conclusions based on surveys will always be somewhat less secure than those based on randomized experiments. We might always overlook a possible confounding factor: it's just possible, after all, that the some of the causes of congenital heart defects are also causes of smoking, and that the correlation between smoking and heart attacks is entirely due to this. But there is surely room for informed judgement here. It would be unduly pessimistic to insist that we can never draw causal conclusions from non-experimental statistics unless we have taken *all* relevant factors into account (as opposed to all factors themselves associated with the putative cause). If such a high degree of security were really required, researchers wouldn't have been able to uncover many of the medical and social causes that we now know about.

So both randomized experiments and well-designed surveys can give us good grounds for judging that certain probabilities, though mixed, are not spurious; and in such cases we can conclude that S is indeed a cause of H. Let us now focus on this last notion of causation: in exactly what sense does a mixed unscreened-off correlation between S and H show us that S is 'a cause' of H? The first point to note is that this causal fact is a fact

about the population, rather than about any single case. We can infer that there is *some* set of background factors in conjunction with which S will increase the chance of H. That is, we can infer that S *can* cause H, that S *sometimes* acts so as to cause H—namely, when the requisite background factors are present and S is followed by H. I shall call this notion of causation—the kind of causation evidenced by non-spurious mixed probabilities —'population' causation. Note that, as I have introduced it, the notion of population causation depends explicitly on the notion of single-case causation: S is a population cause of H just in case S is capable of being a single-case cause of H (namely, when it is conjoined with the requisite background factors and followed by H).

My 'population' causation has some similarities to Sober's 'property' causation. Both require a partition of the reference class into all the different combinations of presence and absence of relevant background factors, and both then depend on what further difference S makes to the chance of H within such cells. But where Sober's property causation requires S to increase the chance of H within *all* such cells, my population causation only demands that S increase the chance of H within at least one such cell. (In other cells it can make no difference; it can even decrease the chance of H in some cells, provided only that the weighted average of the difference it makes over all cells remains positive.) Of course, this difference between Sober's 'property' and my 'population' causation is in the first instance just a matter of definition, not dispute. But note that, whatever the virtues of Sober's definition, we are going to need my notion of population causation anyway, in order to explain what is evidenced by the mixed but non-spurious correlations that empirical researchers characteristically go to such pains to establish. (Moreover, just this notion of 'population' causation is crucial to the logic of decision. Newcomb's paradox and related cases make it clear that spurious correlations are no good for acting on. But if we had to wait for pure probabilities before acting, nobody would ever have given up smoking to avoid illness. The correct strategy is to act on a probability precisely when you have reason to think that, even if mixed, it is not spurious. And this is because such probabilities evidence population causation: they show that S does sometimes make a

difference to the chance of H—and, in particular, they show that S makes a positive difference on weighted average over all the different types of unknown situation you might be in.)

The point I have been leading round to in this section is this. Suppose there is a mixed but unscreened-off correlation between S and H. Then S is a population cause of H. But this doesn't mean that a heart attack has to be caused by smoking in any particular area. Most obviously, Harry might have a heart attack even though he's not a smoker, and then his heart attack couldn't be due to his smoking. But even if Harry does smoke, and has a heart attack, his smoking needn't be a single-case cause of his heart attack. For the population connection between smoking and heart attacks only guarantees that S *sometimes* acts so as to cause H—it only guarantees that there is *some* set of background factors in conjunction with which S increases the chance of H. There may be other sets of background factors together with which S makes no difference to the chance of H, or even decreases that chance. So even though Harry smokes and has a heart attack, he may be the kind of person in whom smoking doesn't increase the chance of a heart attack, and then his smoking won't be a single case cause of his heart attack.

Now imagine that we know that both smoking and cholesterol are population causes of heart attacks, and that we know nothing about other population causes. As above, this can leave us in the dark about single cases. Harry can get a heart attack, but we mightn't know whether or not he smoked, or whether or not he had a high cholesterol level. More interestingly, even if we do know that Harry smoked, and that he had a high cholesterol level, we mightn't know which, if either, was a single-case cause of his heart attack. For Harry might have been the kind of person in whom smoking makes no difference to the chance of a heart attack; or he might have been the kind of person in whom cholesterol makes no difference.

What about Sober's suggestion that we should appeal to the principle of likelihood? Should we prefer that hypothesis about the single case that makes the observed result most probable? (I know Sober's concepts of causation are different from mine. But let me first consider the likelihood suggestion in my own terms. I shall return to Sober's concepts shortly.) Let us take the two

cases in turn. First, the case where we know nothing about Harry, except that he had a heart attack. Second, where we know not only that Harry had a heart attack, but also that he was a smoker and had a high cholesterol level. In both cases we want to know whether Harry's heart attack was caused by his smoking or by his high cholesterol.

In the first case, when we know nothing about Harry except that he had a heart attack, the likelihood principle seems clearly inadequate. This is not just for the reason that Sober gives: namely, that we could know that smoking was positively correlated with a negative cause of heart attacks like exercise (p. 107-8). That's a good enough reason, in its own way, but let's stick to the case where the only population causes we know anything about are smoking and high cholesterol. Even then the likelihood principle doesn't seem to work. Suppose that the probability of heart attacks in general is 10%, that the probability of heart attacks given high cholesterol is 20%, and that the probability of heart attacks given smoking is 90%. This might look like good grounds for thinking that somebody who gets a heart attack was likely to have been a smoker. But those probabilities are consistent with a situation where half the population has high cholesterol, but only one in a million is a smoker. And, if that were the case, it would clearly be misguided to judge that Harry must have been a smoker, just because he had a heart attack. The trouble here is that the probability of heart attacks given smoking indicates what you ought to believe about a heart attack *if* you know someone is a smoker, but abstracts from how likely it is that the person is a smoker in the first place. More generally, the trouble is that the likelihood principle aims to by-pass decisions about prior probabilities. Perhaps there are some inferential contexts where such decisions can sensibly be avoided. But the inference from Harry's heart attack to his smoking isn't one of them. (From a Bayesian point of view, if in addition to knowing that $\text{Prob}(H/S) = .9$, we know that $\text{Prob}(H) = .1$ and that $\text{Prob}(S) = 1/10^6$, then Harry's heart attack should make us believe that he smoked to the degree $9/10^6$.)

What about the second case, where we know that Harry both smoked and had high cholesterol, and want to know which of these his heart attack was actually due to? Here the likelihood principle might seem to make more sense. We don't have to

worry about prior probabilities, because we know that the causal factors are both present. Surely then the sensible thing is to assign the result to the cause that makes the most difference? But I don't accept the presupposition that the heart attack must be due *either* to the smoking *or* to the high cholesterol. As I have it, there are actually four possibilities in the single case: the background factors then present allow smoking to make a difference, but not high cholesterol; the background factors allow high cholesterol to make a difference, but not smoking; the background factors allow both to make a difference; the background factors allow neither to make a difference. All four possibilities are left open by the fact that smoking and high cholesterol are population causes of heart attacks. And I don't see how we can reach any conclusions about the single case without some further assumptions about the structure of background causes. Thus, for instance, if we believe that heart attacks are an indeterministic phenomenon, we will be likely to assume (though we won't have to) that both smoking and high cholesterol increase the chance of heart attacks for *all* sets of background conditions. And this will incline us to judge that Harry's smoking and his high cholesterol are *both* (part of) what caused his heart attack. On the other hand, if we think that heart attacks are always determined, then we will think that smoking only makes a difference given certain very specific background conditions (namely, given conditions together with which smoking determined a heart attack); and we will think the same about the conditions required for cholesterol to make a difference. And so in Harry's case we will probably think that his heart attack is due either to his smoking, or to his high cholesterol, but not to both.

My way of reading the question of whether Harry's heart attack is due to his smoking or to his high cholesterol is different from Sober's. From my point of view, this question only arises when we are *ignorant* of certain facts about the particular case. We know that there are certain background factors in conjunction with which smoking makes a difference, and we know that there are certain background factors in conjunction with which high cholesterol makes a difference; but we don't know what those background factors are, and so we don't know whether Harry has them or not. If we did know what they were,

and whether Harry had them, then we would know what difference Harry's smoking and his high cholesterol made to the chance of his heart attack in the actual circumstances, and then we would know whether his smoking, or his high cholesterol, or both, were single-case causes of his heart attack. Sober, however, thinks that there is a further fact of the matter about what 'token caused' Harry's heart attack, even after we know all the probabilistically relevant facts about the particular case.

It seems initially plausible to say that, even if smoking is a property cause of heart attacks, there is still a further question as to whether Harry's smoking caused Harry's heart attack. But I suspect that is because we read this as the question: were Harry's actual circumstances such as to allow his smoking to make a difference to his chance of a heart attack? This latter question is a good question, and one we might still want to ask even after we know that smoking is a population cause of heart attacks. But it's not Sober's question. For it's fully answered once we know all the probabilistic facts about the single case, whereas Sober's question is supposed to depend on some yet further fact of the matter.

III

Causal Sequences and Probabilities

Why might Sober suppose that there is a further fact of the matter? Here is one possible line of thought. When smoking does cause cancer, presumably it doesn't do so directly. Presumably there is some characteristic sequence of intermediate events, involving, let us say, nicotine in the bloodstream, spasm of the coronary arteries, etc., which occurs when smoking leads to heart attacks. And similarly with a high cholesterol level: presumably there is a characteristic sequence of, say, blood clots, coronary thrombosis, etc., involved when a high cholesterol level leads to a heart attack. But surely this now gives a good sense in which, even if both S and C increase the chance of H in the particular case, it will be either S, or C, but not both, that actually 'token causes' H. For presumably we will either have the first sequence, involving nicotine, or the second, involving a blood clot, but not both, in any particular case. And this will then decide whether it was Harry's smoking or his high cholesterol which caused his heart attack.

But this now changes the rules of the game. So far we have in effect been assuming that in each situation only two times matter—'earlier' and 'later'—and that the factors present at the earlier time (S, C, other relevant background factors) suffice to fix the chances of H at the later time. But this is of course a idealization. What happens in between 'earlier' and 'later' will in general make further difference to the chance of H.

Once we relax the idealization, and recognize that causation involves sequences of events, rather than simple earlier-later pairs of events, the analysis of probabilistic causation becomes much more complicated. But it is still possible. In particular, it is still possible to define single-case causation in terms of single-case probabilities. The trick is to characterize *causal sequences* themselves in terms of single-case probabilities. Once we have done this, we shall see that the probabilistic facts are still sufficient to determine whether or not Harry's smoking led to his heart attack. So bringing in causal sequences still won't give us 'token' causal facts that transcend the probabilistic facts.

At first sight it might seem that the existence of intermediate causes between S and H, like the nicotine in the bloodstream (N, henceforth), undermines the whole analysis of single-case causation. (Let's take the 'single-case' as read from now on, unless I specify otherwise.) I said originally that an instantiation of S is a cause of an instantiation of H if it increases the chance of H given all the other circumstances then present. But of course I meant 'all other circumstances' present at the 'earlier time'. If we include intermediate circumstances that come between S and H, like the presence or absence of N, then it's not clear how S can *ever* be a cause of H. In any particular smoker, either there will be nicotine in the blood stream or there won't be. If there is, then smoking won't make any further difference to the chance of a heart attack: the smoking will, so to speak, already have done its work. And if there isn't, then the smoking certainly won't increase the chance of heart attack. (I'm assuming for simplicity that the only route from S to H is through N, so that smokers who manage to keep N out of their bloodstream aren't at further risk of heart attacks from other effects of smoking.)

Intermediate causes screen prior causes off from their effects. In the presence of N, and in the absence of N, S makes no further

difference to the chance of H. So if causes have to make their effects more likely, given all the circumstances present in the particular case, nobody can ever be caused to have a heart attack by smoking.

What we need here is a distinction between direct and indirect causation. Suppose we define direct causes as causes that aren't screened off from their effects by anything: an instantiation of A is a direct cause of an instantiation of B just in case it increases the chance of B given *all* the factors present on that occasion. Then Harry's smoking isn't a direct cause of his heart attack, because it is rendered probabilistically irrelevant by the presence or absence of nicotine in his blood. But now we can define causation in general (direct or indirect) ancestrally: as a direct cause, or the direct cause of a cause. Then Harry's smoking can still be an (indirect) cause of his heart attack: namely, if it directly causes something which . . . which directly causes N which directly causes something . . . which directly causes H. In effect we have now characterized the notion of a *causal sequence*: a succession of events in which each one is a direct cause of the next. And so we can simply say that one event is a cause of another (direct or indirect) if there is a causal sequence leading from the first to the second.

What about the obvious objection—what if, as seems likely, all sequences of events are causally dense, in the sense that between any two events we can always find an intervening third which screens the first off from the second? Then there won't be any direct causes, and so, given the above definitions, no causes of any kind. Here we need to appeal to some kind of limiting process. Suppose, given some actual sequence of events, we start with a coarse approximation, which pictures it as a series of discrete events, separated by finite time intervals. Then relative to this approximation we can pick out (apparent) direct causes, and hence (apparent) causal sequences. Then we can take series of finer and finer approximations, and specify that X is a genuine causal sequence if and only if for every such series there is some approximation beyond which all further approximations, however fine, present X as an (apparent) causal sequence. If causal sequences are really dense, then at bottom there won't be any direct causes. But apparent direct causation in discrete approximations will do the necessary work. (In what follows I

shall revert to the convenient fiction that causal sequences are discrete.)

Here is a further difficulty. So far I have been assuming that, if two events are causally connected, the first is the cause and the second the effect. But it's not clear that the probabilities justify this. Probabilistic dependence is a symmetrical relation: if Prob(H/S) > Prob(H/-S), then so also is Prob(S/H) > Prob(S/-H), for, after all, both are equivalent to the manifestly symmetrical Prob(H & S) > Prob(H).Prob(S). Nor is the analysis of direct causation and causal sequences of much help here. I have been assuming that if A has a direct causal connection with some succeeding B, and B has a direct causal connection with a succeeding C, then A must cause B which causes C. But again it's not clear why the probabilities should justify this. The probabilities themselves (A screened off from C by B; nothing screening off A from B, or B from C) are just as consistent with B being the common cause of A and C (A←B→C), or indeed with C causing B which causes A (A←B←C), as with A causing B causing C (A→B→C).

Of course, if we build the direction of time into the direction of causation, then there is no problem: we can simply continue to assume, as I have been doing throughout, that, given two causally connected events, the earlier is the cause and the later the effect. But there are good reasons for not reducing the direction of causation to the direction of time (not least being the attractions of the converse reduction). It's relatively straightforward to explain directional causation in terms of probability and time; it would be nice to explain it without time, in terms of probabilities alone. But this is not the place to pursue this issue. (I think that the way forward lies in the fact that, while B's screening off A from C is consistent with B's being causally between A and C, and with B's being a common cause of A and C, it's not consistent with B's being a common effect of A and C. For more on this see Hausman (1984) and Papineau (1985).)

IV

Negative Causes

As I had it in the last section, each stage in a causal sequence needs to be positively relevant to the next. Sober disagrees. 'A gold ball rolling towards the cup is kicked by a squirrel, which

thereby lowers the ball's probability of going into the cup. Improbably enough, though, the ball ricochets here and there and then drops in. Here is a case in which the kick token-caused the ball to drop in, even though a kick of just that kind is not a positive causal factor for producing balls in cups' (p. 99).

If Sober is right here, then we have a strong argument for doubting that causal notions can always be defined in terms of probabilities. Whatever one thinks of the details, it at least makes initial sense to characterize causes as things that increase the chances of their effects. But if we have to allow that causes can sometimes *decrease* the chances of their effects too, then we will need some extra hold on cause-effect relationships: the idea of one thing *either* increasing *or* decreasing the chance of another is surely far too general to pick out causes.

Sober reinforces the point by observing (also p. 99) that in Cartwright's lucky plant example, by contrast with the squirrel example, we don't say that the spraying by a defoliant caused the plant to survive. But the lucky plant and the squirrel case have just the same probabilistic structure. So, if we agree with Sober's intuitions that their causal structure is different, the difference must lie in something other than the probabilities.

I agree that the defoliant didn't cause the plant to survive. But, according to my intuitions, the kick didn't cause the ball to drop in the hole either. After all, the incident with the squirrel scarcely helped the ball get into the cup. Sure, the ball ended up in the cup. But that was just a matter of luck, given the kick. It wasn't *because* of the kick.

More generally, my intuitions are against causes that reduce the chances of their effects. I don't think we ought to allow negative causes. On the other hand, as I shall explain in a moment, there are strong internal reasons within the theory of probabilistic causation for admitting negative causes. I think that this casts doubt on the theory of probabilistic causation. But first it will be helpful to get something else out of the way.

Wesley Salmon (1984) distinguishes between causal processes and pseudo-processes. Where causal processes can transmit marks, pseudo-processes cannot. Thus a moving shadow is a pseudo-process: you can't affect the later stages of a shadow by acting on its earlier stages. (Which is just as well, since moving shadows can travel faster than the speed of light.)

This is a useful and important distinction. The idea of transmitting marks might seem suspiciously anthropocentric. But this is not essential to the distinction. We can say that causal processes are those processes that characteristically carry causal sequences: the later features of a causal process are made more probable by earlier features. In pseudo-processes too there are correlations between earlier and later stages: but these correlations will be screened off by events not themselves part of those pseudo-processes. Paradigm cases of causal processes are the continued existence of a physical object, or the transmission of radiation; other examples would be one species giving rise to another, or the spread of germs through a population.

However, one can accept the idea of causal processes without concluding, as both Salmon and Sober seem to, that the later stages of such a process are *always* caused by the earlier ones. In particular, there is no need to accept this when a later stage of a causal process is made *less* probable by an earlier stage. Consider this case. People are causal processes. They characteristically carry many different causal sequences. Thus, for instance, whether you are a fat or a thin adult is largely dependent on whether you were a fat or a thin child. Suppose I was a fat child but am now a thin man. Should we say, just because I am a person, and people are causal processes, that my present thinness is the result of my childhood fatness? This seems silly. Childhood fatness increases the chance of adult fatness. If I am lucky enough to be thin, even though starting off fat, this is surely despite my initial fatness, not because of it.

Causal processes are certain kinds of space-time lines, certain kinds of sequences of space-time points. We pick them out because lines of those kinds tend to carry causal sequences, sequences of instantiations of properties such that the earlier instantiations make the later instantiations more probable. But when we actually get an improbable later *non*-instantiation, there is surely no need to say that *it* was caused by the earlier features which made it *im*probable.

My argument here presupposes that causation is a relationship between instantiations of properties, not between bare particulars. Causal processes are processes whose earlier properties characteristically cause their later properties. I don't see any point in saying that the earlier points in such a process cause the later

points, independently of whether the properties instantiated at those points cause each other.

Sober argues that token causation is independent of property instantiations: token-causal connections are 'like channels in which information flows; the existence of the channel does not imply what information (if any) actually flows over it' (p. 110). His argument here hinges on his appeal to the likelihood principle: we can't measure the support of the hypothesis that *Fa* token-caused *Gb* by the likelihood Prob(*Gb*/*Fa*, & *Fa* token-caused *Gb*), because '*Fa* token-caused *Gb*' implies that *Gb* actually occurred, so this probability will always be one; instead we have to use Prob(*Gb*/*Fa*, & *a* token-caused *b*). I don't really know what to make of this argument. When Sober first introduces his Connecting Principle, in section II, there is no mention of probabilities *conditional* on hypotheses about token causation. His suggestion is merely that we can measure the support of such token-causal hypotheses *by* (weighted averages of differences between) probabilities of the form Prob(E/C). There is no reference to token causes *inside* the probability functions. It is rather puzzling that they start appearing there in section IV. Token causality is supposed to be an opaque notion, and the point of the Connecting Principle is to give us some evidential hold on it. So the Connecting Principle ought to specify degrees of support for token-causal hypotheses in terms that do not themselves involve the notion of token causation. Which is what happens in section II: the degrees of support for token-causal hypotheses are to vary with the independently understood Prob(E/C)'s. But how are we supposed to understand the Prob(*Gb*/*Fa*, & *a* token caused *b*)'s of section IV? What values do these probabilities have? Are they just equal to Prob(*Gb*/*Fa*)? But then why not leave them as such? There seems no good reason to complicate the story, and therefore no good reason why we should think of token causation as a relationship between bare particulars.

I want to conclude by casting some doubt on the whole idea of probabilistic causation. Now we have separated out Salmon's notion of a causal *process*, there is surely nothing intuitively attractive about negative causes. Yet it is difficult for advocates of probabilistic causation to avoid them. Perhaps we would do better to avoid probabilistic causation instead.

Let me explain. Probabilistic causation is a response to indeterminism. Before twentieth-century physics overthrew determinism, modern philosophy took it for granted that causes had to determine their effects. If Harry's heart attack was caused by his smoking, then the latter determined the former. But quantum mechanics has made us accept that Harry can have a heart attack even though the prior circumstances didn't make this inevitable. So philosophers have responded by arguing that as long as Harry's smoking makes his heart attack more likely, then that's enough for the former to cause the latter.

But probabilistic causation isn't the only possible response to indeterminism. We could instead maintain the old view that causation is *per se* deterministic, and simply say that in so far as Harry's heart attack wasn't determined, it wasn't caused either. There would still be some causation in Harry's case. Harry's smoking would still cause the increased *chance* of a heart attack. Even if the heart attack isn't determined, we can think of this increased chance as itself an objective feature of the situation, and indeed as something determined by Harry's smoking.

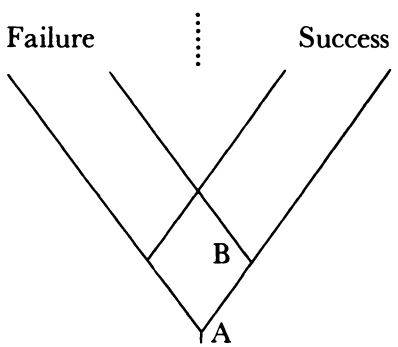
Perhaps this seems unnatural. After all, we're much more interested in actual results, like Harry's heart attack, than in their chances of occurring. So, apart from simple conservatism, what's the point of sticking to the old view that causation is deterministic? The only result seems to be that we are forced to stop talking about the causes of the things we are really interested in.

But it's not all plain sailing for probabilistic causation either. Recall the notion of a causal sequence discussed in the last section. If one thing causes another at a temporal distance, it will be via a sequence of linking events each of which directly causes the next. But if direct causes have to increase the probability of their effects, this is a very strong requirement. When one thing leads to another, there will often be *some* stage in the sequence which in the circumstances reduces the chance of the eventual result.

This difficulty has always been apparent to theorists of probabilistic causation. Suppes (1970) discusses a version of the squirrel case. But I would prefer to avoid the usual examples and instead construct an idealized situation whose probabilistic structure is quite unambiguous. (There is always a danger of

reading more into the squirrel case than you are supposed to, and assuming that given a fuller description of the circumstances the kick would make the sinking highly probable, or even inevitable. I suspect that this is why some people's intuitions are that the squirrel's kick does cause the ball to go into the hole.)

Imagine the following simple maze. You choose whether to go left or right three times in succession, the choice depending on the toss of an indeterministic coin. Success consists in ending at one of the two right termini, failure in ending at one of the two left termini. Suppose your actual path through the maze is as follows. First you go right (A); then you take a left turn (B); but then you take a right turn, and so succeed (S).



Presumably, on the probabilistic theory of causation, A caused the eventual success. After all, the chance of success was greatly increased by A: Prob(S/A) = 75%, Prob(S/-A) = 25%. But if A caused S, then presumably it did so via causing B, which then causes S. Now, there's no problem about A causing B: the chance of B given not-A is zero. But what about B causing S? B made S less likely. In the circumstances Prob(S/B) = 50%, yet Prob(S/-B) = 100%.

There are a number of ways to go here. One possible solution would be to accept that B doesn't cause S, but still insist that A does. But this would be a kind of action at a temporal distance. I take it that this is unacceptable. Another solution would be to accept that A doesn't cause S after all. But the initial attraction of the probabilistic theory was that it allowed us to talk about actual results, like the eventual success, as being caused even when they aren't determined. If we impose the strong requirement

that all intermediate causes have to make their successors more likely, its not clear that we will be left with very many temporally separated cause-effect pairs in a indeterministic world.

So the preferred solution for many advocates of probabilistic causality is to maintain that one thing (B) can cause another (S) even though it makes it *less* likely. This then allows them to have A causing S via a continuous sequence of intermediate causes. (This rationale for negative causes is particularly clear in Salmon, *op. cit.*, Ch. 7.)

But I still don't like negative causes. S doesn't happen because of B, any more than I'm now thin because I was once fat. S happens despite B. If the cost of the probabilistic theory is negative causes, surely we would do better to abandon probabilistic causation altogether.

If we stick to the old-fashioned deterministic view of causation we can describe the goings-on in the maze without getting into tangles. A causes S to have a 75% chance, and B to have a 50% chance. When B happens, S's chance is caused to go down to 50%. That's all the causing there is in the maze. Some other things happen. B happens. But B itself (as opposed to B's having a 50% chance) isn't caused by A. And S happens. But S itself (as opposed to S having a 75% chance, and then, later, a 50% chance) isn't caused either.

More generally, I would suggest that deterministic causation gives us a far simpler way of saying everything we need to say than probabilistic causation. But a full defence of this suggestion will have to wait for another time.

REFERENCES

Hausman, D. (1984), 'Causal Priority', *Nous*, 18.

Papineau, D. (1985), 'Causal Asymmetry', *British Journal for the Philosophy of Science*, 36.

Salmon, W. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.

Suppes, P. (1970), *A Probabilistic Theory of Causality*, North-Holland.