INDETERMINISM, COUNTERFACTUALS, AND CAUSATION*

RICHARD OTTE†

Cowell College
University of California, Santa Cruz

In this paper I wish to argue that counterfactual analyses of causation are inadequate. I believe the counterfactuals that are involved in counterfactual analyses of causation are often false, and thus the theories do not provide an adequate account of causation. This is demonstrated by the presentation of a counterexample to the counterfactual analyses of causation. I then present a unified theory of causation that is based upon probability and counterfactuals. This theory accounts for both deterministic and indeterministic causation, and is not subject to many of the traditional problems facing theories of causation.

Although counterfactual conditionals have always been important to philosophers, one area in which they have proved particularly useful has been in the formulation of a theory of causation. Counterfactual analyses of causation go back at least to Hume, but they have enjoyed a resurgence of popularity lately. I wish to argue that counterfactual analyses of causation are inadequate. I believe that the counterfactuals that are involved in counterfactual analyses of causation are often false, and thus the theories do not provide an adequate account of causation. I will show this by looking at counterfactual conditionals in indeterministic situations. Although counterfactual theories of causation have been limited to deterministic causation, I think that problems that arise with counterfactuals and indeterminism indicate that counterfactual analyses of causation cannot even account for deterministic causation. I will then present a theory that is based on probability and counterfactuals, and I will argue that it is able to account for both deterministic and indeterministic causation.

1. Counterfactuals. Although many theories of counterfactuals have been developed, one of the most popular has been the analysis in terms of possible worlds. In this paper the possible-worlds analysis of counterfactuals will be used as a heuristic device only, and none of the counter-

*Received June 1985; revised September 1985.

†An earlier version of part of this paper was read at the 1985 Pacific Division Meetings of the American Philosophical Association. I would like to thank Davis Baird and Wesley C. Salmon for helpful comments on that paper, and an anonymous referee for comments on an earlier version of this paper. This research was supported in part by Faculty Research funds and funds granted by the Regents of the University of California.

Philosophy of Science, 54 (1987) pp. 45–62.

examples or principles presented will depend upon this particular interpretation of counterfactuals. The basis of possible-worlds theories of counterfactuals is a relation of similarity that orders possible worlds according to how similar they are to a certain world. Although possible-worlds analyses of counterfactuals have been proposed by many philosophers, we will concentrate our attention on the theory of David Lewis (1973b) in this paper. Lewis claims that the counterfactual

$$\text{if } X \text{ were true, } Y \text{ would be true } (X \mathrel{\Box\!\!\rightarrow} Y) \tag{1}$$

is true if and only if either X is impossible, or if at all of the worlds most similar to this world in which X is true, Y is also true.[1] If we ignore the case where X is impossible, this is equivalent to saying that Y must be true in all of the worlds most similar to this world in which X is true. The intuitive idea behind this analysis is that in evaluating the counterfactual $X \mathrel{\Box\!\!\rightarrow} Y$, we want to know if Y would be true if the world were changed just enough to permit X to be true.

Another important type of counterfactual conditional is the "might be" conditional. We often say that if X were true, Y might be true ($X \mathrel{\Diamond\!\!\rightarrow} Y$). We then notice that "would be" conditionals and "might be" conditionals are interdefinable as follows:

$$X \mathrel{\Box\!\!\rightarrow} Y =_{df} \sim(X \mathrel{\Diamond\!\!\rightarrow} \sim Y) \tag{2}$$

$$X \mathrel{\Diamond\!\!\rightarrow} Y =_{df} \sim(X \mathrel{\Box\!\!\rightarrow} \sim Y). \tag{3}$$

In other words, if it is true that if X were true, Y would also be true, it must be false that if X were true, Y might be false. According to Lewis's theory, the counterfactual $X \mathrel{\Diamond\!\!\rightarrow} Y$ is true if and only if Y is true in at least one of the nearest worlds in which X is true. If Y is true in some of the nearest X worlds and $\sim Y$ is true in the other nearest X worlds, then both $X \mathrel{\Diamond\!\!\rightarrow} Y$ and $X \mathrel{\Diamond\!\!\rightarrow} \sim Y$ are true. "Might be" conditionals are particularly important when dealing with indeterministic worlds.

Suppose we have an indeterministic coin and are considering the counterfactual:

$$\text{If the coin were flipped, it would come up heads. } (F \mathrel{\Box\!\!\rightarrow} H) \tag{4}$$

Intuitively, I think that (4) is false. The reason for this is that it is not determined whether the coin would come up heads or tails. The coin

[1] This statement of Lewis's position is not quite accurate, because Lewis does not endorse the Limit Assumption. A precise and accurate statement of Lewis's analysis would not affect any of the arguments that follow, except to make them more difficult to state. Throughout this paper I will use this approximation of Lewis's analysis, although a precise statement of his analysis could be used in any of the arguments that will be presented.

might come up heads and it might come up tails. I propose that the following conditionals are both true:

If the coin were flipped, it might come up heads, $(F \diamondto H)$ (5)

and

If the coin were flipped, it might come up tails. $(F \diamondto T)$ (6)

But if (6) is true, it would seem that (4) is false. If it were true that the coin might come up tails, then it would be false that it would come up heads. Similarly, since we believe that (5) is true, we must also believe that (7) is false:

If the coin were flipped, it would come up tails. $(F \boxto T)$ (7)

These examples demonstrate that many "would be" counterfactuals are not true in indeterministic situations. In indeterministic worlds, it is often impossible to say what would happen if some counterfactual situation were to occur; but we can usually say what might happen if that situation were to occur. Thus "might be" counterfactuals play a central role in indeterministic situations.

2. Causation. Although many philosophers have proposed theories of causation that utilize counterfactuals, one of the most basic theories is that of Lewis (1973a). We will restrict ourselves to discussing Lewis's theory, although it will turn out that our criticisms of Lewis's theory will apply to other theories also. According to Lewis, if C and E both occur and if it is the case that if the cause had not occurred then the effect would not have occurred $(\sim C \boxto \sim E)$, then C is a cause of E. However, C can be a cause of E even when the above counterfactual is false. It is well known that counterfactuals are not transitive, and so it may be the case that it is false that $\sim C \boxto \sim E$, even though there is a causal chain linking C and E together, such that for each of them, if the cause had not occurred, the effect would not have occurred. So Lewis claims that C is a cause of E if and only if C and E occur and either $\sim C \boxto \sim E$ or there is a causal chain linking C and E together, such that each link in the chain satisfies the counterfactual requirement. For our purposes, we shall not be concerned with cases in which transitivity fails and we need to appeal to the existence of a causal chain between C and E. For the rest of this paper, we will be looking at the essence of Lewis's theory, which is that C is a cause of E if and only if C and E both occur and if C had not occurred then E would not have occurred. Other counterfactual analyses of causation generally accept this condition of Lewis's, but add other requirements to the analysis. Thus if we can show that Lewis's requirement is too narrow, then other theories will also be shown to be too narrow.

3. Criticism of Counterfactual Analyses of Causation. Let us first look at a case of indeterministic causation, which will illustrate why counterfactual analyses of causation do not apply in such cases. Suppose we have an atom in an excited state, and bombard it with a photon. In this situation, the atom can decay to the ground state and emit radiation. Physicists say that in this case the photon impinging on the atom caused it to emit radiation. It is important to realize that the atom could emit radiation without the photon impinging on it, and it does not have to emit radiation even when it is bombarded by the photon; the laws governing this reaction are strictly probabilistic. In this situation, it is false to say that if the photon had not impinged on the atom, it would not have emitted radiation. It might be the case that the atom would have spontaneously emitted radiation. According to the possible worlds analysis of counterfactuals, some of the nearest worlds in which the photon does not impinge on the atom are worlds in which radiation is emitted and some of them are worlds in which radiation is not emitted. Thus we see that according to the counterfactual analyses of causation, the photon impinging on the atom is not a cause of the atom emitting radiation. This may not be a defect in those theories, however, because they were developed to handle cases of deterministic causation, and this is clearly a case of indeterministic causation. It is helpful though to see why the counterfactual analyses of causation will not work for indeterministic causation.

I am skeptical about whether there even are any cases of deterministic causation, and am inclined to think that most, if not all, instances of causation are ultimately indeterministic in nature. I do not think that this idea has been appreciated by those who hold counterfactual analyses of causation. If the micro-world is really indeterministic, then we have to say that the macro-world is also. The indeterminism might be very small when we get to the macro-world, but it is still there nonetheless. We generally ignore the very low probability that certain improbable things will occur; this is justifiable in our normal lives, but any indeterminism is enough to cause problems for counterfactual analyses of causation. If there is just one world nearest to the actual one in which the cause does not occur and the effect occurs, we get the wrong result. Thus I feel that the applicability of counterfactual analyses of causation is extremely limited.

Although it certainly is inappropriate to require that the counterfactual analysis of causation handle cases of indeterministic causality, I think it is appropriate to require that it handle all cases of deterministic causality. Let us suppose that there actually are cases of deterministic causation, as well as cases of indeterministic causation. It seems reasonable to believe that the world contains both deterministic and indeterministic causal chains. It also seems reasonable to believe that some causal chains contain some links that are deterministic, and some links that are indeterministic. It might also be true that we have some causal chains that are made up of deterministic causes, and other chains that are made up of indeterministic causes. I see no strong reason to think that the world is made up of only one type of cause. It would be nice to have one unified conception of causation that would account for both deterministic and indeterministic causation; but if this is not possible, we would like our analyses of deterministic and indeterministic causation to mesh with one another nicely when we have a world that involves both types of causes. Thus we expect a correct analysis of deterministic causation to account for the deterministic causes, and a correct analysis of indeterministic causation to account for the indeterministic causes. Probabilistic theories of causation have been previously criticized for failing to account for cases that include both deterministic and indeterministic causes (Otte 1981). Counterfactual analyses of causation fail for a similar reason. Counterfactual analyses of causation cannot account for deterministic causes that occur in a world in which there are also indeterministic causes, or in worlds in which it is physically possible that some events occur by chance.

Suppose that in the actual world D is a deterministic cause of E. We can even say that all of the causes present in the actual world are deterministic causes. Suppose that there are also some worlds with the same laws as the actual ones in which S, which can be either a deterministic or an indeterministic cause of E, occurs spontaneously. In some of the nearest worlds to the actual world in which $\sim D$ occurs, S will spontaneously occur after $\sim D$ and cause E to occur. Thus there is a nearest $\sim D$ world in which E occurs, and so the counterfactual $\sim D \mathrel{\Box\!\!\rightarrow} \sim E$ is false. The reason that D is not considered to be a deterministic cause of E according to the counterfactual analysis is because some other event S may happen by chance which defeats $\sim D \mathrel{\Box\!\!\rightarrow} \sim E$; S may cause E to happen in the absence of D. Thus the counterfactual analysis fails to account for deterministic causation in a world in which it is physically possible that some events happen by chance; it fails even if all of the causes in that world are actually deterministic causes and nothing happens by chance in that world.

One might object to the above counterexample by claiming that it is not true that in some of the worlds most similar to ours in which $\sim D$ occurs that S occurs and causes D. It might be claimed that a world in which S does not occur is more similar to the actual world than to one in which S occurs. This receives some plausibility from the fact that S does not occur in the actual world. Lewis says that in choosing an adequate similarity relation, "[i]t is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails" (1979, p. 472). Thus it might be held that in all of the $\sim D$ worlds most similar to the actual world, $\sim S$ occurs, and so it is true that $\sim D \square\!\!\rightarrow \sim E$.

The problem with this objection is that it excludes the possibility that both $\sim D$ and S might occur; it seems true that if $\sim D$ occurs, S *might* occur. In order for this to be true, at least one of the $\sim D$ worlds most similar to ours must contain S. But then it is false that $\sim D \square\!\!\rightarrow \sim S$, which means that $\sim D \square\!\!\rightarrow \sim E$ is false also. Another way of seeing this is: since we know that if $\sim D$ were to occur, S might occur, and since S is a cause of E, we can conclude that if $\sim D$ were to occur, E might occur. Thus some of the $\sim D$ worlds most similar to the actual world contain S and E, and the above objection fails. We cannot claim that worlds in which $\sim D$ and $\sim S$ occur are more similar to the actual world than worlds in which $\sim D$ and S occur without denying that if $\sim D$ were to occur, S might occur. But it seems obviously true that if D had not occurred, S might have occurred; so we must also claim that if D had not occurred, E might have occurred, even though D is a deterministic cause of E. Counterfactual analyses of causation cannot account for many cases of deterministic causation in an indeterministic world.

Some philosophers may object to this counterexample by claiming that D is not a deterministic cause of E. It might be claimed that since the world we are considering has indeterministic laws, D cannot be a deterministic cause of E. Given the state of that world at the time D occurs, one cannot predict with certainty whether E will occur. Thus the occurrence of D cannot determine or affect the occurrence of E in the way in which a deterministic cause is required to. In response to this objection, I would claim that whether an event is a deterministic cause of another event is not dependent solely upon whether the state of the universe at the time the cause occurs determines the state of the universe when the effect occurs. Consider the following example. Suppose that we have a deterministic world in which D is a cause of E. If we were to change that world slightly, and allow that it is possible that S occur spontaneously and cause E, would we say that D is not a deterministic cause of E? I think not. In changing the world slightly to permit the possibility of S occurring, the tendency of D to produce E was not affected at all. D is still a deterministic cause of E, but now there are other events that make the whole situation indeterministic. The essential connection between D and E remains unchanged, and thus I think that D remains a deterministic cause of E. Hence it is legitimate to expect deterministic theories of causation to account for this example.

4. A Solution to the Problem. If one looks at standard cases of indeterministic causation, such as hadron-scattering reactions, one notices that there are certain well-determined probabilities for different outcomes of a causal interaction. What happens is that the collision of certain particles *determines* the probability that certain states of affairs will occur; usually these probabilities are not 0 or 1, and thus these are not cases of deterministic causation. But the probabilities of the outcomes are different than they would have been if the collision had not occurred. In these cases of indeterministic causation, the cause changes the probability of the effects occurring. The cause affects the occurrence of the effect by changing the probability that the effect will occur. However, I see no reason to think that this holds only for indeterministic causation. I do not think that indeterministic causation and deterministic causation are completely different from one another; I think they are both parts of a single concept of causation. This leads me to believe that this characteristic of indeterministic causation applies to deterministic causation also. In both indeterministic and deterministic causation, it seems that the cause affects the probability that the effect will occur. These observations will play an important role in our theory of causation.

Although we have seen that "would be" conditionals such as (4) are not true in indeterministic situations, some "would be" conditionals are true in indeterministic situations. For example,

If the coin were flipped, the probability of heads would be .5 (8)

appears to be true, assuming we have a fair coin. The reason (8) is true is that although it is not determined whether the coin would come up heads or tails, it is determined that the probability of heads would be .5. In this situation, it is false that if the coin were flipped, the probability of landing heads might not be .5. Although it is not true that if the coin were flipped, it would come up heads, it is true that if the coin were flipped, the probability of it coming up heads would be .5. Thus some "would be" counterfactuals, those that make claims about the probability of a certain event occurring, are often true in indeterministic situations. This suggests that we might be able to use those counterfactuals to analyze causation.

Let us consider the following principle, which is a combination of Lewis's theory and the above ideas:

C is a cause of E if C and E both occur and if the probability of E occurring is x, then if $\backsim C$ were to occur, the probability of E occurring would not be x. (9)

In this principle, and throughout this article, we will assume that causes temporally precede their effects. This principle tells us that C is a cause of E if the probability of E would have been different if C had not occurred. C affects the occurrence of E by modifying the probability with which E occurs. The intuitive idea behind this principle is that if the occurrence of C can affect or modify the probability of E occurring, then C must be a cause of E. I think that this principle solves many of the problems that arose for Lewis's original analysis of deterministic causation, as well as provides us with an analysis of indeterministic causation. Since the consequent of the counterfactual requirement is a probability statement, that counterfactual will be true in many indeterministic situations in which a counterfactual with E as a consequent will be false. Thus we can have a counterfactual analysis of causation that will work in both deterministic and indeterministic situations. It will account for both deterministic causes and indeterministic causes in worlds that contain both.

The above principle easily handles instances of indeterministic causation such as the example in which the photon impinging on the atom in the excited state causes the atom to decay to the ground state and emit radiation. We saw that it is false that if the photon did not impinge on the atom that the atom would not have decayed to the ground state; the atom might have decayed to the ground state and emitted radiation even if the photon did not impinge on it. However, the photon impinging on the atom certainly affects the probability of the atom decaying to the ground state. If the photon had not impinged on the atom, the probability of the atom decaying to the ground state and emitting radiation would have been different. Thus we see that according to the above principle, the photon impinging on the atom is a cause of the atom decaying to the ground state and emitting radiation, which is the correct analysis of this example. Similar considerations show that the principle can account for cases of deterministic causation.

5. Counterfactual Probability. Although I think principle (9) captures the intuitive idea behind a correct analysis of causation, there are ambiguities and problems with it that must be addressed before it will be an adequate analysis. The major problem with the principle is that it combines counterfactuals and probability. It is not clear what a counterfactual that has a probability statement for a consequent is expressing. In particular, what is the probability in the consequent conditional on? For example, one might claim that since E occurs, the probability of E is 1, and hence according to principle (9), C is a cause of E if $P(E \mid {\sim}C) < 1$.

In clarifying the above principle, there are two probability statements that must be made more precise. The first is the claim that the probability of E is x, and the second is the claim that if $\backsim C$ were to occur, then the probability of E would not be x. I propose that in the first of these probability statements, we should conditionalize on everything actual at the time of C; in other words, the entire state of the world at the time C occurs should be conditionalized on. Thus the probability of E will not necessarily be 1, unless that state of the world is sufficient for E occurring. The intuitive idea behind the second probability statement is that we are interested in the probability of E occurring if the world were changed just enough to permit $\backsim C$ to occur. We would like to keep the world the same as much as possible, and yet allow for all of the changes that would occur if $\backsim C$ were to occur. I propose that in the second probability statement, E should be conditional on all that is actual in the world at the time of C minus all that would be not actual if $\backsim C$ were to occur.

We can now use the above intuitions to give a more precise version of the above analysis of causation. The intuition behind the proposed analysis of causation can be expressed as follows:

> C is a cause of E if C and E both occur and if t is the time at which C occurs, then $P(E \mid \text{all that is actual at time } t) \neq P(E \mid \text{all that is actual at time } t \text{ minus all that would be not actual if } \backsim C \text{ were to occur then})$. (10)

This principle states precisely what information is to be conditionalized on. It retains the essential insight of the original principle, and avoids the problems that it faced. The counterexamples to the counterfactual analyses of causation are not counterexamples to principle (10). Counterfactual analyses of causation could not account for the example where D was a deterministic cause of E, and yet if D had not occurred, S might later have occurred spontaneously and caused E. Since S does not occur at time t, but might have occurred later, S is not conditionalized on in either of the probability statements. Thus the probability of E is lowered when we conditionalize on the state of the world without C, which is the correct analysis of this example. One might consider a slightly different example in which an event S' might spontaneously occur at time t and cause E to occur. Since S' does not occur at time t, it is not conditionalized on in the first probability statement. Since it is false that $\backsim S'$ would be not actual if $\backsim D$ were to occur, we do not remove $\backsim S'$ from what is conditionalized on in the second probability statement. In neither case is S' conditionalized on, and thus $\backsim D$ will lower the probability of E, which is the desired result. Thus the occurrence or nonoccurrence of D does affect the probability of E occurring in those two examples. Because principle (10) can handle these cases that counterfactual analyses of causation cannot handle, principle (10) is to be preferred to counterfactual analyses.

6. Positive and Negative Causes. A different sort of problem with principle (10) is that it does not distinguish between positive and negative causes. In theories of indeterministic causation, it is customary and helpful to differentiate between two types of indeterministic causes: those that promote the occurrence of the effect, and those that hinder the occurrence of the effect. We can define a positive cause as follows:

> C is a positive cause of E if C and E both occur and if t is the time at which C occurs, then $P(E \mid$ all that is actual at time $t)$ $> P(E \mid$ all that is actual at time t minus all that would be not actual if $\backsim C$ were to occur then). (11)

A positive cause raises the probability of its effect when the proper information is conditionalized on. We can also define a negative cause:

> C is a negative cause of E if C and E both occur and if t is the time at which C occurs, then $P(E \mid$ all that is actual at time $t)$ $< P(E \mid$ all that is actual at time t minus all that would be not actual if $\backsim C$ were to occur then). (12)

A negative cause lowers the probability of its effect when the proper information is conditionalized on. One advantage of these definitions over other definitions of positive and negative causes is that the above definitions explicitly state what information is to be conditionalized on. Positive and negative relevance is highly dependent upon what is conditionalized on, and it is crucial that what is conditionalized on be specified.

7. Spurious Causes. One problem probabilistic theories of causality face is how to distinguish between genuine indeterministic causes and spurious indeterministic causes. Philosophers have long known that one event can raise the probability of another event without being a cause of that later event. This is because the correlation between them may not be due to a direct causal relationship between them. For example, a falling barometer reading raises the probability of a storm occurring, and yet no one thinks that the falling barometer reading is a cause of the storm. Even though the failing barometer reading raises the probability of a storm occurring, it is not a genuine cause of the storm; it is only a spurious cause of the storm. The genuine cause of the storm would be the dropping atmospheric pressure, which would cause both the barometer reading to fall and the storm to occur. One of the most difficult problems for probabilistic theories of causation is to distinguish between genuine and spurious causal relations.

Reichenbach (1956), and later Suppes (1970), had the intuition that once it was known that the genuine cause occurred, knowledge of the occurrence of a later spurious cause would not affect the probability of the effect occurring. Given knowledge of the genuine cause, knowledge of spurious causes are irrelevant to predicting the occurrence of the effect. In the above example, if it is known that the atmospheric pressure is dropping, knowledge of the falling barometer reading does not help one to predict the occurrence of a storm. In formalizing this intuition, Reichenbach proposed to distinguish between genuine and spurious causes by means of the screening-off relation. The screening-off relation is defined as follows:

$$\text{Event } C \text{ screens off event } SC \text{ from event } E \text{ iff } P(E \mid SC \,\&\, C) = P(E \mid C). \quad (13)$$

Given this definition, we can express Reichenbach's intuition about spurious causes by saying that a spurious cause will always be screened off from the effect by an earlier genuine cause. Thus we can distinguish genuine causes of an effect from spurious causes by whether there is an earlier event that screens off the cause from the effect. If the cause is a genuine cause of the effect, there will be no earlier event that screens it off from the effect.

Although Reichenbach's criterion accounts for many cases of spurious causation, there is an important class of causal relations that it does not give an adequate account of. Salmon (1978, 1980, and 1984) has discussed two basic types of causal forks: conjunctive forks and interactive forks. Without going into the details of these causal relations, we can say that Reichenbach's analysis works for spurious causes that are part of a conjunctive fork, but it does not account for spurious causes that arise because of an interactive fork. In an interactive fork, the common cause produces both the effect and the spurious cause, and yet the common cause does not screen off the spurious cause from the effect. Thus the intuition that genuine causes can be distinguished from spurious causes by means of the screening-off relation is not correct when the causes are part of an interactive fork.

I believe that previous theories of indeterministic causation have been unable to account for the difference between genuine and spurious causes because they have not seriously considered what information should be conditionalized on. I think that principle (11) can handle all cases of spurious causation that deal solely with indeterministic causation. When we consider deterministic spurious causes, the problem becomes more complicated. In the following, I will also argue that my account of causation can handle any case of spurious causation that the counterfactual analyses of causation can handle. Thus problems concerning spurious causation do not provide a reason to prefer a counterfactual analysis of causation to the above account of causation.

Let us consider a simple case of indeterministic spurious causation to illustrate how principle (11) will deal with the difference between spurious and genuine causes. Let *CC* (common cause) be an indeterministic cause of both *SC* (spurious cause) and *E* (effect) as is diagramed in figure 1. It does not matter whether this is a conjunctive fork or an interactive fork: the analysis handles both types of forks in the same way. Suppose that *SC* occurs at time *t*. We are interested in the values of $P(E \mid \text{all that is actual at time } t)$ and $P(E \mid \text{all that is actual at time } t \text{ minus all that would be not actual if } \backsim SC \text{ were to occur then})$. One event that is actual at time *t* is an intermediary event *IE* that is in the causal chain between *CC* and *E* (see figure 2). Furthermore, it is not true that if $\backsim SC$ were actual, then *IE* would not occur. Thus *IE* is among the information that is conditionalized on in both of the probability statements in the above principle. When *IE* is among the information conditionalized on, then the occurrence of *SC* is irrelevant; *IE* screens off *SC* from *E*. Hence *SC* does not raise the probability of *E*, and according to principle (11), the spurious cause *SC* is not a genuine cause of *E*. The above analysis is able to distinguish adequately between genuine and spurious causes when the causes are indeterministic.[2]

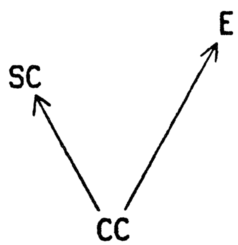

Figure 1

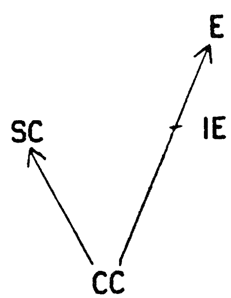

Figure 2

[2]Patrick Maher has suggested (in conversation) that the existence of an intermediary event earlier than *IE* may allow Suppes's theory to avoid the problem of interactive forks.

Let us now consider a case of spurious causation that involves deterministic causes. Although I think that principle (11) handles most cases of spurious causation involving deterministic causation in the same way in which it handled the indeterministic case, there is one example that is a problem for the above principle. Suppose that in the example diagramed by figure 2, *CC*'s occurring is sufficient for the occurrence of *SC* and necessary for the occurrence of *IE* and *E*. In order to determine what information to conditionalize on, we must determine what states of affairs must be deleted from all of the actual states of affairs at time *t*. This depends upon what states of affairs would be not actual if $\sim$*SC* were to occur. The crucial questions are whether *CC* would not occur if *SC* did not occur, and whether *IE* would not occur if *SC* did not occur.

According to Lewis (1979), on the standard way of interpreting counterfactuals, counterfactuals that say how the past would be different if a certain event were to occur now are usually false. If Lewis is correct, then it would be false that if $\sim$*SC* were to occur, then *CC* would not occur; and hence it would also appear to be false that if $\sim$*SC* were to occur, then *IE* would not occur. Principle (11) would then be able to handle this case, because *IE* would be conditionalized on in both probability statements in the proposed analysis. The occurrence or nonoccurrence of *SC* would not make any difference to the probability of *E* occurring.

However, many philosophers disagree with Lewis and think that some backwards-directed counterfactuals are true. It certainly is difficult to find true backwards-directed counterfactuals, but some philosophers claim that under certain circumstances backwards-directed counterfactuals are true. Pollock (1984) claims that a backwards-directed counterfactual is true if it is an instantiation of a law of nature. Thus if the counterfactual, if $\sim$*SC* were to occur, then $\sim$*CC* would occur, were an instantiation of a law of nature, then it would be a true counterfactual. However, even if it were also true that if $\sim$*CC* were to occur, then $\sim$*IE* would occur, one could not conclude that if $\sim$*SC* were to occur, then $\sim$*IE* would occur. That inference would require counterfactuals to be transitive, and it is well known that counterfactuals are not transitive.

Even though it does not logically follow that if $\sim$*SC* were to occur, $\sim$*IE* would occur, it does seem plausible that this is a true counterfactual in the situation we are considering. The fact that it is not derivable from other counterfactuals does not imply that it is not a true counterfactual. Perhaps an example could be constructed in which it seemed intuitively true that if $\sim$*SC* were to occur, then $\sim$*IE* would also occur. In an example in which this was true, principle (11) would give the wrong result and would claim that *SC* is a cause of *E*. Although principle (11) has problems with this example, so do counterfactual analyses of causation. Counterfactual analyses of causation cannot provide an analysis of examples of this sort that is better than the above analysis. Examples of this sort are extremely puzzling, and it is not clear that the inability of the above analysis to adequately handle them is a major defect in the analysis. Further investigation is required into whether examples of this sort exist, and into ways of dealing with them if they exist.

8. Simpson's Paradox. Simpson's paradox arises when an indeterministic cause is correlated with a third causal factor in such a way that the cause either lowers the probability of the effect or is probabilistically irrelevant to the effect. Consider the following fictitious example, in which eating egg yolks causes heart disease, and exercise prevents heart disease. People who exercise regularly may also eat egg yolks, because they think that the exercise will offset the effects of eating food high in cholesterol. In this situation, it may be the case that eating egg yolks actually lowers the probability of a person getting heart disease, because exercising regularly is better at preventing heart disease than eating egg yolks is at causing it. Even though eating egg yolks contributes to the occurrence of heart disease, it actually lowers the probability of heart disease. This is an example of Simpson's paradox.

In Otte (1985), I distinguished between three different forms of Simpson's paradox. These are diagramed in figures 3, 4, and 5. In those figures, C is the cause of E, and F is a third event that is correlated with C and E in such a way that the probabilistic relevance relations are reversed. A "+" or a "−" is placed by the causal connections to indicate whether the cause contributes to or hinders the effect. The above example concerning exercising regularly, eating egg yolks, and heart disease has the structure of figure 3. Exercising regularly (F) prevents heart disease (E) and contributes to the ingestion of egg yolks (C).

We could modify the above example, which would give it the structure of figure 4. Suppose that the reason that people who eat egg yolks (C) also exercise regularly (F) is that there is a common genetic characteristic

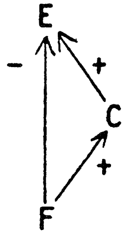

Figure 3

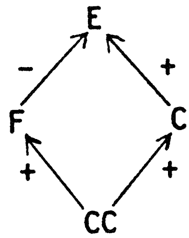

Figure 4

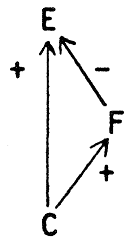

Figure 5

(*CC*) that contributes to both exercising regularly and eating egg yolks. In this situation, we could again have the eating of egg yolks lowering the probability of heart disease, even though it contributes to heart disease.

An example that has the structure of figure 5 would be if eating egg yolks (*C*) causes people to exercise regularly (*F*) because they worry about getting heart disease (*E*). In this situation, it would appear that eating egg yolks contributes both to the occurrence of heart disease, and to the non-occurrence or heart disease, even though eating egg yolks lowers the probability of heart disease.

Our analysis of causation easily handles cases of Simpson's paradox diagramed by figures 3 and 4, and a very minor addition will enable it to account for cases that have the structure of figure 5. In cases diagramed by figure 3, the probability of *E* will be conditional on some intermediate event *IE* which lies between *F* and *E* and is actual at the time *C* is actual (see figure 6). If *C* does not occur, the probability of *E* will certainly be lower than it would be if *C* occurred. Conditional on the event *IE*, *C* is positively relevant to *E*. Thus according to principle (11), *C* is a positive cause of *E*, even though $P(E \mid C) < P(E)$. Similar considerations apply to cases of Simpson's paradox that have the structure diagramed by figure 4. When the proper information is conditionalized on, *C* becomes a positive cause of *E*.

When we apply the above analysis to cases of Simpson's paradox that

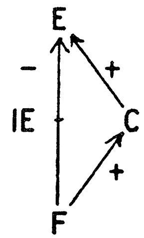

Figure 6

are diagramed by figure 5, we find that C is a negative cause of E. I believe that is a correct result, because C certainly does contribute to the prevention of E. Since C is a cause of F, which tends to prevent E, C also tends to prevent E. Our analysis of causation correctly claims that C is a negative cause of E. Any analysis of Simpson's paradox that has the result that C is not a negative cause of E is inadequate.

Although our analysis is partly correct, it does not appear to account for the fact that in addition to being a negative cause of E, C is also a positive cause of E in examples with this structure. One easy and natural way to get this result from the above analysis is to claim that positive causation is transitive: if X is a positive cause of Y, and Y is a positive cause of Z, then X is a positive cause of Z. Although this principle cannot be derived from our analysis of a positive cause, it can be added to it without inconsistency. Indeed, this requirement must be added in order for the analysis to account for many cases of deterministic causation, because it is widely recognized that deterministic causation is transitive. Lewis (1973a) realized that counterfactuals were not transitive and that causation is transitive. As we saw earlier, he solved this problem by adding a transitivity clause to his analysis of causation.

Our analysis of a positive cause gave only sufficient conditions for C being a positive cause of E; we will also claim that C is a positive cause of E if there is some intermediate event IE such that C is a positive cause of IE and IE is a positive cause of E. More formally, we add the following recursive requirement to our analysis:

If C is a positive cause of IE and IE is a positive cause of E, then C is a positive cause of E. (14)

This stipulation that positive causes are transitive enables the above analysis of causation also to handle cases of Simpson's paradox that have the structure diagramed by figure 5. In cases that have the structure of figure 5, there will be some intermediate event IE in the causal chain between C and E (see figure 7). When the proper information is conditionalized

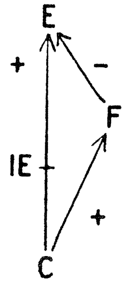

Figure 7

on, C will be a positive cause of IE, and IE will be a positive cause of E. Given the actual state of the world, C raises the probability of IE, and IE raises the probability of E. Since we claim that the relation of being a positive cause is transitive, C is a positive cause of E. It does not matter that C is also a negative cause of E. Upon our analysis, an event C can be both a positive and a negative cause of E. Most analyses of indeterministic causation do not allow for this possibility, and are unable to account for all cases of Simpson's paradox. Thus we see that the analysis of causation presented can account for our intuitions in cases involving Simpson's paradox.

9. Models. One objection that has been raised to the analysis presented here is that it seems to ignore the role that models play in scientific practice. Researchers normally use probability models to determine the probability of events and to determine which events are causes of other events. It might be claimed that an analysis of the sort given above is unnecessary because of the existence of probability models. In response to that objection, I would like to claim that both the above principles and probability models are important to science. We have proposed an analysis, which is very different from a model. An analysis of causation provides truth conditions; whereas it is seldom claimed that a model gives truth conditions or an analysis of causation. I think that an analysis of causation will tell us what information should be conditionalized on in a correct model. Neither the above analysis nor the existence of probability models implies that the other is unnecessary.

In this article we have examined a basic problem with counterfactual analyses of causation. We then investigated how this problem could be solved by a theory of causation that was based on both counterfactuals and probability. The analysis presented made explicit the role that probability and counterfactuals will have in a theory of causation. Theories that analyze causation only in terms of probability or only in terms of counterfactuals are missing an important part of causation. The theory presented here is able to account for many problems with deterministic and indeterministic causation that previous theories have been unable to account for. In addition, the theory presented is a unified theory of causation. It accounts for both deterministic and indeterministic causation in essentially the same way, which is to be preferred to having two separate theories that treat them as totally different types of causation.

REFERENCES

Lewis, D. (1973a), "Causation", *Journal of Philosophy, 70*: 556–567.
———. (1973b), *Counterfactuals.* Cambridge: Harvard University Press.
———. (1979), "Counterfactual Dependence and Time's Arrow", *Noûs 13*: 455–76.
Otte, R. (1981), "A Critique of Suppes' Theory of Probabilistic Causality", *Synthese 48*: 167–189.
———. (1985), "Probabilistic Causality and Simpson's Paradox," *Philosophy of Science 52*: 110–125.
Pollock, J. (1984), *The Foundations of Philosophical Semantics*. Princeton: Princeton University Press.
Reichenbach, H. (1956), *The Direction of Time*. Berkeley and Los Angeles: University of California Press.
Salmon, W. (1978), "Why ask, 'Why?'?", *Proceedings and Addresses of the American Philosophical Association 51*: 683–705.
———. (1980), "Probabilistic Causality", *Pacific Philosophical Quarterly 61*: 50–74.
———. (1984). *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton University Press.
Suppes, P. (1970), *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.