ON COUNTERFACTUAL PROBABILITIES AND CAUSATION: A NOTE

Nils-Eric SAHLIN

1. *Suppes' probabilistic theory of causality*

As an alternative to the variety of theories of causality which assume determinism, Suppes [10] suggests a probabilistic theory of causality. The basic idea behind the theory is that a cause increases the probability of its effect, an idea captured in the following definition:

D1 c is a *prima facie* cause of e if and only if

(i) c precedes e in time,
(ii) $P(c) \neq 0$ and
(iii) $P(e/c) > P(e)$. (1)

Condition (i) excludes such possibilities as backwards-causality, (ii) states that the cause is not impossible, and (iii) expresses the idea that the cause increases the probability of its effect.

It is easy enough to find examples of causal relations which satisfy conditions (i)-(iii) of *prima facie* cause. Such examples will not be examined here. Rather, an interesting counterexample to Suppes' theory will be considered, one which represents a crucial example for the theory to be presented here.

(1) P(.) is assumed to be an ordinary probability measure defined over some Boolean algebra or propositional field. It is assumed to be normalized, non-negative, and finitely additive. In a latter section I will assume some form of possible world semantics. P(.) will then be referred to as the probability measure defined over all possible worlds (cf. note 8). The probability of a proposition (set of possible worlds) equals the sum of the probability of those and only those worlds where the proposition holds. A finite set of possible worlds is not assumed (cf. note 6), thus allowing for non-standard measure theory to be used.

2. *A crucial example*

Such a crucial example is the 'contraceptive-pill' example given by Hesslow [2] and [3]. Physicians have argued that contraceptive pills can cause *t*hrombosis, that *p*regnancy can cause *t*hrombosis, and that pregnancy much more frequently causes thrombosis than contraceptive pills do. According to Suppes' theory, we should thus have $P(t / c) > P(t)$ and that $P(t / p) > P(t)$. However, since contraceptive pills lower the probability of pregnancy, it rather seems to be the case that $P(t / c) < P(t)$, contrary to what Suppes' theory states. Or at least this should be true for a population lacking other contraceptives.

What this example intends to reveal is that it is not always the case that a cause raises the probability of its effect. I doubt that the example can be satisfactorily handled within Suppes' theory, at least not within the realm of that theory in the form just described. (2) The aim of the present paper is to outline a probabilistic theory of causality based on counterfactual probabilites, a theory for which the example under consideration is no problem.

3. *Counterfactual probability and causality*

As Lewis [5] points out, Hume defines causation in two different ways when he states that

> '... we may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second.* Or in other words *where, if the first object had not been, the second never had existed*'. (3)

Lewis argues that we ought to develop Hume's second definition of causality, i.e. base a theory of causation on the statement that 'if the first object had not been, the second never had existed'. Such an approach suggests a counterfactual analysis of causation. However,

(2) For a different suggestion for solving the contraceptive pill example see Rosen [8]. Rosen's approach, which can be seen as a modification of Suppes' theory, is quite different from the one presented here.

(3) See Hume [4], p. 76.

to apply counterfactuals straight-forwardly, as Lewis does, leads to a deterministic theory of causality. Quantum mechanics tells us, on the other hand, that the world is indeterministic and that we thus need a probabilistic theory of causality. Hume, to be sure, did not know this, but had he known it would probably have said 'if the cause had not been, the effect would *probably* never have existed'. (4)

In what follows we need a connective, $\textcircled{\alpha}\!\!\rightarrow$, of counterfactual probability, with a numeral α written inside the circle. The sentence $C \textcircled{\alpha}\!\!\rightarrow E$ should be read 'if C had been the case, E would with probability greater than α have been the case', where C and E are the propositions expressing the truth of the events c and e, respectively. Assuming a possible-world semantics, this means that C and E are the propositions that hold at all and only those worlds where c and e occur.

The sentence $C \textcircled{\alpha}\!\!\rightarrow E$ is true if and only if in the set of worlds where C holds and most similar to the actual world (given some similarity relation) the probability of E is greater than the number denoted by the numeral α. Another way of expressing the same thing is to say that $C \textcircled{\alpha}\!\!\rightarrow E$ is true if and only if the counterfactual probability of E given C is greater than α, where the counterfactual probability of E given C is defined as the conditional probability of E given C obtained by considering a restricted set of possible worlds, i.e. those and only those worlds where C holds and which are most similar to the actual world. Let $P_C(E)$ stand for the counterfactual probability that E given that C. (5)

We are now in the position of being able to revise Suppes' probabilistic theory of causality. My suggestion is that we ought to

(4) It should be noted that the demand for some sort of probabilistic theory of causality requires no reference to physics. It is quite enough to consider an ordinary agent's state of belief.

(5) I have assumed that there always exists a set of closest worlds to the actual world and I thus avoid convergence problems. If there is no world where C holds, the counter-factual probability, $P_C(.)$, is undefined. It should be noted that $C \textcircled{\alpha}\!\!\rightarrow E$ is not identical with $C \,\Box\!\!\rightarrow P(E) = \alpha$, i.e. with a counterfactual with the overall probability of E as a probabilistic consequent. This latter counter-factual is read 'if C had been the case, $P(E) = \alpha$ would have been the case'.
For a discussion of counterfactual probabilities see Lewis [6] and for a discussion of the connection between $C \textcircled{\alpha}\!\!\rightarrow E$ and $C \,\Box\!\!\rightarrow E$ see Lewis [7].

have a counterfactual probabilistic theory of causality. The following definition is a first step towards such a theory.

D2 C is a *prima facie* cause of E if and only if

(I) C precedes E in time,
(II) $P(C) \neq 0$ and
(III) C ⓐ→ E, where $\alpha = P(E)$.

Conditions (I) and (II) are the counterparts of (i) and (ii) in *D1*. Condition (III) states that if the cause had been, the effect would with probability greater than α have been the case, where α equals the overall probability of E. This condition can also be stated in terms of counterfactual probabilities and we thus have (III′) $P_C(E) > P(E)$. (6)

Let us once again consider the crucial example and see how it can be handled by a counterfactual probabilistic theory of causality.

4. *The crucial example and the counterfactual probabilistic theory of causality*

The problem was that, although contraceptive pills can cause thrombosis, the conditional probability of thrombosis given that one takes contraceptive pills, is lower than the probability of thrombosis *per se*, contrary to what Suppes' theory implies. However, the counterfactual probability that T, given that C, is greater than the overall probability of T, since in those worlds in which women take contraceptive pills and which we regard as most similar to the actual world, the probability of thrombosis is greater than the overall

(6) Two things should be noted. Firstly, that if there is only finitely many worlds, linerarly ordered without ties by a similarity relation, then there is a unique closest C-world (if there is any C-world) and thus the counterfactual probability is either equal to zero or one. Further, C ⓐ→ E holds if and only if C □→ E holds. Therefore, either sets of finitely many worlds or an infinite set of worlds should be assumed. Secondly, $P_C(E)$ is, in general, not equal to $P(E/C)$. This is easily seen by drawing some sort of possible world diagram. For those interested in imaging as an alternative to conditionalization it should be pointed out that the counterfactual probability $P_C(E)$ is identical to the one we get by general imaging, i.e. where $P_C(E)$ is taken as the general image of P on C, given some similarity relation. See Gärdenfors [1] for a discussion of general imaging and for further references.

probability of thrombosis. Or, stated differently, C @→ T, where $\alpha = P(T)$, holds.

There is nothing obscure about this fact. Intuitively it can be seen that we are considering those and only those worlds that really matter, or, if you prefer, that we consider the conditional probability of T given C under quite a realistic belief revision. One such closest C-world, from our point of view, would be one in which women started to use contraceptive pills instead of some other contraceptives and where the other contraceptives do not cause thrombosis.

The type of worlds used to introduce the example, where the population lacks contraceptives, will not be considered here as one of the closest C-worlds. Put more generally, we consider populations of women who were given contraceptive pills and where each individual, for some reason or other, could not become pregnant.

I do not deny that this assumes a pragmatics of conditionals, nor that the pragmatic rules needed are far from established. However, the counterfactual probabilistic theory of causality outlined should benefit from its pragmatic content since the theory is thus in accordance with the way problems of causality are usually dealt with in science. ([7]), ([8])

([7]) There are several crucial examples which I will not be able to discuss within the realm of the present note, but let me briefly mention one problem. Assume that C causes first E and then F, but E does not cause F. This is the problem of epiphenomena. To handle this type of problem within Suppes' theory a 'screening off' method is needed. See, for example, Salmon [9]. I believe that a similar device is needed for a counterfactual probabilistic theory of causality, however, formulated in terms of counterfactual probabilities. But it should also be noted that some of the examples based on the problem of epiphenomena only seem to need a more realistic consideration of which worlds are closest to the actual world to avoid the problem. Such an approach should be compared with Lewis' [5] suggestion of rejecting the counterfactuals that cause the problem.

([8]) Counterfactual probability has been defined in terms of a probability measure P(.) and we thus have a problem of interpretation. My suggestion is that the measure P(.) should be interpreted as epistemic or subjective probability. Condition (II) would then demand that the cause is epistemically possible. However, the theory here outlined can preferably be based on some other interpretation of the measure P(.). P(.) can, for example, be read as objective chance. Given such a reading it might be possible to say by pure calculation of chance whether or not C is a cause of E. My choice of an epistemic measure is, of course, based on the fact that I believe that such a measure is needed and maybe the only reasonable one. However, we must leave these questions until a later date.

5. *Conclusions*

I believe that this paper shows that the contraceptive pill example is not a problem to a probabilistic theory of causality, if such a theory is based on counterfactual probabilities. It is also evident that a counterfactual probabilistic theory of causality can handle all the cases covered by Suppes' theory and must therefore be regarded as a better probabilistic theory of causality.

Department of Philosophy
University of Lund, Sweden
Nils-Eric SAHLIN

REFERENCES

[1] Gärdenfors P. Imaging and Conditionalization, forthcoming *Journal of Philosophy* 1982.
[2] Hesslow G. Two Notes on the Probabilistic Approach to Causality, *Philosophy of Science* 1976; 43: 290-292.
[3] Hesslow G. Causality and Determinism, *Philosophy of Science* 1981; 48: 591-605.
[4] Hume D. *Enquiries*, Clarendon Press, 1951.
[5] Lewis D. Causation, *Journal of Philosophy* 1973; 70: 556-567.
[6] Lewis D. Counterfactuals and Comparative Possibility, *Journal of Philosophical Logic* 1973; 2: 418-446.
[7] Lewis D. Causal Decision Theory, *Australasian Journal of Philosophy* 1981; 59: 5-30.
[8] Rosen D. In Defence of a Probabilistic Theory of Causality, *Philosophy of Science* 1978; 45: 604-613.
[9] Salmon W.S. *Statistical Explanation and Statistical Relevance*, University of Pittsburgh Press, 1971.
[10] Suppes P. *A Probabilistic Theory of Causality*, North-Holland, 1970.