NOÛS 30:4 (1996) 508–526

Causal Decision Theory and Decision-theoretic Causation

CHRISTOPHER READ HITCHCOCK
Rice University

1. Introduction

Probabilistic theories of causation and probabilistic theories of rational decision both face difficulties from spurious probabilistic correlations. Both types of theory handle these difficulties in the same manner: the spurious correlations are made to disappear by conditionalizing upon the elements of a carefully chosen partition. The structural similarity between the two types of theory suggests a systematic connection between them. One view—the view reflected in the name 'causal decision theory'—has it that the theory of causation is conceptually prior to that of decision: causal decision theory has the structure it does because it aims to tell us about the expected *effects* of our actions. But then we may ask from whence probabilistic theories of causation inherit their mathematical structure. In this paper, I will explore the prospects for a 'decision-theoretic causation' that explains the mathematical structure of a probabilistic theory of causation using a conceptually prior decision theory.

My approach has similarities to that of Hugh Mellor (1988).[1] Mellor argues, *pace* Salmon (1984) and others, that causes must *raise* the probabilities of their effects. He is unwilling to settle this dispute in what has come to be the normal manner: consulting intuitions on a case by case basis. Rather, he searches for a principled connection between causation and probabilities. Part of what it means to say that C causes E is that C is a potential means for accomplishing end E: if we want to achieve E, we may do so by bringing about C. According to standard decision theories, we should act so as to maximize expected utility. Suppose E is a desirable outcome: our utility for E is greater than our utility for $\sim E$. We have a

choice between two actions, one of which will realize C, and one of which will realize $\sim C$. What should we do? Barring complications, the act that brings about C will maximize expected utility when $P(E|C) > P(E|\sim C)$. Thus given our desired end E, C is a means to that end only if it raises the probability of E. So Mellor appeals to decision theory to defend a certain feature of probabilistic theories of causation: that causes must raise the probabilities of their effects. I will be concerned with a different feature of probabilistic theories of causation. According to most such theories, we do not assess the causal relevance of C for E by comparing probabilities conditional upon C and $\sim C$ *simpliciter;* rather, we must compare probabilities conditional upon the members of an elaborately constructed partition. Why is this so? Like Mellor, I will not accept the normal justification: that conditionalization upon members of the relevant partition reproduces our intuitions on a case by case basis. Instead, I will turn to decision theory to provide a deeper explanation.

This project falls within the tradition of Douglas Gasking (1955), who presented a 'manipulability' theory of causation, according to which C is a cause of E if C provides a "recipe" for achieving E (but not vice versa). As such it must face two of the standard objections that have been pressed against manipulability theories. The first is that they are too anthropocentric: they seem to have the consequence that there would be no causation if there had not been any rational agents, and that events beyond the scope of human intervention cannot be causes. The second objection is that manipulability theories are ultimately circular, since the relation of means to end is itself causal. In particular, there is some reason to think that the most reliable decision theory has causal concepts built into it. If so, then the resulting decision-theoretic account of causation runs the risk of circularity.

In the following section I present a probabilistic theory of causation. According to this theory, the causal relevance of C for E does not depend upon the simple conditional probabilities $P(E|C)$ and $P(E|\sim C)$, but rather upon conditional probabilities within an elaborately constructed partition. It is this feature of the theory that we wish to understand. In section 3, I present a version of causal decision theory. It turns out that the algorithm for computing causal expected utility involves conditionalization on the same sort of partition that figured in the probabilistic theory of causation. The trick is then to show that causal decision theory can be articulated and motivated without recourse to probabilistic causation—that is the burden of section 4. Thus we may, without circularity, appeal to causal decision theory to justify the mathematical structure of our probabilistic theory of causation. Section 5 draws some connections with controlled experiments, and section 6 summarizes my conclusions.

2. Probabilistic Causation

I will present a probabilistic theory of causation based on that of Ellery Eells (1991). In Eells' theory, the causal relevance of C for E depends upon conditional probabilities within cells of a partition $\{B_1, B_2, \ldots\}$, which we will call the *c-partition* for C and E. Each cell B_i corresponds to a causally homogeneous background context. C is a (positive) cause of E if and only if $P(E|C\wedge B_i) > P(E|{\sim}C\wedge B_i)$ for each cell B_i; C is a negative cause of E if and only if $P(E|C\wedge B_i) < P(E|{\sim}C\wedge B_i)$ for each cell B_i; and C is a mixed cause of E if one of the inequalities $P(E|C\wedge B_i) > P(E|{\sim}C\wedge B_i)$ or $P(E|C\wedge B_i) < P(E|{\sim}C\wedge B_i)$ holds for some cell B_i, but not for all cells. C is *causally relevant* for E if it is one of these types of causes.[2] The probabilities are to be understood as objective, although we will not broach the question of whether they are propensities, limiting relative frequencies, or what have you.

Let us now look a little more closely at how and why the partition $\{B_1, B_2, \ldots\}$ is constructed. The probability $P(E|C)$ may be high or low as a result of *spurious* correlations. Suppose, for example, that C is coughing and L lung cancer; then $P(L|C) > P(L|{\sim}C)$. Coughing raises the probability of lung cancer because it is correlated with a genuine cause of lung cancer: smoking. Such spurious correlations are one of the central difficulties facing a probabilistic theory of causation. In order to avoid the erroneous conclusion that coughing causes lung cancer, we evaluate the conditional probabilities of lung cancer while 'holding fixed', or conditionalizing on, smoking and not smoking. If $P(L|CS) = P(L|{\sim}CS)$, and $P(L|C{\sim}S) = P(L|{\sim}C{\sim}S)$—if smoking and not smoking 'screen off' coughing from lung cancer—this would suggest that coughing is not a cause of lung cancer. The suggestion toward which this is heading is that in order for C to be a cause of E, it must raise the conditional probability of E while holding fixed all other causes of C and E.[3] (Where 'causes' is meant to include any causally relevant factors, not only positive causes.)

This suggestion is not quite right. Tar in the lungs is a cause of lung cancer, and let us make the assumption that smoking causes lung cancer exclusively by causing the presence of tar in the lungs. Then the presence of tar in the lungs will screen off smoking from lung cancer, i.e. $P(L|ST) = P(L|{\sim}ST)$ and $P(L|S{\sim}T) = P(L|{\sim}S{\sim}T)$. We should not conclude from this that the correlation between smoking and lung cancer is spurious, as we did in the case of coughing and lung cancer: smoking causes lung cancer *by* causing the presence of tar in the lungs. The moral is that when comparing the probabilities of E conditional upon C and upon ${\sim}C$, we do not want to hold fixed *all* of the causes of E, but all of those causes that are independent of C (i.e. those for which C is not causally relevant).[4]

Let $\{F_1, F_2, \ldots\}$ be the set of all factors that are to be held fixed

according to these criteria. Then the partition $\{B_1, B_2, \ldots\}$ that is relevant to the evaluation of the causal relevance of C for E is the coarsest partition such that for any i and any j, either $B_i \subseteq F_j$ or $B_i \cap F_j = \phi$. That is, B_i may not overlap with both F_i and $\sim F_i$; each B_i holds fixed each F_i either positively or negatively.

I have argued elsewhere (Hitchcock 1993) that this theory is best seen as a special case where the cause and effect variables are both binary: the possible causes are C and $\sim C$; the possible effects are E and $\sim E$. More generally, causal claims describe the functional dependence of one variable upon another (relative to cells of a partition)—that is, causal claims describe *conditional distribution functions*. These functions can be quite complex: Eells' trichotomy of positive, negative, and mixed causation will have to be replaced by a much richer taxonomy. The language of causation provides a resource for providing partial and qualitative descriptions of these conditional distribution functions.

One prominent feature of this construction is that it involves causal concepts. The factors that are to be held fixed are characterized in terms of their *causal* relations. As a result, the theory sketched will not provide a reductive analysis of causation in terms of probabilities. Instead, the theory describes the relationship between probabilities and causation: a specification of the probabilities of various factors restricts what the causal relations between those factors can be, and vice versa. While some writers—notably David Papineau (1989, 1993)—have argued that it is possible to provide a suitable theory of probabilistic causation that avoids reference to causal relations, and thus to provide a reductive analysis, most are not so sanguine.[5] I have elsewhere (1993) defended a compromise view. A probabilistic theory of causation need not presuppose (as Eells' does) the various relations causal relevance, such positive, negative, or mixed causation (in the binary case). Instead it suffices to treat *causal relevance* (which was defined above, in the binary case, as the disjunction of positive, negative, and mixed causation) as a primitive, and then use probabilities to provide non-circular definitions of the various *types* of causal relevance in terms of this primitive relation.

The moral is that we must remain modest in our expectations of a probabilistic theory of causation. It can provide a reductive analysis of *something,* but not of causation *in toto.* The concept of causation has too many dimensions for any one theory to adequately characterize it—Skyrms (1984a) has presented a compelling argument for this conclusion. What the probabilistic theory of causation is able to provide is a *taxonomy* of causal relevance relations. We say that smoking *promotes* heart attack and that regular exercise *prevents* it. Promotion and prevention are both causal concepts, but the distinction between them is important. The probabilistic theory of causation may be unable to provide a reductive analysis of the

core causal component that they share, but it is well suited to providing an analysis of the distinction between them.

So let us now raise the question: why is the type of causal relevance that one variable has for another (or, in the binary case, that one factor has for another) determined by probabilities *conditional upon members of the relevant c-partition*? At first glance, this question does not seem to be very pressing: don't the various examples described above provide reason enough? But the query can become worrisome if we inquire into the purpose of making causal claims. Eells (1991, 94–97) has defended his formulation of the theory of probabilistic causation against various rivals by invoking considerations of *expressive power:* on his theory, one can use causal language to describe the probabilistic facts with greater precision than would be possible on rival theories. This line of defense only makes sense, however, if the *point* of making causal claims is to describe the structure of the underlying probability relations. I have argued (1993) that by adopting this view, we can account for the defectiveness of certain causal claims involving disjunctive causes. These considerations lend strong support to the following thesis about the purpose of causal claims: causal claims, insofar as they ascribe types of causal relevance such as promotion and prevention, provide qualitative information about conditional distribution functions within a probability space that adequately represents some aspect of the world.

Within this picture of probabilistic causation, the question that we have posed gains urgency. It is not hard to understand why we might be interested in conveying qualitative information about a probability space that adequately represents some aspect of the world, but why should we be so interested in the conditional distributions that obtain *relative to the c-partition* $\{B_1, B_2, \ldots\}$, which is so baroque in its construction? It will not do to answer that this is the partition relative to which conditional distribution functions give us information about *causal relations,* for we are viewing talk of causal relations as a way of communicating qualitative information about probabilities. Compare: we have expressions like 'tall' and 'short' for conveying qualitative information about the heights of things, but why are we so interested in talking about the heights of things? It is no answer to say that we are interested in the heights of things because it is these heights that determine whether things are tall or short.

Our question will be answered with the aid of decision theory: *c*-partitions are salient because analogous partitions play a prominent role in our deliberations as agents in the world. We are now in a position to respond to the charge of anthropocentrism. The probabilistic relations that constitute distinct types of causal relevance are objective features of the world, independent of the existence of rational agents. (Or better: it is an objective, agent-independent matter that features of the world can be ade-

quately represented by these probabilistic relations.) However, many noncausal probabilistic relations also have an objective existence. Our role as agents is not to *project* causal relations onto the world, but rather to *select* these relations, among all of the probabilistic relations that exist in the world, as being of particular interest to us.

3. Evidential and Causal Decision Theory

Suppose that an agent is contemplating performing one of the (mutually incompatible) actions A_1, A_2, . . . , which have as possible outcomes O_1, O_2, . . . The outcomes are specified in sufficient detail to incorporate all differences that the agent cares about. The agent has a subjective probability function, representing her degrees of belief, defined over an algebra of propositions that includes A_1, A_2, . . . , and O_1, O_2, . . . The agent also assigns utilities to each of the outcomes O_1, O_2, . . . , representing the relative desirability of those outcomes. Then, according to a standard version of decision theory, as formulated by Jeffrey (1983), the desirability of an act A_j is its *evidential expected utility:*

$$EEU(A_j) = \Sigma_i P(O_i|A_j) \times U(O_i).$$

The expected utility is 'evidential' because the conditional probability $P(O_i|A_j)$ reflects the extent to which the performance of act A_j provides evidence that outcome O_i will obtain. The rational choice for the agent is then to perform that act that has the highest evidential expected utility: that is the recommendation of 'evidential' decision theory.

Evidential decision theory is widely believed to run into difficulties with Newcomb problems — versions of a problem initially posed by William Newcomb and presented in Nozick (1969). Here is a simple example. Suppose that Fred periodically suffers from a certain vitamin deficiency. This deficiency has two effects: it inclines him to eat pickles, which he enjoys doing in any event, and it gives him severe muscle cramps whose pain far outweighs the simple joy of eating pickles. As a result of their common origin, Fred's pickle-eating and his muscle cramps are statistically correlated. Fred knows all this about himself. Let M stand for muscle cramp, E stand for eating a pickle, and V for vitamin deficiency. Then Fred's subjective probabilities might be:

$P(E|V) = .4$
$P(E|\sim V) = .25$
$P(M|V) = .75 = P(M|EV) = P(M|\sim EV)$
$P(M|\sim V) = .12 = P(M|E\sim V) = P(M|\sim E\sim V)$
$P(V) = .333 \ldots$

Together, these probabilities entail that:

$P(M|E) = .4$
$P(M|{\sim}E) = .3.$

His utilities are:

$U(ME) = 10$
$U(M{\sim}E) = 0$
$U({\sim}ME) = 1010$
$U({\sim}M{\sim}E) = 1000.$

He is now faced with a choice of whether or not to eat a pickle. Calculating the evidential expected utility of the two acts using the conditional probabilities $P(M|E)$ and $P(M|{\sim}E)$ yields $EEU(E) = 610$ and $EEU({\sim}E) = 700$. Eating the pickle has a lower evidential expected utility because of its correlation with muscle cramps. Intuitively, however, it seems that Fred's decision not to eat a pickle is irrational. Eating a pickle in this context is a harmless pleasure. Fred either has the vitamin deficiency right now or not—he has no control over that. Either way, his eating a pickle will not cause him to suffer from muscle cramps.

Newcomb-type problems arise because of spurious correlations: Fred's eating a pickle is correlated with muscle cramps even though eating a pickle does not cause muscle ache. This difficulty is very similar to the *prima facie* difficulty that spurious correlations pose for probabilistic theories of causation: recall the example involving coughing and lung cancer. Thus we may expect to find that the proper solution to Newcomb problems is along the same lines as the solution to the problem of spurious correlations in probabilistic theories of causation. Causal decision theorists advocate precisely this sort of solution: one should evaluate an act by computing expected utility within each cell of a partition, just as the theory of probabilistic causation evaluates the causal relevance of a factor in terms of probabilistic dependencies within each cell of a partition.

The causal decision theorist would advise Fred to compute the expected utility of his eating a pickle as follows: He should first assume that he does have the vitamin deficiency, and calculate the expected utility of eating a pickle, using probabilities conditional upon V. Then he should assume that he does not have the vitamin deficiency, and calculate the expected utility of eating a pickle using his degrees of belief conditional upon ${\sim}V$. Finally, he should compute an overall value to his eating a pickle by weighting the two evidential expected utilities with the corresponding absolute probabilities. So his expected value for eating the pickle will be:

$$\begin{aligned} KEU(E) &= P(V)\{P(M|EV)U(ME) + P(\sim M|EV)U(\sim ME)\} + \\ &\quad P(\sim V)\{P(M|E\sim V)U(ME) + P(\sim M|E\sim V)U(\sim ME)\} \\ &= P(V)\{P(M|V)U(ME) + P(\sim M|V)U(\sim ME)\} + \\ &\quad P(\sim V)\{P(M|\sim V)U(ME) + P(\sim M|\sim V)U(\sim ME)\} \\ &= 1/3\{.75(10) + .25(1010)\} + 2/3\{.12(10) + .88(1010)\} \\ &= 1/3(260) + 2/3(890) \\ &= 680. \end{aligned}$$

(We assume that V and $\sim V$ make no contribution to the utility of an outcome beyond that made by M, $\sim M$, E, and $\sim E$.) Likewise, his expected value for not eating the pickle will be $KEU(\sim E) = 670$. This yields the intuitively correct result that Fred would be better off (by 10 units of utility) eating the pickle. (Note that 'KEU' is an inelegant abbreviation for 'causal expected utility'. The abbreviation 'CEU' is avoided, as that is sometimes used for 'conditional expected utility', which is another name for what we have called 'evidential expected utility'.)

In general the causal decision theorist recommends that the values of actions be evaluated according to the algorithm:

$$KEU(A_j) = \Sigma_k \Sigma_i P(O_i|A_jB_k) \times U(O_i) \times P(B_k),$$

where $\{B_1, B_2, \ldots\}$ is a partition into what Lewis (1981) calls "causal hypotheses." In the case of Fred, there are two causal hypotheses: that he is suffering from the vitamin deficiency, and that he is not.

Defenders of evidential decision theory have developed some clever responses to Newcomb problems; we will catch a glimpse of some in subsequent sections. I will not attempt to argue that these responses fail, but will simply note that if evidential decision theory can be successfully defended, that would surely make the *general* project of articulating a decision-theoretic account of causation easier (although the details of such an account would differ from those of the account developed here). For if evidential decision theory can be defended, decision-theoretic causation will not face the threat of circularity.[6]

The similarity between the problems that spurious correlations pose for probabilistic approaches to causation and those they pose for decision theory might lead us to conjecture that the partition to be used in each case is essentially the same. The language used by advocates of causal decision theory lends some support to this conjecture. Skyrms (1980, 133) characterizes the cells of the relevant partition as ". . . maximally specific specifications of the factors outside of our influence at the time of the decision which are causally relevant to the outcome of our actions. . . " This looks much like the characterization of the cells used in the probabilistic theory of causation; indeed Skyrms claims "[t]he partition, $\{B_k\}$, here [in causal

decision theory] is just the partition relevant to saying whether and how much the act, A_j, has a causal tendency to produce the consequence, O_i, in a world according to the probabilistic theory of causation of the last section."[7] (Skyrms (1988, 63), with minor changes to maintain consistency in notation.) Let us formulate this idea in terms of a precise conjecture: Let X be a variable ranging over an agent's potential actions, so that $X = i$ corresponds to act A_i, and let Y similarly range over the relevant outcomes $O_1, O_2, \ldots$ Similarly, let other factors be represented by variables. Then the B_k's that are to be used in evaluating the causal expected utility of the actions $A_1, A_2, \ldots$ are maximally specific specifications of the values of variables Z, such that Z is causally relevant to X or to Y, but X is not causally relevant to Z.

There are a couple of reasons for thinking that the partitions used in the two theories are not exactly the same, however. One difference between the two theories is that causal decision theory involves an *averaging* over members of the partition, while the theory of probabilistic causation sketched in the previous section does not. To see that this matters, suppose that A is a contemplated action, O is a potential outcome, and B is a factor that is causally relevant to O and probabilistically independent of A—both objectively, and according to our rational agent's degrees of belief. There is thus no spurious correlation between A and B. Nonetheless, the probabilistic theory of causation requires us to hold B fixed when evaluating the causal relevance of A for O (for otherwise we may fail to recognize a case of mixed causal relevance). When deliberating about whether or not to perform act A using the causal decision theorist's algorithm, however, it *does not matter* whether the agent computes her expected utility relative to the partition $\{B, \sim B\}$. In general, when the agent's potential actions are probabilistically independent of the background states, causal decision theory and evidential decision theory yield the same recommendations. This suggests that in our characterization of causal decision theory, we have overstated what needs to be held fixed. Thus the partition of background states that is *needed* for causal decision theory appears to be coarser than that required by probabilistic causality (although no harm comes from using the finer partition).

One difficulty with this objection follows from an observation made by Eells (1991, 75–78). The variable X, representing an agent's range of choices, may be uncorrelated with each of two variables Z and W, and nonetheless be correlated with the joint values of these variables. That is, although knowing the value of Z alone may provide no information about the value of X, and likewise for W, it is nonetheless possible that knowing the values of both Z and W will provide information about X. In this sort of case, failure to hold fixed the values of Z and W may result in spurious correlations.

In any event, the alleged difference between the partitions used by the two theories may simply be an artifact of the particular forms of those theories that have become most popular. Dupré (1984) has advocated a theory of probabilistic causality which effectively involves averaging over background contexts. Alternatively, it may be possible to motivate versions of causal decision theory that do not involve averaging over background contexts. One of the central concerns of Nozick's paper on Newcomb's problem was to determine when it is appropriate to apply *dominance* as a rule of decision. The conclusion toward which he headed was that an agent should prefer one act over another if the first dominates the second over members of the appropriate *c*-partition. For our purposes, it suffices that the partition that is used to determine causal relevance in the probabilistic theory of causation *figure prominently* in our rational deliberations; the foregoing considerations suffice to show that considerations about averaging over background contexts do not undermine this connection.

A second consideration does show that the partitions used in the two theories do not in general have the same form. The partitions will be the same only if the agent knows (or believes with certainty) which factors are potentially relevant to her action and to its outcome (although perhaps not which of these factors are actually present). Put another way, the discussion so far has presupposed that the agent knows which causal variables are relevant to her act and to its outcome, but not which values of those variables obtain. But as Skyrms (1980) and Lewis (1981) have pointed out, the agent may not know this. Consider a variant on the example of Fred the pickle-lover. In this variant, Fred recognizes that there is a correlation between his eating a pickle, and his suffering a muscle ache, but he is not quite sure why this is so. He has two hypotheses: according to hypothesis H_1, the situation is just as described above: a vitamin deficiency is a common cause of Fred's eating a pickle and suffering a muscle ache. According to hypothesis H_2, Fred's eating a pickle causes him to suffer a vitamin deficiency, which in turn causes him to suffer a muscle ache. According to H_1, Fred should hold V or $\sim V$ (the presence or absence of vitamin deficiency) fixed when deliberating about whether or not to eat a pickle; according to H_2, he must *not* hold V or $\sim V$ fixed. So should he or shouldn't he? The solution is to incorporate the hypotheses H_1 and H_2 into the partition, so that he calculates the expected value of eating a pickle relative to H_1V, $H_1\sim V$, and H_2. In general, when evaluating his choices of action, he must hold fixed propositions of the form $H_k\&B_{kl}$, where B_{kl} is the lth maximally specific specification of factors that are beyond the agent's control *according to hypothesis* H_k. Once again, this observation does not undermine the very general point that *c*-partitions play a prominent role in our deliberations, and thus does not undermine the strategy of appealing to decision theory to motivate our peculiar interest in *c*-partitions. In the

discussion that follows, we will retain the assumption that the agent puts all of her credence in just one hypothesis about which factors are beyond her control, while remembering that this is a special case.

Recall our problem: According to the probabilistic theory of causation advocated in the previous section, causal claims — at least those that ascribe different types of causal relevance — provide qualitative information about probability relations. But the probability relations that are so described are of a particularly baroque sort: conditional distribution functions within cells of the *c*-partition. Why should we find these probability relations, among all of the probably relations instantiated in the physical world, so interesting? The answer we have been exploring is that these are the types of probabilities that enter into our evaluations of potential actions. Of course, the probabilities that figure in our deliberations are subjective, while those that constitute the different relations of causal relevance are objective. Is our interest in objective probabilities conditional upon members of the *c*-partition motivated only by a *formal analogy* with the subjective probabilities that guide our choices? Mellor (1971), Skyrms (1980, 1984b) and Lewis (1981) have all argued that rational degrees of belief will be guided by beliefs about objective probabilities. If this is correct, then the subjective probabilities that figure in the rational evaluation of potential actions, being probabilities conditional upon members of the *c*-partition, will be informed by beliefs about objective probabilities conditional upon members of the *c*-partition. In other words, our beliefs about the objective probabilities that constitute the various relations of causal relevance will guide our deliberations. No wonder, then, that we have a disproportionate interest in just these objective probabilities.

Can this appeal to causal decision theory be made without circularity? There are two potential sources of circularity. First, the construction of the *c*-partition that figures in the causal decision-theorist's algorithm makes reference to causal concepts. If one believes, like Papineau, that it is possible to construct the *c*-partition without any reference to causality, then this threat of circularity will not arise. But even if this is not possible, the circularity involved need be no worse than that already admitted by many advocates of probabilistic theories of causation. The theory sketched in the previous section invoked a primitive relation of causal relevance to construct the *c*-partition, and then defined additional, important, causal concepts in terms of this primitive and probabilities. If only this causal primitive is required for the construction of the *c*-partition, then only this causal primitive is required in the formulation of the causal decision theorist's algorithm. Thus the appeal to causal decision theory may invoke some causal notions, but not those very notions that the probabilistic theory of causation is intended to explicate.

The second potential source of circularity concerns the rationale for

applying the algorithm recommended by the causal decision theorist. One rationale might be as follows: acts are valuable insofar as they tend to *cause* or *promote* desirable outcomes, and, according to the probabilistic theory of causation, we must assess these causal tendencies by looking at conditional probabilities within the cells of the *c*-partition. If this is the rationale, then the appeal to causal decision theory will indeed be circular. But perhaps it is possible to motivate the evaluation of acts using the *c*-partition in some other manner. In the next section, I will explore the possibility of providing such an alternative rationale.

4. Agent Probabilities

Huw Price (1991) has attempted to provide an agency based theory of causation in terms of evidential decision theory. He argues that a rational agent's evidential probabilities will *not* exhibit the sorts of spurious correlations that give rise to Newcomb problems. Thus if one understands the probabilities figuring in a probabilistic theory of causation as *agent probabilities,* probabilities as an agent ought to figure them, then the problems with spurious correlations that haunt probabilistic theories of causation need not arise.

It would be nice to have a clearer mathematical picture of what Price's agent probabilities are.[8] Price (1993, 261) does offer the following suggestive idea:

> Ramsey [identifies] what he takes to be the crux of the agent's perspective, namely the fact that from the agent's point of view contemplated actions are always considered to be *sui generis,* uncaused by external factors. As he puts it, "my present action is an ultimate and the only ultimate contingency." [Ramsey 1978, 146] I think this amounts to the view that free actions are treated as probabilistically independent of everything except their effects . . .

There is something deeply right about the idea that deliberation requires that contemplated actions be viewed as *sui generis.* Indeed, part of what is puzzling about Newcomb problems is that one is asked to deliberate about which course of action to pursue in the knowledge that one's actions are, in part, caused by factors beyond one's control. Nozick suggests that in the extreme case where one's action is completely beyond one's control, the very idea of choice becomes incoherent:

> To get the mind to really boggle, consider the following [table of outcomes]:

	B_1	B_2
A_1:	10	4
A_2:	8	3

> Suppose that you know that either B_1 or B_2 already obtains, but you do not know which, and you know that B_1 will cause you to do A_2, and B_2 will cause you to do A_1. Now choose! ("Choose?") (Nozick (1969, 141), with some modifications in notation.)

This line of thought is also behind Eells' (1982, 1985) defense of evidential decision theory. Eells argues that for a rational agent, Newcomb problems cannot even arise, for part of what it is to be rational is to have one's actions determined exclusively by one's beliefs and desires (in conjunction with a decision rule). If factors such as vitamin deficiencies influence the actions of a rational agent, it can only be by causing her to have certain beliefs and desires. Moreover, a rational agent ought to know her own beliefs and desires. For such an agent, choices can never provide new information about background states in such a way as to give rise to Newcomb problems. Lewis rightly responds that we should seek a decision theory for rational agents of a less idealized sort:

> [Eells' d]efence does establish that a Newcomb problem cannot arise for a fully rational agent, but . . . decision theory should not be limited to apply only to the fully rational agent . . . Not so, at least, if rationality is taken to include self-knowledge. May we not ask what choice would be rational for the partly rational agent, and whether or not his partly rational methods of decision will steer him correctly? (Lewis (1981, 10).)

The Ramsey-Price proposal strikes the right balance: it incorporates the insight that there is something funny about deliberating about actions that one knows to be caused by external factors, without withholding decision theory from agents that occasionally have their actions so caused. The idea is that rational deliberation demands of an agent that she entertain the *fiction* that her acts are *sui-generis;* deliberation is the determination of what an agent would do if she were completely free to act in accordance with her interests.

I propose to render this piece of 'as if' reasoning precise with the aid of a mathematical trick originally performed by Brian Skyrms (1980, 135–136). Let the probability function P_d represent the rational agent's 'disinterested' degrees of belief. This includes the agent's degrees of belief about various factors beyond her control, but also her degrees of belief about her own actions, when assessed in a disinterested and non-self-deceptive way, much as an independent observer might. As an intuitive crutch, one can think of P_d as providing the odds that an agent would use if she were to bet on the actions of a twin, known to have identical behavioral dispositions. P_d will reflect all of those influences on the agent's actions that she believes to be beyond her control; it incorporates all of the spurious correlations that give

rise to Newcomb problems. But P_d is *not* the probability function that the agent uses in the evaluation of her own possible courses of action. For this purpose, she must entertain the fiction that her actions are *sui generis,* which, following Price, we will take to mean probabilistically independent of factors that are beyond her control.[9] When evaluating her various courses of action, then, she will use what Skyrms calls a "fictitious" probability distribution, P_f, in which her acts are probabilistically independent of the cells of the c-partition (which specify the factors that are beyond her control). This fictitious probability will then play the role of Price's agent probability.

What properties should the fictitious probability distribution have? Since the acts $A_1, A_2, \ldots$ are being treated as *sui-generis,* the probability assigned to these acts need not correspond to the realistic probabilities $P_d(A_1)$, $P_d(A_1)$, . . . Instead, we will allow these probabilities to be arbitrarily assigned. So let: 1) $P_f(A_j) = p_j$, with $\Sigma_j\, p_j = 1$. We will see below that the evaluation of actions will be independent of the values of the p_j's. Relative to P_f, the actions $A_1, A_2, \ldots$ should be independent of the members of the c-partition $B_1, B_2, \ldots$; thus P_f should satisfy: 2) $P_f(A_jB_k) = P_f(A_j)P_f(B_k)$ for all j and k. But P_f should otherwise be as similar to P_d as possible. This condition is difficult to formulate precisely; for our purposes, however, it suffices that P_f agree with P_d about the relative likelihood of various factors beyond the agent's control, and about how her acts, in conjunction with those factors, influence the occurrence of the various possible outcomes. So P_f should satisfy: 3) $P_f(B_k) = P_d(B_k)$ and 4) $P_f(O_i|A_jB_k) = P_d(O_i|A_jB_k)$. It is shown in the appendix that such a P_f always exists; P_f will satisfy conditions 1 through 4 if it is constructed from P_d according to the following formula:

$$P_f(\bullet) = \Sigma_j\Sigma_k\, P_d(\bullet|A_jB_k)p_jP_d(B_k), \text{ where } \Sigma_j\, p_j = 1.^{10}$$

What is the upshot of evaluating the evidential expected utility of an act A_j using a fictitious probability space P_f satisfying conditions 1 through 4? In the appendix, it is shown that this evidential expected utility is identical to the causal expected utility of A_j as calculated using P_d. In this way, we can make precise Price's claim that evidential decision theory suffices if one uses agent probabilities.

The c-partition is of prime interest to us because it figures in our algorithm for evaluating various courses of action. Although the construction of the c-partition does involve a primitive relation of causal relevance, the use of the c-partition can be motivated without appeal to the probabilistic theory of causation. We can provide an account of why we evaluate our actions in terms of probabilities relative to elements of the c-partition without saying: 'because those probabilities tell us in what way our actions are causally relevant to the outcomes of interest to us'. Rather, we evaluate

our actions in terms of these probabilities because they allow us to simulate the independence of our actions from factors beyond our control. Rational deliberation requires that we view our actions in this way, even if only in fiction.

5. Experimentation

The process of deliberation, as described above, involves the creation of a fictitious probability space in which acts are probabilistically independent of factors that are beyond the agent's control. But there is a context in which we actually bring about objective probability distributions that are analogous to the fictitious probability distributions described above: controlled experimentation. Suppose, for example, that a research team is interested in determining the efficacy of a new drug designed to reduce hypertension. One strategy they might follow would be to make the drug commercially available, and then take measurements of the blood pressure of those that take the new drug, and compare them with those of individuals who do not take it. This would be a very poor research strategy. The choice to take the new drug is bound to be correlated with all sorts of factors that are themselves causally relevant to blood pressure: socioeconomic status, access to a research hospital, overall quality of medical care, and so on. Any discovered correlation between choice of drug and drop in blood pressure could well be a spurious correlation of the sort that we have become quite familiar with.

The accepted procedure is to conduct a controlled experiment. A pool of subjects is divided *randomly* into a treatment group and a control group (who will receive a placebo, or perhaps an established form of treatment if the denial of treatment would be unethical). The frequency of blood pressure decrease will then be compared across these two groups. The effect of randomly assigning subjects to the two groups is to render the causal variable—in this case treatment with the new drug vs. 'treatment' with a placebo—probabilistically independent of those variables to which it's not causally relevant. Although, under normal circumstances, there are a great many factors that are causally relevant to or otherwise correlated with the decision to take the new drug, the effect of randomization is to destroy these correlations. Thus, the creation of artificial probability distributions is not something that only takes place in the imagination.

6. Conclusion

Evidential decision theorists such as Eells have argued that a rational agent's subjective probabilities will always be such that evidential and causal decision theory yield the same prescriptions. A more modest proposal, due to Price and Ramsey, is that the context of deliberation demands

of a rational agent that she adopt a certain perspective from within which the prescriptions of evidential decision theory are the same as those of causal decision theory. I have made this proposal concrete by suggesting that the agent deliberates using a fictitious probability measure that differs from her realistic, or disinterested probability measure. The agent's evidential deliberations using her fictitious probability measure are demonstrably equivalent to her causal deliberations using her disinterested measure. The construction of the fictitious measure, like the computation of causal expected utility, invokes the relation of causal relevance. The appeal to causation is mitigated by two factors, however. First, the primitive relation appealed to is not one of the causal relations that is to be *defined* within a probabilistic theory of causation; it is rather the relation that must be presupposed by such a theory. Second, the claim that the relation of causal relevance figures in an agent's deliberations in the manner described is not motivated by appeal to probabilistic theories of causation. That is, the agent does not evaluate her various courses of action while holding fixed certain causally relevant factors *because* this is the way to assess the expected effects of her actions. The rationale is rather that deliberation requires the assumption of a freedom of action that may, as a matter of fact, not exist.

Because of these mitigating factors, the theorist interested in probabilistic causality may help himself to causal decision theory without vicious circularity. With causal decision theory safely in hand, the theorist may provide an account of the peculiar mathematical structure of his probabilistic theory of causation. Causal decision theory requires that we evaluate our acts using probabilities that are conditional upon the cells of a certain partition, the *c*-partition. The world contains events and properties that are related to each other in ways that can be represented using probabilities; our beliefs about these objective probabilities drive our subjective probabilities. We are particularly interested in those objective probabilities that inform the subjective probabilities that figure in our rational deliberations, that is, objective probabilities that are conditional upon members of the *c*-partition. It is hardly surprising, then, that we have developed a particular mode of discourse—the ascription of types of causal relevance—to talk about them.[11]

Appendix

Let P_d be a probability function defined on a finite algebra of events that includes the following three families of exclusive and exhaustive events: $A_1, A_2, \ldots, A_m$ (representing various acts); $B_1, B_2, \ldots, B_n$ (representing causal background conditions); and $O_1, O_2, \ldots, O_r$ (representing possible outcomes). Assume that $P_d(A_jB_k) \neq 0$ for all $1 \leq j \leq m$, $1 \leq k \leq n$. Let P_f be a function on the same algebra defined as follows:

$P_f(\bullet) = \Sigma_j \Sigma_k\ P_d(\bullet|A_jB_k)p_jP_d(B_k)$, where $\Sigma_j\ p_j = 1$.

Then P_f has the following properties:

1. P_f is a probability function.
Proof: $\Sigma_j \Sigma_k\ p_jP_d(B_k) = 1$, and for all j, k, $P_d(\bullet|A_jB_k)$ is a probability function. Thus P_f is a mixture of probability functions, and hence a probability function itself.

2. For fixed j, $P_f(A_j) = p_j$.
Proof: $P_f(A_j) = \Sigma_k\ P_d(A_j|A_jB_k)p_jP_d(B_k)$
$= \Sigma_k\ p_jP_d(B_k) = p_j$.

3. For fixed k, $P_f(B_k) = P_d(B_k)$.
Proof: $P_f(B_k) = \Sigma_j\ P_d(B_k|A_jB_k)p_jP_d(B_k)$
$= \Sigma_j\ p_jP_d(B_k) = P_d(B_k)$.

4. For fixed i, j, k, $P_f(O_i|A_jB_k) = P_d(O_i|A_jB_k)$.
Proof: $P_f(O_i|A_jB_k) = P_f(O_iA_jB_k)/P_f(A_jB_k)$
$= P_d(O_iA_jB_k|A_jB_k)p_jP_d(B_k)/P_d(A_jB_k|A_jB_k)p_jP_d(B_k)$
$= P_d(O_iA_jB_k|A_jB_k)/P_d(A_jB_k|A_jB_k) = P_d(O_i|A_jB_k)/1$

5. For fixed j, k, $P_f(A_jB_k) = P_f(A_j)P_f(B_k)$
Proof: $P_f(A_jB_k) = P_d(A_jB_k|A_jB_k)p_jP_d(B_k)$
$= p_jP_d(B_k)$
$= P_f(A_j)P_f(B_k)$ by 2 and 3.

Let P_d be as above, and let P_f be any measure on the same algebra of events satisfying 1 through 5. Then P_f has the following properties.

6. Let U be a function that assigns real numbers (utilities) to O_1, O_2, . . . , O_r.
Then for fixed j, $\Sigma_i\ P_f(O_i|A_j)U(O_i) = \Sigma_k\Sigma_i\ P_d(O_i|A_jB_k)U(O_i)P_d(B_k)$
Proof: For fixed i, j, $P_f(O_i|A_j) = \Sigma_k\ P_f(O_i|A_jB_k)P_f(B_k|A_j)$
$= \Sigma_k\ P_f(O_i|A_jB_k)P_f(B_k)$ (by 5)
$= \Sigma_k\ P_d(O_i|A_jB_k)P_d(B_k)$ (by 3 and 4)
Thus $\Sigma_i\ P_f(O_i|A_j)U(O_i) = \Sigma_i\ \{\Sigma_k\ P_d(O_i|A_jB_k)P_d(B_k)\}U(O_i)$
$= \Sigma_k\Sigma_i\ P_d(O_i|A_jB_k)U(O_i)P_d(B_k)$

Notes

[1]I do not mean to deny, of course, that there are important differences between my approach and Mellor's.

[2]According to Eells (1991), causal claims are made relative to a population and a population-type. While these relata are important in squaring the theory with certain intuitions, I will ignore them in what follows.

[3]Requiring that causes of C, as well as causes of E, be held fixed allows us to avoid some difficulties raised by Eells (1991; 140–142, 245–246). In order to avoid all problem cases, it

may become necessary to hold fixed all *causal ancestors* of C and E, where A is a causal ancestor of C iff there exists a chain of factors $F_1, F_2, \ldots, F_n$ such that A is causally relevant to F_1, F_1 is causally relevant to $F_2, \ldots, F_n$ is causally relevant to C. (On Eells' theory, causal relevance may fail to be transitive due to fortuitous cancellation of probabilities.)

[4]In order to avoid some of the counterexamples described by Eells (1991, 202–206), we must refrain from holding fixed any factor that is a causal descendent of C, where F is a causal descendent of C iff C is a causal ancestor of F. (See the previous footnote for the definition of causal ancestry.)

[5]See, e.g., Cartwright (1979) and Eells (1991).

[6]The reader is referred to Price (1991) and Jeffrey (1993) for examples of this sort of program.

[7]Of course, Skyrms was referring to the last section of his own paper, but since a similar probabilistic theory of causation was also introduced in the last section of this paper, the quote remains appropriate.

[8]Alternative mathematical approaches to this sort of problem can be found in Kyburg (1980) and Meek and Glymour (1994).

[9]The principle invoked here is that causal independence implies probabilistic independence. Perhaps this is making a minimal appeal to a probabilistic theory of causation. Since we are not presupposing anything about probabilities conditional upon members of the c-partition, however, we are not invoking that feature of the probabilistic theory of causation presented in section 2 that we wish to justify.

[10]This is the unique probability measure satisfying 1 through 4 if we take P_d and P_f to be defined only on the Boolean algebra generated by $\{O_i\}$, $\{A_j\}$, and $\{B_k\}$.

[11]For their comments upon earlier drafts, I would like to thank Hugh Mellor, David Papineau, Brian Skyrms, an anonymous referee from *Noûs*, audience members at the Silver Jubilee meeting of the Society of Exact Philosophy in Calgary, and audience members at Princeton University (particularly Dick Jeffrey and David Lewis). I regret that spatial considerations have prevented me from giving fair treatment to all of their suggestions and challenges.

References

Cartwright, Nancy. (1979) "Causal Laws and Effective Strategies," *Noûs* 13: 419–437.

Dupré, John. (1984) "Probabilistic Causality Emancipated," in French, Uehling, and Wettstein (1984), pp. 169–175.

Eells, Ellery. (1982) *Rational Decision and Causality* (Cambridge, U.K.: Cambridge University Press).

———. (1985) "Causality, Decision, and Newcomb's Paradox," in *Paradoxes of Rationality and Cooperation* edited by Richmond Campbell and Lanning Sowden (Vancouver: University of British Columbia Press), pp. 183–213.

———. (1991) *Probabilistic Causality* (Cambridge, U.K.: Cambridge University Press).

French, Peter, Theodore Uehling, Jr., and Howard Wettstein. (1984) *Midwest Studies in Philosophy IX* (Minneapolis: University of Minnesota Press).

Gasking, Douglas. (1955) "Causation and Recipes," *Mind* 64: 479–487.

Hitchcock, Christopher. (1993) "A Generalized Probabilistic Theory of Causal Relevance," *Synthese* 97: 335–364.

Hull, David, Mickey Forbes, and Kathleen Okruhlik. (1993) *PSA 1992, Volume Two* (East Lansing: Philosophy of Science Association).

Jeffrey, Richard. (1983) *The Logic of Decision,* 2nd edition (Chicago: University of Chicago Press).

———. (1993) "Probability Kinematics and Causality," in Hull, Forbes, and Okruhlik (1993), pp. 365–373

Kyburg, Henry. (1980) "Acts and Conditional Probabilities," *Theory and Decision* 12: 149–171.

Lewis, David. (1980) "A Subjectivist's Guide to Objective Chance," in *Studies in Inductive Logic and Probability, Vol. II,* edited by Richard Jeffrey (Berkeley: University of California Press), pp. 263–294.

——. (1981) "Causal Decision Theory," *Australasian Journal of Philosophy* 59: 5–30.

Meek, Christopher, and Clark Glymour. (1994) "Conditioning and Intervening," *British Journal for the Philosophy of Science* 45: 1001–1021.

Mellor, D. Hugh. (1971) *The Matter of Chance* (Cambridge, U.K.: Cambridge University Press).

——. (1988) "On Raising the Chances of Effects," in *Probability and Causality: Essays in Honor of Wesley C. Salmon,* edited by James Fetzer (Dordrecht: Kluwer), pp. 227–239.

Nozick, Robert. (1969) "Newcomb's Problem and Two Principles of Choice," in *Essays in Honor of Carl G. Hempel,* edited by Nicholas Rescher (Dordrecht: Reidel), pp. 114–146.

Papineau, David. (1989) "Pure, Mixed, and Spurious Probabilities and Their Significance for a Reductionist Theory of Causation," in *Scientific Explanation,* edited by Philip Kitcher and Wesley Salmon (Minneapolis: University of Minnesota Press), pp. 307–348.

——. (1993) "Can We Reduce Causal Direction to Probabilities?", in Hull, Forbes and Okruhlik (1993), pp. 238–252.

Price, Huw. (1991) "Agency and Probabilistic Causality," *British Journal for the Philosophy of Science* 42: 157–76.

——. (1993): "The Direction of Causation: Ramsey's Ultimate Contingency," in Hull, Forbes and Okruhlik (1993), pp. 253–267.

Ramsey, Frank P. (1978) "General Propositions and Causality," in *Foundations: Essays in Philosophy, Logic, Mathematics and Economics,* edited by D. Hugh Mellor (London: Routledge and Kegan Paul), pp. 133–151.

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

Skyrms, Brian. (1980) *Causal Necessity* (New Haven: Yale University Press).

——. (1984a) "EPR: Lessons for Metaphysics," in French, Uehling, and Wettstein (1984), pp. 245–255.

——. (1984b) *Pragmatics and Empiricism* (New Haven: Yale University Press).

——. (1988) "Probability and Causation," *Journal of Econometrics* 39: 53–68.