Conditioning and Intervening

Christopher Meek; Clark Glymour

Conditioning and Intervening

Christopher Meek and Clark Glymour

ABSTRACT

We consider the dispute between causal decision theorists and evidential decision theorists over Newcomb-like problems. We introduce a framework relating causation and directed graphs developed by Spirtes *et al.* (1993) and evaluate several arguments in this context. We argue that much of the debate between the two camps is misplaced; the disputes turn on the distinction between conditioning on an event E as against conditioning on an event I which is an action to bring about E. We give the essential machinery for calculating the effect of an intervention and consider recent work which extends the basic account given here to the case where causal knowledge is incomplete.

1 *The Markov condition*
2 *Intervening and conditioning*
3 *Inquiry and decision*

Two essays have come to our attention. One, a recent piece in *Science*, claims that despite the expense of 200 million dollars the Agency for Health Care Policy and Research (AHCPR) has failed to extract from records of medical data a single correct conclusion relevant to decisions to alter medical practice.[1] The other, an essay of Robert Nozick's long familiar to most philosophers [1969], concerns difficulties in the theory of rational decision-making that arise when actions and their outcomes are both influenced by some feature of the world not under the agent's control. We will argue that both essays turn on the same logical facts, and that partial resolutions of the problems to which they call attention are found in recent work on causal inference and the design of empirical studies. Our exposition invokes a parallelism in the philosophical discussion of rational decision-making and the statistical discussion of experimental design. We begin with Nozick's problem.

Suppose you believe, truly and with justification, that a genetic factor causes people both to smoke and to contract cancer of the lung, and that smoking itself has no influence on disease. Suppose further that you believe you would enjoy smoking but it is very much more important to you that you not die of cancer. The probabilities you assign accord with these beliefs: the probability that you will get cancer, given that you smoke (in

---

[1] *Science* [1994], **263**, pp. 1080–2.

the future) is greater than the corresponding probability given that you do not smoke (in the future), and cancer and smoking are independent in probability conditional on the value of your genotype. Your utilities likewise accord with these preferences. A simple calculation shows the expected utility of not smoking is greater than the expected utility of smoking; but, whatever your unknown genotype, you are better off smoking than not. Ought you to smoke?

Nozick's essay considers variants of this question, which he views as indications of a conflict between two decision theoretic principles: if, in every possible circumstance, one action gives outcomes at least as good as another and in some circumstances better, then the first action should be preferred to the second (weak dominance), and if the sum over all possible states of the world of the probability of the state multiplied by the value of the outcome produced by one action in that state is greater than the like sum for a second action, then the first action is to be preferred to the second (expected utility). Nozick's conclusion is that the expected utility principle must give way to the dominance principle 'if the actions or decisions to do the actions do not affect, help bring about, influence, and so on, *which* state obtains ...'[2]

Since Nozick's essay appeared, a considerable literature has developed around 'causal decision theory'. Following a suggestion of Robert Stalnaker, Alan Gibbard and William Harper claimed the issue is not between dominance and expected utility, but rather between two forms of the expected utility principle, one using probabilities of states conditional on actions and the other using probabilities for subjunctive (Gibbard and Harper say 'counterfactual') claims as to the consequences were some particular action to be taken. A related theory has been published by David Lewis.[3] Avoiding subjunctives, Brian Skyrms proposed that one should choose the action that maximizes the sum, over each outcome of interest and each 'maximally specific specification of factors outside our influence', of the product of the probability of the outcome given the action and the factor value, the probability of the factor value, and the utility of the outcome, action, and factor value. Where no unique set of such specific factors is known, Skyrms proposed that the probabilities be mixtures over

[2] Newcomb's problem—you must decide either to open only the second of two boxes or to open both, knowing that a perfect predictor of your decision put a thousand dollars in the first box and one million dollars in the second if he predicted you would open only the second box, and nothing in the second box if he predicted you would open both—is formally a special case of a problem of this sort, in which the decision-maker believes that both the probability of smoking and the probability of contracting cancer given the appropriate genotype are one, and strict dominance holds.

[3] R. Stalnaker [1972]; A. Gibbard and W. Harper [1978]; D. Lewis [1981]. Stalnaker's suggestion is anticipated by P. Fishburn [1964].

alternative sets of specific factors, and he showed his theory to be formally equivalent to the account proposed by Lewis. Brad Armendt has given a representation theorem for Skyrms' theory in terms of axioms on preferences over appropriate objects.[4]

No idea is without objections. Ellery Eells [1982] argued that in assessing choices the probabilities used must be conditional on the total evidence, and that in the relevant cases Nozick and others consider the disposition to take a particular action is itself evidence conditional on which the state of the world and the action are independent. By using 'evidental' decision theory and claiming the situation contains additional information, Eells obtains, in almost all cases, the same decisions as do advocates of causal decision theory. At least in the most celebrated case of this kind, Newcomb's problem, Teddy Seidenfeld rejects altogether Nozick's conclusion that conditional expected utility must give way to dominance, and rejects the causal decision theorists' calculations: for example, in the extreme case of Newcomb's problem with a perfect predictor, one is faced with a decision under *certainty*: you are certain to be better off if you open only one box.[5]

1 The Markov condition

Most of the discussion of causal decision theory, both for and against, has shared assumptions about the connection between beliefs about causality and beliefs about probability (or about the connection between causality and probability) that are perfectly reflected in the statistical literature on causal inference. Most of the philosophical discussion, from Nozick on, supposes that the causal relations in the smoking example, which we may represent graphically as

$$S \nearrow \overset{G}{} \searrow C$$
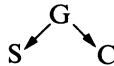
imply a 'factorization' of the probabilities;[6] in obvious notation:

$$P(S, C, G) = P(S|G)P(C|G)P(G).$$

All of the philosophical commentators appear to agree that if smoking and cancer have no effect on one another and have only genotype as their

[4] B. Skyrms [1980], [1982] and [1984]; B. Armendt [1988]. For a view similar to Skyrms, see D. Papineau [1989].
[5] T. Seidenfeld [1990].
[6] There are exceptions. In their [1986], Eells and Sober appeal to the notion 'interactive fork' in which the causal relations are represented as in the graph above but the factorization does not hold. In our view all putative examples of 'interactive forks' that are not from quantum mechanics are simply cases where further causal connections, as between S and C or between other causes and S and C, have been omitted. That of course was the point of Simon's [1954]. For a discussion, see Spirtes, Glymour, and Scheines [1993].

common cause, then they are independent conditional on genotype, or in the notation now common among statisticians:

$$S \perp\!\!\!\perp C | G$$

which is an immediate consequence of the factorization formula. Related claims are made more generally in the statistical literature on factor analysis, which supposes that when two or more variables have no influence on one another they are independent conditional on the set of all of their common causes.[7] Herbert Simon's once influential paper on 'spurious correlation' made the same point.[8]

No philosophical commenator appears to disagree with Eells that if in the same circumstance genotype only produces smoking through the occurrence of a conscious desire to smoke, they smoking and cancer are independent conditional on the desire to smoke. Eells represents the circumstance graphically, as can we.
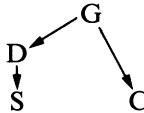
If the Causal Markov condition (see below) holds, then the graph of causal relations implies that the probabilities satisfy the factorization

$$P(SDCG) = P(S|D)P(D|G)P(C|G)P(G)$$

and that Eells' independence claim

$$S \perp\!\!\!\perp C | D$$

follows. Again, related claims are made in the statistical literature on experimental design. Consider an example due to Donald Rubin. In an educational experiment in which reading program assignments T are assigned on the basis of a randomly sampled value of some pre-test variable X which shares one or more unmeasured common causes, V, with Y, the score on a post-test, we wish to predict the average difference τ in Y values if all students in the population were given treatment $T = 1$ as against if all students were given treatment $T = 2$. The situation in the experiment is represented by the graph

---
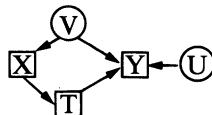
[7] For a particularly clear statement of this assumption, see D. Bartholomew [1987].
[8] H. Simon [1954].

Rubin's analysis assumes that if the influence from T to Y does not exist, then T and Y are independent conditional on X.[9]

A special case arises when an action has no influence on an outcome, and neither is there a common cause of the action and the outcome. Then we should expect the action and the outcome to be independent. The same principle is of course at the center of experimental design. The 'null hypotheses' that Ronald Fisher introduced in *The Design of Experiments* are hypotheses of independence (or consequences of such hypotheses) associated with the hypothesis that the treatment has no influence on the outcome in experiments arranged so that it is known that no causes act to determine directly both the treatment a unit receives and the outcome it exhibits.[10] Randomization of treatment assignments serves two purposes in Fisher's theory: one is to help ensure that treatment assignment and outcome have no common cause and are independent if treatment has no effect on outcome; the other is to determine a definite joint probability distribution for treatment and outcome under the assumption of no effect. Even Kadane and Seidenfeld [1990],[11] the most articulate critics from a subjectivist viewpoint of Fisher's views on experimental design, assume that Fisher is correct that when there is no causal relation of any kind between treatment and outcome they should be regarded as independent.

Hans Reichenbach [1956] formulated various general connections between causal hypotheses and probability constraints, but the relation that seems correctly and fully to generalize these and other uncontroversial examples was first formulated by Kiiveri and Speed [1982], who titled it, perhaps unfairly to Reichenbach, the 'Markov condition': We will give it in two versions:

Causal Markov Condition (Frequency Version): Let *G* be a directed acyclic graph describing the causal relations among a set V of variables, where every common cause of variation of two or more variables in V is itself in V, and let P be a population of units all sharing the same causal relations *G*. Let *P* be the frequency distribution of variables in V. Then every variable X in V is independent of its nondescendants (i.e., of variables it does not affect) conditional on its parents.

Causal Markov Condition (Subjective Version): Let *G* describe the causal relations among a set V of variables, where every common cause of

[9] D. Rubin [1977].
[10] R. Fisher [1951].
[11] C. Howson and P. Urbach [1988] also criticize Fisher from a subjectivist viewpoint, and appear to agree with the independence claim. Their chief complaint against Fisher seems to be that randomization does not guarantee the absence of an unintended causal connection between the outcome and some concomitant of treatment, which is correct but seems an odd argument against randomizing at all.

variation of two or more variables in **V** is itself in **V**, and let P be the distribution of degrees of belief over **V** conditional on G. Then $\langle G, P \rangle$ ought to be so related that every variable X in **V** is independent of its nondescendants conditional on its parents.

The independence relations in the philosophical examples follow from treating the relevant properties—smoking and cancer, for example—as binary variables, and from the fact that for sets of random variables **X**, **Y**, **Z**, if **X** is independent of **Y** given **Z** then each member of **X** is independent of each member of **Y** given **Z**. It is important to stress that the causal structure constrains the probabilistic structure, e.g. independence relationships, if the Causal Markov condition obtains.

The Markov condition entails a version of what has been called the common cause principle, specifically: if in a system of variables satisfying the Markov condition, neither X nor Y influences one another and they are statistically dependent, then there exists a set **Z** of variables not containing X or Y but causing both, and conditional on **Z** X and Y are independent. Elliott Sober [1987] objects that the common cause principle typically appears to be violated in mutually increasing (or decreasing) time series, where variables without apparent causal connection seem to be probabilistically dependent. Sober's point was made early in this century by G. Udny Yule [1926], who attributed such statistical dependencies either to an unobserved common cause or to an unrepresentative sample or to mixing populations with different causal structures and different probability distributions, all consistent with holding that the Markov condition describes the causal relations among an appropriately extended set of variables in such time series. Frank Arntzenius[12] has formulated a version of the common cause principle that requires causes to precede their effects, and has given a number of counterexamples, either quantum mechanical or else involving purely logical relations—as in correlations among two descriptions of the same event—or violating his temporal order requirement. Only the quantum mechanical cases clearly violate the Markov condition, and presumably such circumstances rarely if ever apply in making decisions. The Bayesian version of the Markov condition is perhaps more difficult to justify, since, as Brian Skyrms has noted,[13] nothing prevents one from believing that an odd lot of dissimilar appearing coins of different manufacture and history all have exactly the same bias. In that case the outcomes of flips of one coin are not independent of flips of others, although there is no direct influence and no common cause, and the subjective Markov

---

[12] F. Arntzenius, 'The Common Cause Principle', preprint, Department of Philosophy, University of Southern California.
[13] (Hitherto) private communication.

condition is not satisfied. Such beliefs might appear very odd but they are not incoherent; still, as the examples of Kadane, Seidenfeld, and Rubin illustrate, and as examples in the next section will further show, statisticians writing from the subjective Bayesian perspective like to set up probabilities in agreement with the Markov condition.

We emphasize that the Causal Markov condition, in either version, *does not* imply that probabilities uniquely determine causal relations. In general, quite distinct directed graphs can, according to the Causal Markov condition, imply exactly the same constraints—the same conditional independence relations—on admissible probability distributions. For example, the graphs $X \to Y \to Z$ and $X \leftarrow Y \leftarrow Z$ both imply that X and Z are independent conditional on Y, and imply no other constraints on probabilities. Characterizations of the classes of graphs that imply the same constraints on probabilities according to the Markov condition, as well as according to other general hypotheses about the connection between causal graphs and probabilities, are given in Spirtes, Glymour, and Scheines [1993].

2 Intervening and conditioning

There is a difference between predicting and deciding, even between predicting one's own fate and choosing it, and the difference is more than resignation. We find something odd in questions about what an agent, no matter whether oneself or another, *ought to do* when one knows the agent's action, whatever it is, will be necessitated by circumstances that cannot be influenced by any response to the question. Both deliberation and advice then seems pointless, and their benefits illusory. Even so, two possible reasons to deliberate or advise suggest themselves. Deliberation, when one believes one's actions are caused by circumstances not under one's control, may have a kind intellectual value: You may wish to know—perhaps to satisfy a curiosity or to regret your limitations— what someone otherwise like you but free to choose among alternative actions would rationally choose to do. Alternatively, one may view decisions, one's own or another's, as the result of the action of a dual system with a default part and an extraordinary part—the default part subject to causes that may also influence the outcome through another mechanism, but the extraordinary part not so influenced and having the power to intervene and displace or modify the productions of the default part. For brevity we will describe the extraordinary part as the Will, although one need only assume, as in an example Isaac Levi [1985] uses, that, without influence by causes of the outcome of interest, decisions can be made to take prior actions that will influence the decision made

subsequently in the original problem. Either account—and there may of course be others that have not occurred to us—requires that we consider two related structures: First, the system of causal and stochastic relations in which the actual agent, or the default part of the actual agent, is enmeshed; second, the system of causal and stochastic relations that would govern an agent freed of the influences on the decision-making of the actual agent, but otherwise like the actual agent, or that would obtain were the Will to act to overrule, or partially to overrule, the causal influences on the default part of the agent.[14]

Any number of scenarios can be—and have been[15]—imagined for the counterfactual causal and probabilistic properties of an agent and a circumstance of decision-making, but what is supposed to happen when the Will intervenes seems comparatively definite. In the smoking case, for example, the Will may intervene to alter the probability of smoking conditional on genotype, and, in the simplest case, the Will simply determines whether or not one smokes, and smoking becomes independent of genotype.

Suppose we consider the simple case, which is complicated enough: When the Will intervenes, the Will alone determines whether one smokes or not; when the Will does not intervene, whether one smokes or not is influenced by genotype. Representing the Will by W, then, the full causal situation is like this:
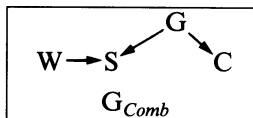
According to the Causal Markov condition, the probability distribution P_{Comb} in this circumstance must satisfy:

$$(1) \qquad P_{Comb}(S, C, G, W) = P(S|W, G)P(C|G)P(G)P(W)$$

We have supposed that three states of the Will are relevant: (i) to intervene for smoking, (ii) to intervene against smoking, or (iii) not to intervene. By elementary calculations, these three states of the Will issue in three conditional probability distributions over G, S, and C, namely:

$$(2) \qquad P_{Comb}(S, C, G|W = i) = P(S|W = i, G)P(C|G)P(G)$$

$$(3) \qquad P_{Comb}(S, C, G|W = ii) = P(S|W = ii, G)P(C|G)P(G)$$

$$(4) \qquad P_{Comb}(S, C, G|W = iii) = P(S|W = iii, G)P(C|G)P(G)$$

[14] We do not wish to suggest that either account is without puzzles.
[15] See T. Horgan [1981].

These conditional distributions are not the same. By assumption, if $S = 1$ stands for smoking, $S = 0$ for not smoking, $P(S = 1|W = i, G) = P(S = 1|W = i) = 1$; $P(S = 1|W = ii, G) = P(S = 1|W = ii) = 0$, and $P(S|W = iii, G) = P(S|G)$. Again by elementary calculations, these three conditional distributions give different distributions for C given S:

(4) $\qquad P_{Comb}(C|S = 1, W = i) = P(C|W = i) = P(C)$

(5) $\qquad P_{Comb}(C|S = 1, W = ii) =$ undefined

(6) $\qquad P_{Comb}(C|S = 1, W = iii) = P(C|S = 1)$

$\qquad = [\Sigma_G P(C|G)P(S = 1|G)P(G)]/P(S = 1)$

(4) is the conditional probability recommended by causal decision theorists for calculating the probability of cancer given smoking; (6) is the conditional probability causal decision theorists claim is required by 'evidential' decision theory. The two recommendations both calculate the maximum expected utilities, but they do so using different conditional probabilities. *The conditional probabilities differ because different events have been conditioned on.* The causal decision theorist conditions on the event of an intervention, willing to smoke, and the stalking horse 'evidential' decision theorist conditions on an event, smoking, that is not an intervention. The difference in the two recommendations does not turn on any difference in normative principles, but on a substantive difference about the causal processes at work in the context of decision making— the causal decision theorist thinks that when someone decides whether to smoke, an intervention occurs, and the 'evidential' decision theorist thinks otherwise.

The calculations in this example illustrate a general theorem that provides the real reason for interpreting as casual hypotheses the directed graphs that encode conditional independence relations according to the Markov condition. Causal claims are generally thought to entail claims—perhaps only *ceteris paribus* claims—about interventions or manipulations. Let G_{Unman} be the directed graph describing the causal relations in a system or population of systems. In the example just considered, G_{Unman} is
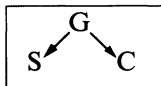
In order to manipulate features of G_{Unman} something must be done, some state of affairs M must be changed in such a way as to force values on some of the variables in G_{Unman}. (In the example just considered, M is of course W.) Expand the graph G_{Unman} by introducing an additional variable, M—

representing the manipulation—and directed edges from the variable M to whatever variables in G_{Unman} are directly controlled or influenced by the manipulation under consideration. M must have a special value, say 0, representing the unperturbed state of affairs, the state of affairs without the intervention or manipulation. Call the expanded graph, G_{Unman} plus M and the additional edges, G_{Comb}. Suppose P_{Unman} is the probability distribution generated by the causal structure G_{Unman} and satisfying the Markov conditon for G_{Unman}. Suppose P_{Comb} is the probability distribution generated by the expanded causal structure G_{Comb} and satisfying the Causal Markov condition for that graph. Then $P_{Comb}(\ |M = 0) = P_{Unman}$. Let $P_{Man = i}$ be the probability distribution of the original variables (those in G_{Unman}) upon a manipulation $M = i$. $P_{Man = i}$ is the thing one wants to predict, the distribution that a genuinely causal hypothesis, G_{Unman}, and P_{Unman} are supposed to provide information about. They do. It follows from the Causal Markov condition that for any value i of M, $P_{Comb}(\ |M = i)$ can be computed from G_{Unman}, P_{Unman}, and the probabilities, conditional on $M = i$, of the variables in G_{Unman} that are directly influenced by M.

A general rule for calculating the manipulated probability distributions from the unmanipulated distribution and causal graph and information about the nature of the intervention is given by the Manipulation theorem.

Manipulation Theorem: With the definitions just given, the manipulated distribution, $P_{Man = i}$, is obtained by replacing in the factorization of P_{Unman} the conditional probabilities for the variables X directly influenced by M with $P_{Comb}(X|M = i)$.[16]

The Bayesian version of the Manipulation Theorem is obvious enough, and specifies what the probabilities ought to be conditional on an intervention. Graphically, in the case of an ideal manipulation—one where a particular value is forced on the manipulated variable (as is the case in the example above)—the theorem amounts to saying that we can compute the probabilities on an intervention by removing all edges directed into a manipulated variable, and using the factorization with the same conditional probabilities as before but in accord with the new, abbreviated graph.

The significance of the Manipulation Theorem for the issue before us is that, if a decision to act is regarded as an intervention in a system—and it

---

[16] The theorem follows from the Markov condition; the rather trivial derivation is given in Spirtes *et al.*, [1993]. The same principle was given as an independent condition in P. Spirtes, C. Glymour, R. Scheines, C. Meek, S. Fienberg, and E. Slate [1991], 'Prediction and Experimental Design with Graphical Causal Models', unpublished, and still earlier in a different but equivalent formalism by J. Robins [1986].

seems that the examples in the literature are all intended to involve interventions of this sort—then there is nothing in the causal decision theorist's examples that threatens ordinary Bayesian decision theory. Nothing at all. The differences in judgments about cases are entirely explained by the fact that different events—sometimes interventions, sometimes not—are conditioned on.

Once again, the philosophical cases parallel a statistical literature. Pratt and Schlaifer, for example, have given *ad hoc* rules for when the conditional probability of Y conditonal on X = x equals the probability of Y when X is forced to have the value x. Their rules follow from the Markov condition and the Manipulation theorem.[17] An equivalent of the Manipulation theorem has for some years been used by James Robins to obtain predictions from non-randomized experiments of the outcomes of randomized trials that were not, but could have been, performed. Consider, for example, an experiment in which patients with HIV virus are randomly assigned to treatment groups given differ dosages of AZT. Some of the patients develop pneumonia, and at the determination of their respective physicians the patients may be given a drug D which acts against pneumonia. For all one knows, AZT may influence mortality through its influence on pneumonia, through its influence on other symptoms that induce physicians to give or withhold the drug D, or directly. Pneumonia influences mortality directly, and also influences whether a patient is given D, which in turn may influence mortality. So the apparent dependencies in the experiment look something like this:

$$AZT \longrightarrow Pneumonia \longrightarrow D \longrightarrow Death$$
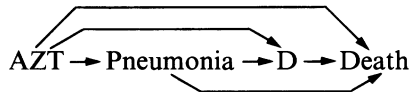
One might put the question of the efficacy of AZT this way: what is the probability of death in a similar experiment in which administraton of D is also randomized? It isn't the probability in the actual experiment without conditioning on pneumonia, and neither is it the probability in the actual experiment conditional on pneumonia. Instead we must think of a randomized experiment as breaking the edges into D and assigning it a new probability independent of pneumonia and AZT, but leaving unchanged the probability of pneumonia and conditional on AZT and the probability of death conditional on D, pneumonia, and AZT. And that is exactly the probability Robins' algorithm computes in answer to the question.[18]

---

[17] J. Pratt and R. Schlaifer [1988]. A derivation of their conditions is given in P. Spirtes, C. Glymour, and R. Scheines [1993].

[18] See J. Robins [1986] and J. Robins, D. Blevins, G. Ritter, and M. Wulfoshn [1982]. In the case Robins *et al.* discuss, there are repeated applications in substantially more difficult than in our simplified example.

Another well-known example that illustrates some of the issues is given by Lindley and Novick [1981]. They consider the two following Simpson-like cases:

Case 1: We obtain data on recoveries for samples of males and females who have received a treatment (t) and a control (c).

Males	R = 1	R = 0
T = t	18	12
T = c	7	3

Females	R = 1	R = 0
T = t	2	8
T = c	9	21

Combined	R = 1	R = 0
T = t	20	20
T = c	16	24

Case 1.

The sample is understood to be very, very large. We are not given information as to how persons were chosen for the sample or for treatment value. Note that the recovery rate is higher for T = c for both males and females but the recovery rate is higher for T = t for the combined group.

In view of this information, if we are presented with a new subject whose gender is unknown, which treatment should we prefer, t or c? Lindley and Novick say we should prefer T = c.

Case 2: We obtain analogous data on yields and heights for samples of black and white plants.

Tall	Y = 1	Y = 0
C = w	18	12
C = b	7	3

Short	Y = 1	Y = 0
C = w	2	8
C = b	9	21

Combined	Y = 1	Y = 0
C = w	20	20
C = b	16	24

Case 2.

This time we must decide whether to plant a white (C = w) or a black variety of plant, in ignorance of the height the plant will grow to. Lindley and Novick say we should prefer C = w.

Lindley and Novick make it quite plain that the difference they see in the two cases is causal, but say they prefer not to talk about that. Instead they say that in the second case, but not the first, the new unit about which the decision must be made is 'exchangeable' in De Finetti's sense with previous units.[19] Lindley's and Novick's idea is that in Case 1 in the sample data the treatment T and gender (which they denote by M) are statistically dependent, but since the decision-maker does not know the gender of the new subject, the new subject's gender can have no 'effect' (their word), no influence, on the choice for the value of T for the new subject, and therefore (?) for the new subject one should treat T and M as statistically independent. By contrast, in Case 2 the decision as to which variety of plant to grow does nothing to alter the causal processes that produce in the sample the statistical dependency between plant height and colour, and therefore (?) one's probability distribution for the new subject should treat plant height and colour as statistically dependent just as they are in the sample. So in the second case the sample conditional probability of Y on C should equal the probability of Y given that variety C is planted. The principles on which the inferences we have queried are grounded are nowhere explained.

The Markov condition and the Manipulation theorem give a smooth account of what is going on in Lindley's and Novick's two cases and of the grounds for the inferences they wish to make. In Case 1 we need not know—or even have detailed views about—the actual causal structure. We need only believe that gender (M) is not an effect (a descendant) of treatment T, and that the decision to treat or not to treat does not otherwise alter the influence of other factors on recovery. We then find that when we decide to impose a treatment on the new subject, T and M must be independent, whereas for subjects in the data sample T and M are not independent, presumably because for them gender in some way influenced the treatment they received. In Case 2, the decision to plant one variety or another does not interfere in the causal processes (e.g. the genetic features) that produce the association between height and colour in the sample, and whatever the causal structure may be, no processes are altered that terminate in colour and connect to height and yield, and by the two conditions we again have Lindley and Novick's solution.

---

[19] The terminology seems to us only to suggest that judgments of exchangeability are related, in a way that remains to be clarified, to judgments about uniformity of causal structure, and that an explicit account of the interaction of causal beliefs and probabilities is necessary to understand when exchangeability should and should not be assumed.

All of the realistic cases in the philosophical disputes that have evolved from Nozick's paper appear to turn on the difference between changing probabilities by conditioning on an event and changing probabilities by intervening—or, in the subjectivist version, the difference between conditioning on an event E and conditioning on an event I which is an intervention to bring about E. Our analysis explains in these cases, and all others, how to calculate the probabilities that result from an intervention when the relevant parts of the causal structure are known, when probabilities are available agreeing with the unperturbed causal structure through the Markov condition, and when the intervention is ideal in the sense of the Manipulation theorem. The probability of an outcome on an intervention can be worked out by using the Markov condition to find the factorization of the joint distribution, substituting according to the Manipulation theorem the distributions forced directly on variables by the intervention, and then computing the marginal probability of the outcome.

We think the Markov condition explains as well some of the views of those who oppose the very idea of causal decision theory. Isaac Levi (*op. cit.*, p. 244), writes:

> Suppose that genotype G yields lifetime smokers 100 per cent of the time and likewise yields sufferers of cancer 100 per cent of the time. Genotype not-G never yields lifetime smokers and never yields sufferers of cancer. Jones can know this about himself and also regard himself as free to choose whether to break the habit or not without contradiction or incoherence.

We read Levi's view as an affirmation of 'soft determinism' and a denial of libertarianism,[20] and likewise a denial that the decision to smoke constitutes an intervention. Using slightly different terminology, Seidenfeld is explicit about the matter in the case of Newcomb's problem, and he is a one-boxer. Given that the actions in the Newcomb case do not alter the relevant causal relations or probabilities, Seidenfeld's judgment is fully in accord with the Markov and Manipulation theorems; were it stipulated with Seidenfeld that there is no intervention, his judgment is also that which causal decision theory ought to give.[21] The same is true of Eells if he does not view action that results from deliberation as an intervention.

Our analysis of the dispute between causal and 'evidential' decision theory does not put us neatly on either side, exactly because we think

---

[20] We are indebted to T. Seidenfeld for suggesting this reading to us.
[21] Things are a bit more complicated. Levi and Seidenfeld do not allow unconditional probabilities over one's own acts to be used in rational deliberation. (See I. Levi [1992], and J. Kadane and T. Seidenfeld [1992]). We are unclear as to what constitutes such a use, but it seems to us that our entire discussion, including the Manipulation theorem, can be carried out in terms of probabilities conditional on acts.

the dispute has been misdescribed, from Nozick on. From our perspective, whether causal or 'evidential' recommendations should be followed depends only on what one believes about the causal character of a context of decision. We agree with 'evidential' decision theorists that nothing but an ordinary calculation of the maximum expected utility is required; we agree with causal decision theorists that sometimes the relevant probabilities in the calculation are not the obvious conditional probabilities; we agree with Skyrms in avoiding an appeal to *sui generis* subjunctive relations, but we retain the subjunctive flavour of causal decision theory by showing how causal hypotheses constrain probabilities and by showing how to reproduce the causal decision theorists' calculations using nothing but (less obvious) conditional probabilities. Our suggestion is that the differences in recommendations offered by causal decision theorists and most of their critics do not result from whatever differences they may have about principles of rational choice, which is not to say they have no such differences. Where they recommend different decisions in particular cases causal decision theorists have discussed it is because they differ about whether an action is an intervention; whether the manipulated and unmanipulated distributions are different. If so, then a different event must be conditioned on than if not, and a different calculation results.[22]

Of course, the methods by which the various decision theorists reach their recommendations may differ. The method that we suggest is to use the combined causal graph, the knowledge of the direct effects of a manipulation, and the Manipulation theorem. To this extent we are causal decision theorists. But if one believes that a decision will not produce an intervention in the system, our calculation will agree with 'evidential' decision theorists; we only disagree with those critics of causal decision theory who believe that an action *is* an intervention in the sense we have described. With our method, uncertainties, and therefore disagreements in recommendations for action, can occur in many circumstances: from a lack of knowledge about the side-effects of an intervention—G_{Comb} is not known—or from a lack of knowledge about how the manipulated variables will be effected by the manipulation. Both of these epistemic problems are familiar to researchers; the usual solution is to run pilot studies (e.g. drug safety and efficacy tests in the pharmaceutical

---

[22] Whether this happy reconciliation is news is unclear. Bill Harper tells us that causal decision theorists always thought they were addressing the difference between conditioning on an event and conditioning on an intervention to bring about that event, and that causal decision theory was intended to give the ordinary results when the actions considered are not interventions. Skyrms' assessment is that while there is a correspondence between an analysis in terms of his K-partitions, probabilities of counterfactuals, and probabilities calculated conditional on interventions, causal decision theorists did not think of the matter in exactly the last way. We thank them both.

industry) to identify possible side-effects and quantitative knowledge about the intervention.

3 Inquiry and decision

Eells objects that Skyrms' proposal can seldom be applied, since we rarely know all of the most specific factors influencing an outcome, or even, Eells claims, a complete set of alternative most specific factors. Ordinary, evidential decision theory, Eells claims, gives the right answers without requiring any such special knowledge. In our view, of course, the issue is not about the proper decision theory, or about two kinds of utilities, or about dominance versus expected utility, but about the difference between computing probabilities by conditioning or an event and computing probabilities upon (or by conditioning upon) an intervention to bring about that event. What is required for the latter sort of computation? If, when causal relations are known, probabilities should satisfy the Markov condition and probabilities on appropriate interventions should satisfy the Manipulation theorem, what properties should probabilities before and after intervention have when only fragments of the relevant causal structure are known? For example, what are the probabilities on an intervention if a causal structure is known, but the probability relations are known only for a proper subset of the features? What are the probabilities on an intervention if the causal structure is unknown but probabilities among a set of variables are known? On our reconstruction of what is at issue in causal decision theory, these questions ought to be among the most urgent in the subject, and the subject ought therefore to engage some of the most difficult problems in statistics. For example, the problem of determining the probability on an intervention when the causal structure is known but the probabilities are known only for a proper subset of features contained within it is the whole of what econometricians call the 'identification problem'.

Peter Spirtes[23] has provided an important part of the answer to how to calculate the probabilities that result from an intervention when one has only very incomplete causal knowledge. When probability relations are known among a set of variables or features Spirtes has described a graphical object—a 'partially ordered inducing path graph'—that encodes all of the graphical properties that are shared by all causal structures (that is, all directed acyclic graphs) that, individually, accord with the known probability relations under the Markov condition. Using the Manipulation theorem, he has also described for such circum-

---

[23] Spirtes, Glymour and Scheines [1993].

stances sufficient conditions for the probability of an outcome upon an intervention to be calculable from the known probabilities before intervention and from the probabilities to be imposed upon a manipulated feature by an intervention. Of course, in many cases more causal knowledge than the initial probabilities and the Markov condition provide is required to calculate the probability of an outcome given an intervention. And that brings us to the difficulties described in the essay in *Science*.
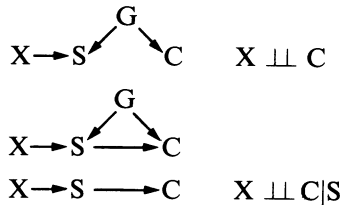
When, as in the case of public health and other public policy decisions, for all one knows some interventions may worsen circumstances, others may entail great costs but yield no benefits, while still others may markedly improve things, and no one knows which interventions have which properties, there may be considerable utility in further inquiry. One aim of such inquiry ought to be to enable us to compute the probabilities of outcomes upon interventions in accord with the two conditions we have described. Essential to that aim is providing evidence that will help us decide the causal structure behind associations among features that policy might manipulate and feature representing outcomes of interest. In applied research, the most common form of approaching the problem is to determine whether an association is due to the influence of one feature on another, or to some unrecorded common cause, or both. Only in the first case will conditioning give the probabilities appropriate for policy decisions. And, as Ronald Fisher insisted in the 1950s, one problem with research on medical records is that statisticians and epidemiologists have found and adopted no methods to distinguish any of these three cases from the others.[24] That is the chief problem alleged with research undertaken with the auspices of the AHCPR. The research sponsored by the AHCPR attempts to infer causal structure from non-experimental data, envisioning using that knowledge to guide decisions about changes in medical practices. The chief problem with research of this kind is exactly to know when—or how much of—a statistical association between a feature which policy might manipulate and a desired outcome is due entirely to an influence of the first on the second.

Fisher claimed that nothing but experimental controls can tell us whether smoking causes cancer or some third unrecorded thing causes both. As everyone knows, his opinion did not carry the day, and epidemiologists concluded that when everything plausible is measured and

---

[24] Kadane and Seidenfeld [1990] point out, correctly, that Fisher thought that observational studies of smoking did not meet Fisher's conditions for a reference class sufficiently well defined to permit tests of statistical hypotheses. We agree, but we also read Fisher, and those who in the 1960s commented on Fisher on smoking, as giving an underdetermination argument.

conditioned on and an association between smoking and lung cancer still remains, smoking causes cancer. It is worth observing that, under somewhat stronger assumptions about the connections between causality and probability, Fisher was wrong.

Suppose some feature X is measured which is known to influence cancer, if at all, only through smoking. Let S and C also be measured, but G be some unmeasured feature. The three alternative causal structures Fisher and other statisticians[25] envisioned and the associated independence relations that follow from the Markov condition are:

$$X \rightarrow S \nearrow^{G} \searrow C \qquad X \perp\!\!\!\perp C$$

$$X \rightarrow S \nearrow^{G} \searrow \rightarrow C$$

$$X \rightarrow S \longrightarrow C \qquad X \perp\!\!\!\perp C|S$$

The structures can be distinguished by the independence relations they entail among X, S, C. It is not essential even to know that factor X causes smoking if two features, X and Y, can be found that are independent but not independent conditional on smoking. The general theory of the structures that can and cannot be distinguished by the conditional independence relations they entail among a subset of features has been worked out in detail[26] and the completeness of the theory is shown in an unpublished paper by Spirtes and Verma.[27] The general theory can be used to discover structure if one assumes that all (and only) the observed independence relations result from the Markov condition applied to an (unknown) causal structure. Whether the theory is a guide in inquiry depends, for Bayesians, on their prior probabilities. Paul Teller somewhere remarked that if two hypotheses, H_1 and H_2, each entail a piece of evidence, then the ratio of the posterior probabilities of the hypotheses on the evidence equals the ratio of their prior probabilities. When X precedes Y precedes Z and X and Z are independent conditional on Y, the latter fact is entailed either by (H_1) the Markov condition and the absence of any unrecorded common cause of Y and Z, or (H_2) by the Markov condition and the existence of two or more unrecorded common causes of Y and Z that perfectly balance. If one is sufficiently close to certain of the existence of unrecorded common causes, then the prior probability of H_2, while very low, will be greater than the prior probability

---

[25] Compare K. Brownlee [1965].
[26] Spirtes, Glymour, and Scheines [1993].
[27] P. Spirtes and T. Verma [1992].

of H_1 and will remain so no matter the evidence of the conditional independence.[28]

This is not the place further to describe disputes over causal inference, but it is perhaps the place to conclude with two remarks. First (with the notable exception of Brian Skyrms), advocates of causal decision theory seem to have missed that their calculations—if not their descriptions—are part and parcel of both Bayesian and non-Bayesian statistical practice. Isaac Levi is correct in saying that causal decision theory entangles Bayesian decision theory with controversial 'causal metaphysics' that 'must seem questionable to all but rabid partisans of such views'.[29] But the entanglement already runs through and through scientific practice; nothing is gained, philosophically or scientifically, by declining to work out the principles and consequences of the connection. Second, causal decision theorists and their critics tend in each controversial case to share an assumption, the Causal Markov condition, and their differences and their agreements turn on different sensibilities about how to apply consequences of the condition to the cases. The Causal Markov condition does not obviously follow from coherence alone, or from the expected utility principle, or from axioms on prefererences. Why then are instances of the condition so ubiquitous, so uncontroversially assumed and applied in otherwise controversial cases? Does the condition follow in some *unobvious* way from principles of rationality, or is it, like John Stuart Mill's arithmetic, simply a substantive regularity so widely exemplified that we usually assume it without acknowledgment or dispute?

Acknowledgements

Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213
USA

References

Armendt, B. [1988]: 'Conditional Preference and Causal Expected Utility' in W. Harper and B. Skyrms (*eds*), *Causation in Decision: Belief, Change and Statistics*, Volume II, Kluwer.

Bartholomew, D. [1987]: *Latent Variable Models and Factor Analysis*, Oxford: Oxford University Press.

---

[28] We owe this observation to James Robins.
[29] Levi [1992], p. 244.

Brownlee, K. [1965]: 'A Review of "Smoking and Health"', *Journal of American Statistical Association*, **60**, pp. 722–39.

Eells, E. [1982]: *Rational Decision Theory and Causality*, Cambridge, Cambridge University Press.

Eells, E. and Sober, E. [1986]: 'Common Causes and Decision Theory', *Philosophy of Science*, **53**, pp. 223–45.

Fishburn, P. [1964]: *Decision Theory and Value*, New York, Wiley.

Fisher, R. [1951]: *The Design of Experiments*, Edinburgh, Oliver and Boyd.

Gibbard, A. and Harper, W. [1978]: 'Counterfactuals and Two Types of Expected Utility', in C. A. Hooker, J. J. Leach, and E. F. McClennen (*eds*), *Foundations and Applications of Decision Theory*, Dordrecht, Reidel, pp. 125–62.

Horgan, T. [1981]: 'Counterfactuals and Newcomb's Problem', *Journal of Philosophy*, **78**, pp. 331–56.

Howson, C. and Urbach, P. [1988]: *Scientific Reasonsing: The Bayesian Approach*, La Salle, IL, Open Court.

Kadane, J. and Seidenfeld, T. [1990]: 'Randomization in a Bayesian Perspective', *Journal of Statistical Planning and Inference*, **25**, pp. 329–45.

Kadane, J. and Seidenfeld, T. [1992]: 'Equilibrium, Common Knowledge, and Optimal Sequential Decisions' in C. Bicchieri and M. Chiara (*eds*), *Knowledge, Belief and Strategic Interaction*, Cambridge, Cambridge University Press, pp. 27–45.

Kiiveri, H. and Speed, T. [1982]: 'Structural Analysis of Multivariate Data: A Review' in S. Leinhardt (*ed.*), *Sociological Methodology*, San Francisco, Jossey-Bass.

Levi, I. [1985]: 'Common Causes, Smoking and Lung Cancer' in R. Campbell and L. Sowden (*eds*), *Paradoxes of Rationality and Cooperation*, UBC Press.

Levi, I. [1992]: 'Feasibility' in C. Bicchieri and M. Chiara (*eds*), *Knowledge, Belief and Strategic Interaction*, Cambridge, Cambridge University Press, pp. 1–20.

Lewis, D. [1981]: 'Causal Decision Theory', *Australasian Journal of Philosophy*, **59**, pp. 5–30.

Lindley, D. and Novick, R. [1981]: 'The Role of Exchangeability in Inference', *Annals of Statistics*, **9**, pp. 45–58.

Nozick, R. [1969]: 'Newcomb's Problem and Two Principles of Choice', in N. Rescher *et al.* (*eds*), *Essays in Honor of Carl C. Hempel*, Dordrecht, Reidel, pp. 114–46.

Papineau, D. [1989]: 'Pure, Mixed and Spurious Probabilities and Their Significance for a Reductionist Theory of Causation', in W. Salmon and P. Kitcher (*eds*), *Scientific Explanation*, Minneapolis, University of Minnesota Press.

Pratt, J. and Schlaifer, R. [1988]: 'On the Interpretation and Observation of Laws', *Journal of Econometrics*, **39**, pp. 23–52.

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

Robins, J. [1986]: 'A New Approach to Causal Inference in Morality Studies with Sustained Exposure Periods: Application to the Healthy Worker Survivor Effect', *Mathematical Modeling*, **7**, pp. 1393–2152.

Robins, J., Blevins, D., Ritter, G., and Wulfoshn, M. [1982]: 'G-Estimation of the Effect of Prophylaxis Therapy for Neumocystic Carinii Pneumonia (PCP) on the Survival of AIDS Patients', *Epidemiology*, **3**, pp. 319–36.

Rubin, D. [1977]: 'Assignment to Treatment Group on the Basis of a Covariate', *Journal of Educational Statistics*, **2**, pp. 1–26.

Simon, H. [1954]: 'Spurious Correlation: A Causal Interpretation', *Journal of American Statistical Association*, **49**, pp. 467–79.

Skyrms, B. [1980]: *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*, Yale University Press.

Skyrms, B. [1982]: 'Causal Decision Theory', *Journal of Philosophy*, **79**, pp. 695–711.

Skyrms, B. [1984]: *Pragmatics and Empiricism*, Yale University Press.

Sober, E. [1987]: 'The Principle of Common Cause', in J. Fetzer (*ed.*), Probability and Causality, Dordrecht, Reidel.

Spirtes, P., Glymour, C., and Scheines, R. [1993]: *Causation, Prediction and Search, Lectures Notes in Statistics*, New York, Springer-Verlag.

Spirtes, P. and Verna, T. [1992]: 'Equivalence of Causal Models with Latent Variables', *Philosophy, Methodology and Logic Technical Report 33*, Carnegie-Mellon University.

Stalnaker, R. [1972]: Letter to David Lewis', in W. Harper, R. Stalnaker & G. Pearce (*eds*), *Ifs*, Dordrecht, Reidel.

Yule, G. U. [1926]: 'Why Do We Sometimes Get Nonsensical Relations Between Time Series? A Study in Sampling and the Nature of Time Series', *Journal of the Royal Statistical Society*, **89**, pp. 1–64.