AGENCY AND PROBABILISTIC CAUSALITY

Huw Price

ABSTRACT: Probabilistic accounts of causality have long had trouble with 'spurious' evidential correlations. Such correlations are also central to the case for causal decision theory – the argument that evidential decision theory is inadequate to cope with certain sorts of decision problem. However, there are now several strong defences of the evidential theory. Here I present what I regard as the best defence, and apply it to the probabilistic approach to causality. I argue that provided a probabilistic theory appeals to the notions of agency and effective strategy, it can avoid the problem of spurious causes. I show that such an appeal has other advantages; and argue that it is not illegitimate, even for a causal realist.

1. INTRODUCTION

In her influential paper 'Causal laws and effective strategies'[1] Nancy Cartwright considers the relationship between two kinds of laws of nature; between causal laws, on the one hand, and statistical laws, or laws of association, on the other. She argues for two conclusions: firstly that causal laws cannot be reduced to laws of association, and secondly that causal laws cannot be done away with. Her case for the second claim is a version of the now-familar argument that orthodox Bayesian decision theory is inadequate to cope with certain sorts of decision problem, and needs therefore to be supplemented by a theory that makes explicit reference to causal judgements. As Cartwright puts it, 'causal laws cannot be done away with, for they are needed to ground the distinction between effective strategies and ineffective ones. ... [T]he difference between the two depends on the causal laws of our universe, and on nothing weaker.'[2]

This paper covers similar ground to Cartwright's, but in the opposite direction. Causal decision theory is falling on hard times. There are now a number of strong defences of the orthodox theory. Here I present a version of what I regard as the most powerful and general defence, and thus refute Cartwright's second conclusion.[3] More importantly, I argue that in doing so we go a long way towards refuting the first. Provided that we may appeal to the notions of agency and effective strategy, a probabilistic account of causation escapes the worst of the problems to which Cartwright's first claim appeals. I show that such an appeal has other striking advantages; and argue that it is not illegitimate, even for someone who is inclined to be strongly realist about causation.

There are a number of attractions in the idea that causation be analysed in terms of probability. For one thing, it seems to allow a plausible relaxation of the Humean condition that causes show constant conjunction with their effects - a relaxation arguably needed (inter alia) to cope with the possibility of indeterministic causation. For another thing it may seem to have the advantage of explicating a problematic notion in terms of a less problematic notion. The perceived character of this latter advantage will depend to some extent on one's philosophical viewpoint. Thus a metaphysical realist might be attracted by the promise of ontological economy, and of an understanding of a puzzling feature of reality in terms of a less puzzling feature. Probability might thus seem more attractive than causation as a basic constituent of the world - as a piece of metaphysical furniture. From a less robustly realist point of view, on the other hand, the prospect will seem to be that of a useful piece of conceptual analysis, and thereby the promise of an account of causation in the approved style - be it now projectivist, pragmatist, or whatever - in terms of such an account of probability.

Whatever the philosophical motivation, such an analysis would ideally appeal to the principle that an event **A** causes an event **B** if and only if it raises the probability of **B** - if and only if **B** is more probable given **A** than it would be otherwise. In its unrestricted form, however, this principle has seemed to face certain devastating counterexamples. The most serious of these are cases in which the biconditional fails from right to left: i.e., in which an event **A** raises the probability of an event **B** but does not cause **B**. There are two main kinds of such case, both having familiar counterparts in constant conjunction accounts of causation. The first turns on the fact that probabilistic dependence is in general a symmetric relation, whereas causation is asymmetric. Effects thus raise the probability of their causes (or in Humean terms are constantly conjoined with their causes), but clearly do not cause their causes. The second kind of example is really a consequence of the first: because events may be correlated with their causes, they may be correlated also with other effects of those causes. We may thus find probabilistic dependence or constant conjunction between the joint effects **A** and **B** of a common cause **C**, when neither **A** nor **B** is a cause of the other.

The usual treatment of such cases is exemplified in what is perhaps the best-known probabilistic theory of causation, that of Patrick Suppes.[4] Suppes begins by defining what he calls a prima facie cause, as follows:

> Definition 1. An event **A** is a prima facie cause of an event **B** if and only if (i) **A** occurs earlier than **B**, (ii) the conditional probability of **B** occurring when **A** occurs is greater than the unconditional probability of **B** occurring. (1984, p. 48)

This definition incorporates the central idea of the probabilistic approach to causation: that causes raise the probability of their effects. In condition (i), however, the definition already incorporates a restriction imposed by the first of the above kinds of apparent counterexample to an analysis based simply on this central idea. These counterexamples rest on the fact that effects are evidence of the occurrence of their causes - effects make it more likely that their causes have occurred. Like Hume himself, Suppes meets this difficulty by stipulating that a prima facie cause must occur earlier than its effect. He is not entirely happy with this restriction, saying that it is a simplifying assumption, eventually requiring 'more extensive and leisurely discussion' from the standpoint of discussion 'regarding the direction of time and the direction of causality.' (1970, p. 80)

We shall see that one of the advantages of introducing the notion of agency into a probabilistic account of causation is that such a condition becomes unnecessary. Agent probabilities restore the required asymmetry between cause and effect, without explicit reference to time ordering. Thus they provide an analysis of causation that allows us to contemplate such possibilities as backward causation, instantaneous action at a distance, and a causal foundation for time itself.

Even with condition (i) in place, however, Suppes feels that there is a range of cases in which 'increases the probability of' cannot be equated with 'causes'. These are cases of the second of the two kinds we distinguished above, in which correlation between two events is explained by their joint correlations with the same earlier event. Suppes calls this spurious causation:

> Definition 2. An event **A** is a spurious cause of **B** if and only if **A** is a prima facie cause of **B**, and there is a partition of events earlier than **A** such that the conditional probability of **B**, given **A** and any element of the partition, is the same as the conditional probability of **B**, given just the element of the partition. (1984, p. 50)

Suppes then proposes that we 'characterize genuine causes as prima facie causes that are not spurious.' (1984, p. 50)

There has been considerable discussion in the literature as to whether Suppes' is the appropriate definition of the notion of a spurious cause, and as to whether any such characterization can draw the boundaries of the class of genuine causes in exactly the right place (i.e., whether it can exclude all and only those cases of prima facie causation that do not meet the standards of our ordinary intuitions about causality). And even if there is a characterization that gets the boundaries in the right place, the probabilistic analysis is not out of trouble. In the paper I mentioned above, Cartwright argues that the appropriate characterization itself refers to causal relations, and hence cannot provide an analysis of causation. This is the basis of her claim that causal laws are not reducible to statistical laws. I want to show that the introduction of the notion of agency enables us to sidestep this problem. So long as we concentrate on agent probabilities, and assess these correctly, there are no spurious cases of prima facie causation of the sort that Suppes' definition is intended to exclude.

We shall be thus be concerned with judgements of probability as properly made from an agent's perspective. This use of the notion of agency may itself seem problematic. For one thing, it may seem that agency is itself a causal notion, and hence that an account of causality in terms of agent probability will necessarily be circular. For another thing, any such account may seem to ground causality in the wrong place. This objection is particularly to be expected from the metaphysical realist I mentioned earlier. The realist wants an account of the basic constituents of the physical world, and suspects that causation, whatever it is, is one of these constituents. An analysis of causation in terms of an agent-dependent notion of probability would seem to conflict with the intuition that whatever the basic constituents are, they do not depend on the (surely contingent) existence of human or other agents. (The projectivist, in contrast, expects our ordinary ways of talking about the world to be coloured by the contingencies of our position in the world.)

I shall return to these points later. I shall try to show that even for the realist there is an interesting and non-trivial explication of causation in terms of agent probability, albeit not an analysis in the strict sense. Even for a realist, therefore, the notion of agency is not out of place in a probabilistic account of causality; and the account is very much simpler for having it there.

As I noted, causal realists such as Cartwright have themselves appealed to agency and related notions in recent years, in order to argue for the objectivity of causation. Our first concern will be to show that this appeal rests on a mistake. Given the constraints of the agent's perspective, ordinary procedures of evidential reasoning can draw the distinctions they are said to be unable to draw. The distinction between effective and ineffective strategies needs probabilistic reasoning by agents, and nothing stronger. So although (as I shall explain) an emphasis on agent probability does not exclude a realist explication of causality in terms of probability, a proper understanding of agent probability does serve to undermine a popular recent argument for causal realism. It shows that the factors the realist takes to require objective causation can be adequately explained without it.[5]

2. MEDICAL NEWCOMB PROBLEMS.

Some of the most striking cases that have been thought to exhibit spurious prima facie probabilistic causality are those involved in the 'medical' Newcomb problems. These decision problems, like the related original Newcomb problems and Prisoners' Dilemmas, have been claimed to show the inadequacy of orthodox evidential or Bayesian decision theory. That is, they have been said to provide cases in which the evidential theory recommends actions clearly contrary to our common sense intuitions - and hence to illustrate the need for a decision theory grounded on causal judgements. In making this point, the 'medical' examples have two considerable advantages over Newcomb problems of other kinds: they exist (indeed they are common), and virtually everyone agrees on the 'right' decision in such a case (on what a decision theory ought to prescribe, so to speak).

Our present interest in these cases stems from two factors. The first is that as objections to evidential decision theory, medical Newcomb problems depend on the claim that spurious correlations translate into spurious evidential dependencies between contemplated actions and other events. It is these evidential dependencies that are supposed to lead the evidential decision theorist astray. As I have said, however, I want to show that one of the advantages of introducing agency into a probabilistic theory of causation is that spurious correlations disappear from an agent's perspective (and thus don't have to be dealt with by a possibly vicious restriction on the general principle that only causes increase probabilities). Thus in these medical Newcomb cases I want to show that the objectors are wrong: from the agent's perspective there are no spuriously-grounded evidential dependencies of the kind that would mislead a Bayesian agent.

That's the first reason for looking at these cases - if agent probabilities are to handle spurious causes in general then they must do so here (contrary to the still-common view that they do not). The second reason is that these cases are really the only cases we need consider. As I'll explain later on, the agent's perspective makes any case of spurious causation a medical Newcomb problem. In dealing with these cases we thus deal with the general problem.

Let me describe a typical medical Newcomb problem. It has long been recognized that in people susceptible to migraine, the onset of an attack tends to follow the consumption of certain foods, including chocolate and red wine. It has usually been assumed that these foods are causal factors, in some way triggering attacks. This belief has been the source of much mental and physical anguish for those susceptible both to migraines and to the attractions of these substances. Recently however an alternative theory has come to light. It has been discovered that eating chocolate is not a cause of migraine, but a joint effect of some pre-migrainous state (or 'PMS', as we doctors say). The physiological changes that comprise PMS thus typically increase a subject's desire for chocolate, as well as leading, later, to the usual physical symptoms of migraine. Clearly this is good news for a migraine-afflicted chocaholic. There is now no point in giving up chocolate in order to try to avoid the greater deprivations of the migraine itself.

Or so intuition tells us. But advocates of causal decision theory argue that evidential decision theory still recommends that the migraine sufferer decline chocolate. The argument goes like this. Call such a person 'Coco', and suppose that he finds himself tempted by a Mars Bar. Coco knows that in his case, in general, eating chocolate is positively correlated with PMS. Hence such behaviour provides positive evidence that he is in PMS, and Coco should therefore reason that if he accepts a Mars Bar, it will be more likely than otherwise that he is already in PMS. Given his strong preference not to be in PMS, evidential decision theory will then recommend that he decline the Mars Bar. This will be the choice that maximizes expected utility. Causal decision theorists rightly point out that this advice would be foolish. Whether Coco is in PMS or not, and whether that is good or bad from his point of view, accepting the Mars Bar won't make any difference. Its evidential bearing ought to be counted irrelevant to his decision.

I want to show, however, that rational evidential reasoning cannot work in the way it would have to work, in order to establish a contrast between the advice offered by evidential and non-evidential decision theories. Thus in a case such as this, only someone whose evidential reasoning was already irrational would find their evidential decision rules leading them astray. Hence their irrational behaviour would be traceable to their defective evidential reasoning, rather than to their subscription to the wrong decision rule.

The essential point is that in the context of the kind of decision problem in question, the relevant spuriously-grounded evidential judgement is self-defeating. To see this, consider Coco again. We are told that his case illustrates that evidential decision theory 'commends an irrational policy of managing the news so as to get good news about matters which you have no control over'.[6] But can Coco really manage the news? Suppose that he attempts to follow that 'irrational policy'. Knowing that PMS inclines him to chocolate, he judges that declining a Mars Bar would indeed be evidence that he is (probably) not in PMS. Hence he does decline, hoping thereby to come by this good news. Does he succeed?

It depends on how good he is at self-deception. He has to overlook the fact that his choice to decline is fully explained by the very judgement in question (namely the judgement that if he were to decline, that would constitute evidence that he is not in PMS). This judgement has provided him with a reason to decline which is quite independent of whether he is in PMS. Given this judgement, anyone with the same background beliefs and desires would make the same choice, regardless of whether he or she was in PMS. Given that Coco has made this judgement, in other words, he now has no grounds for taking it to be true. Remember that it rested on his belief about the correlation between PMS and chocolate consumption; and its effect is to destroy that correlation. In similarly motivated agents who make this judgement, PMS is simply irrelevant to their resulting decision to choose chocolate.

The policy of managing the news thus does require that Coco be systematically irrational, but not in the sense that he subscribe to the wrong decision theory. Given that he does subscribe to the evidential theory, the policy only works if he can ignore the proper effects of new beliefs on the grounds for old beliefs (and hence the fact that judgements may be self-undermining).

Here is a case which is in some ways analogous. Consider Scrooge, who has never received a Xmas present. He has heard of the pleasure of unexpected gifts, and as the festive season approaches would dearly love to experience that pleasure for himself. Now the crucial thing about an unexpected gift is that its prior probability be low. However, Scrooge reasons that since he has never been given a Xmas present, any present he receives will have low prior probability. He concludes that he can send himself a little token of his self-esteem, safe in the knowledge that when it arrives it will be completely unexpected.

Obviously the trick only works if Scrooge is able to ignore certain relevant information which is at his disposal. To find his self-sent present unexpected, he must ignore what he knows about its history - what he knows in virtue of which, unlike presents in general, it does not have low prior probability. Scrooge's contemplated action is attractive in the light of the belief that the gift will be unexpected when he receives it; but to perform the action would be to undermine that belief.

In the next section I want to go through the case of Coco in more detail. In particular I want to show that it survives two kinds of objection. One of these is to try to modify the case so that the relevant evidential judgement becomes non-self-defeating - more on this in section 4. The other is to object that my account of how Coco should assess the relevant probabilities relies on his causal beliefs. For example, it is because the judgement concerned is (or would be) a cause of his decision to eat chocolate that the evidence on which it would be based is no longer applicable. Doesn't this involve some sort of circularity?

This objection needs to be handled with some care. True, it is no use to the advocates of causal decision theory or to others who hold that our causal judgements and our evidential probability judgements come apart in cases such as this. Against such an opponent the present argument can be phrased as a reductio: assuming these two kinds of judgement are conceptually distinct, the argument shows that in the medical Newcomb cases they nevertheless coincide. We thus defeat the case that is said to show that these two kinds of judgement must be conceptually distinct.

The problem that remains is this: if causation is really to be analysed in terms of evidential probability then as ordinary speakers and reasoners we will lack the very conceptual distinction that gets evidential decision theory off the hook in cases such as this. Without the conceptual distinction, how do we get it right? I shall come back to this problem. First then to the easier task of refuting the usual case for causal decision theory (and hence for the need for restrictions to cope with spurious causation).

3. COCO'S CASE IN MORE DETAIL.[7]

At the heart of the problem is the issue of the applicability of statistical generalizations to individual cases. Coco believes something like this:

(1a) I choose to eat chocolate more often when I am in PMS than when I am not in PMS;

or perhaps like this:

(1b) I am more likely to choose to eat chocolate when I am in PMS than when I am not in PMS.

Either way, what he believes is a generalization, and the frequency formulation (1a) makes it harder for us to ignore this crucial fact. On the occasion in question, the issue is whether (1a) or (1b) provides Coco with grounds for the judgement

(2a) If I eat this Mars Bar, then it will be more probable than it would otherwise be that I am in PMS.

We don't want to exclude evidential probabilities by fiat, so let us also include

(2b) If I eat this Mars Bar, that will be (positive) evidence that I am in PMS.

Why does it matter whether (1a) and/or (1b) provide grounds for (2a) and/or (2b)? Because (2a), (2b) or something similar is the judgement that would incline Coco, if he follows an evidential decision theory, to do what we have agreed would be irrational: to decline the Mars Bar. If (1a) and/or (1b) do not license (2a) and (2b) - if indeed we can show that in Coco's circumstances, such an inferential step would be irrational - then we would have an alternative explanation of irrationality of that behaviour. It would be attributable not to an inappropriate decision principle, but to an inappropriate statistical inference.

The argument that Coco is not entitled to (2a) or (2b) on the basis of (1a) or (1b) turns on two assumptions. The first concerns Coco's view of the causal connection between PMS and his choice behaviour. Roughly, we need to assume that he believes that PMS is a cause rather than an effect of choosing to eat chocolate. As I have explained, this does not beg the question against causal decision theory, for of course this is an assumption that the causal decision theorist shares. (After all, if Coco believes that eating chocolate could cause him to be (already) in PMS, then the causal and evidential theories agree that from his point of view it is better to abstain.) We are entitled to assume what causal decision theory accepts, in order to show that there is a fallacy in the argument on which it is standardly taken to rest.

The second assumption we need is that in general Coco's probabilistic judgements do play the rôle in his decision behaviour that the evidential theorist claims they do (or should) - and that Coco has some reflective awareness of this, to the extent of recognizing that in a given case, a particular probabilistic judgement is or would be among his reasons for acting in a certain way.[8]

Given these assumptions, we can show that in the imagined case Coco cannot reasonably infer (2a) or (2b) from (1a) or (1b). (1a) and (1b) are statistical generalizations. They tell Coco at best that (for him) eating chocolate is normally evidence of PMS. He has the problem that faces anyone who would reason by instantiation from a statistical generalization: he has to decide whether the instance in question would be 'normal', in the required sense. There are complexities here, but one thing is clear. Such an instantiation is blocked or undermined by the information that the case in question is in some way exceptional, in such a way as to fall within the scope of some conflicting generalization.

This is what happens in Coco's case, though with one extra twist. He recognizes not that the instantive inference from (1a)/(1b) to (2a)/(2b) is blocked by some conflicting generalization, but that it would be undermined, were he in fact to accept (2a)/(2b). For by the first assumption, he believes that (1a)/(1b) holds in virtue of the fact that chocolate consumption is typically caused by PMS. What matters is therefore the causal history of the contemplated consumption. Does Coco know anything about (what would be) the causal history of that action that prevents him from instantiating (1a)/(1b)? Not directly, it seems. However, he sees that if he were to infer that (2a)/(2b) - to accept that eating the Mars Bar would be evidence that he is in PMS - he would then know something further about the causes of his ensuing decision to decline. Given his decision principles and background motivations (his reasons, after all, for being interested in (2a)/(2b) in the first place), he can see that in this case, his Mars Bar avoidance would be attributable to his acceptance of (2a)/(2b). He would thus have an alternative causal explanation of his decision to decline - an explanation that makes no mention of PMS. He has no reason to think that in cases in which he has such a motivation for eating or declining chocolate, there is any correlation between Mars Bars and PMS.

4. FIRST OBJECTION: CAN THE EXAMPLE BE STRENGTHENED?

The causal decision theorist might respond at this point by attempting to strengthen Coco's statistical beliefs, in order to counter the self-undermining character of his judgement that (2a)/(2b) holds. Perhaps Coco's beliefs are more detailed, and actually cover the kind of case in which his immediate reasons for acting include a belief of the form of (2a) or (2b). He believes not only (1a)/(1b) but also

(3) The positive correlation between my eating chocolate and my being (already) in PMS survives in the sub-class of cases in which my decision to eat (or not) is

influenced by a belief of the form of (2a) or (2b), grounded in my awareness of the correlation described in (1a) and (1b).

As it stands, however, this simply delays the inevitable. Coco considers inferring that (2a)/(2b) on the basis of (1a)/(1b). He sees that were he to do so, (3) would then apply. This confirms his judgement that (2a)/(2b). He has the same judgement, but it is based on different evidence. He has located himself, as it were, in a sub-class of the class which formed the basis of his initial judgement that (2a)/(2b). But in virtue of applying (3) to his present case he now finds himself in a sub-class of that sub-class. (3) tells him nothing about cases in which accepts that (2a)/(2b) on the basis of (3) itself. He thus finds himself with a belief which in the circumstances would lead him to abstain, and with no reason to invoke his not being in PMS in order to explain that abstension. As before, it is a belief that undermines the statistical instantiation on which it itself would have been grounded.

To avoid this kind of argument altogether, I think the causal decision theorist must credit Coco with a belief in a correlation between chocolate-taking and pre-existing PMS that is stable under a range of assumptions about his reasons for acting. Coco must believe something like

(4) Whatever my reasons for consuming or not consuming chocolate, there is a positive correlation between my doing so and pre-existing PMS.

A belief of this form will remain applicable to Coco's case - remain a valid basis for inference by instantiation - whatever he comes to believe about the reasons for a contemplated action (of the kind in question). However, I think that in crediting Coco with a belief of this kind, the causal decision theorist throws the baby out with the bath water. For if Coco believes (4), he has grounds for believing that a pre-existing PMS can be a (probabilistic) effect of chocolate eating - and hence that declining a Mars Bar is an effective strategy for bringing it about that he is not already in PMS.

One way to show that (4) leads to this conclusion is to note that if it were true, it would seem to enable us to influence Coco's physiological state by providing him with sufficient motivation either to accept or to decline a Mars Bar. If this seems implausible, consider the limiting case: suppose Coco believes (and we concur) that under certain background conditions (not themselves dependent on his motivations) it is true that whenever he eats chocolate he turns out to be in the early stages of PMS. Suppose he also believes (and we again concur) that on such occasions he has the relevant freedom of choice: it is up to him whether he eats, and his decision can be expected to turn as usual on his relevant beliefs and desires. Suppose finally that we want to ensure that he is not in PMS. It is enough (we should believe) to offer him a Mars Bar, while at the same time providing a sufficient motive to ensure that he declines.

Of course, it is hard to imagine ourselves (or any rational person) actually accepting the beliefs that we have here supposed that we share with Coco. But this implausibility is that of (4) itself (in conjunction with the assumption of free action). Because to accept (4) is to accept the possibility of 'backward' causation, (4) conflicts with the principles that normally rule out such causation. In particular, it conflicts with the principle that it is always possible (in theory, at any rate) to find out whether the relevant earlier event (the supposed effect) has taken place, before one settles on a later action (the supposed cause). For if Coco believes

(5) In the relevant background circumstances, it is possible for me to find out whether I am in PMS, before I decide whether to eat chocolate

then he can design an experiment to refute (4). All he has to imagine is that in a randomly selected range of cases of the relevant kind, he should follow the policy of eating chocolate when and only when he has already discovered that he is not in PMS. He is thus able to generate experimental data that is guaranteed to conflict (to any desired degree of certainty) with the statistical claim embodied in (4). (In the limiting case refutation is much easier, of course.) Note the importance of the 'irrelevance of reasons' aspect of (4). If the correlation concerned holds whatever Coco's reasons for acting, then it holds in the sub-class of cases in which he is motivated by the desire to refute (4).

As readers may have noticed, I have here drawn on Michael Dummett's analysis of the conditions under which, without inconsistency, we might claim to be able to bring about past events.[9] Dummett shows that we can accomodate a belief in backward influence, so long as we are prepared to give up the assumption that before we decide how to act, it is possible for us to find out whether the past event in question has already occurred. The importance of this assumption turns on its rôle in 'causal loop' arguments of the above kind. In showing that to believe (4) is to reject (5) - the relevant instance of this general assumption - we have shown that is open to someone who accepts (4) to interpret it in terms of an ability to affect a pre-existing state of affairs.

As a result, the causal decision theorist finds in (4) no answer to our earlier objection. We can agree that if Coco accepts (4), evidential decision theory will recommend that he decline a Mars Bar (in order make it less probable that he is in PMS). But we can add that this is now the right recommendation. Whether a causal decision theory agrees will depend on the notion of cause that theory invokes. If it allows for backward causation (and accepts that (4) involves a case of it), then it will agree that Coco should decline. If it rejects backward causation (though still maintaining that (4) is a coherent belief) then it will disagree, recommending that Coco accept the Mars Bar. In this latter case, however, I maintain that it is the causal decision theorist who is in trouble. The example shows that with such a notion of causation, it is the causal decision principle that sometimes gives the wrong results. I don't think I can do more than simply to maintain this. If someone's intuitions went the other way, I think that short of trying to improve their understanding of causation we could do little to dissuade them.

For present purposes, however, the important point is that to rest a medical Newcomb problem on a belief such as (4) is to deprive it of its most telling ingredient: the obvious absurdity of acting so as to influence a pre-existing physiological state. Given (4), or something like it, such a course is no longer absurd. Intuition begins to fail us here; but with it fails the causal theorist's case against the evidential theory.

5. SECOND OBJECTION: DOES COCO NEED PRIOR CAUSAL BELIEFS?

There is an air of unreality about the convoluted process of hypothetical reasoning by means of which Coco establishes that in view of what he knows about the causal structure of the case, his decision as to whether to eat is probabilistically irrelevant to whether he is in PMS. 'Surely ordinary speakers don't go through that,' one might object. Indeed one might, and justly so, I think. I want to emphasize, however, that what we have described need not be considered a normal episode of probabilistic reasoning. It flows from the assumption that causality and agent probability are not intrinsically linked - that they can and do come apart in such cases. Having refuted the standard argument for that assumption, we are now in a position to suggest a much simpler story about Coco's case. We can suggest that in believing that PMS is not an effect of chocolate consumption, he already believes that accepting the Mars Bar is probabilistically irrelevant to whether he is in PMS. Whatever leads him to the one belief leads him also to the other, and there is no need for the convolutions described above.[10]

This will be the case, in particular, if beliefs about agent probability are constitutive of causal beliefs. In dealing with the medical Newcomb problems I think we have removed the main obstacle to such a view. That obstacle was the problem of spurious causes - the existence of cases in which an event **A** seems to be (positively) probabilistically relevant to an event **B**, without being a cause of **B**. We now see that these supposed counterexamples do not survive the move to agency probability. If we think of **A** as a contemplated action, then we have the basis of a medical Newcomb problem. To make it appear 'problematic' - to produce a case in which evidential decision theory appears to yield the wrong prescription for action - we need only add an appropriate assignment of subjective utilities to **A** and **B**. However, we have seen that the problematic appearance of such cases rests on a fallacy of probabilistic reasoning. From the agent's point of view probabilistic relevance and causal relevance cannot diverge. To introduce the agent is in effect to assume an independent causal history to the event **A**. Those probabilistic correlations that survive this assumption seem to have claim to be counted as genuine effects of **A**.

6. CAUSE AND TEMPORAL ORDER.

I said at the beginning that the problem of spurious causes was one of two difficulties commonly felt to afflict probabilistic accounts of causation (as indeed to afflict their ancestor, the constant conjunction account). The other problem was that in contrast to causation, probabilistic relevance seems a symmetric relation. This is usually tackled, as it is by Suppes and was by Hume, by specifying that causes occur earlier than their effects. To see that the introduction of agency also removes the need for this restriction, consider

(6) An event **A** is a cause of a distinct event **B** if and only if ensuring that **A** rather than not-**A** would be an effective means-end strategy for a free agent whose overriding desire is that it should be the case that **B** (and whose concern is thus to act so as to maximise the probability that **B**).

In the light of our discussion of Coco's case, it would clearly be inappropriate to add to (6) the requirement that **A** occur before **B**. But we need to show that without this condition, (6) nevertheless guarantees the asymmetry of the cause-effect relation.

We need to confirm that no agent could coherently regard it as possible both (i) to raise the probability of **B** by ensuring that **A** rather than not-**A**; and (ii) to raise the probability of **A** by ensuring that **B** rather than not-**B**. The argument is simple, however. To accept (i) is to regard **A** and not-**A** as alternative outcomes of a conceivable free action. To accept (ii) is similarly to regard **B** and not-**B** as conceivable choices. To regard (i) and (ii) as compatible is thus to regard (**A** and **B**), (**A** and not-**B**), (not-**A** and **B**) and (not-**A** and not-**B**) as alternative outcomes of a free compound action. Given this much, however, it is easy to design an experiment to refute the claimed correlation between **A** and **B**: one simply needs to add sufficient motivation for choosing (**A** and not-**B**) and (not-**A** and **B**) over the other two alternatives.

It might be objected that an agent could accept (i) and (ii) without accepting that the actions of choosing between **A** and not-**A** and of choosing between **B** and not-**B** are compatible. Perhaps the background conditions for the former action exclude those for the latter, although both are individually conceivable. However, to admit this is simply to admit that under certain general descriptions of kinds of events, the causal connection between events of two kinds may go in one direction in some cases and in the other direction in others. Smoking causes cancer, but cancer also causes smoking (by distraught victims and their friends, for example). The asymmetry of causation is not besmirched by this fact.

So the agent's perspective guarantees causal asymmetry, and has the consequent advantage of not excluding prematurely the possibilities of simultaneous and backward causation. Moreover, the point provides a new argument against the possibility of spurious causation. Because (6) guarantees causal asymmetry, it respects the 'directedness' of the causal links in any claimed case of spurious causation. In other words, it respects ordinary intuitions, and cannot find a probabilistic relation that 'backtracks' across a causal link. The remaining possibility we need to deal with is that (6) might find a causal connection not in the wrong direction but where intuition finds no causal link at all. For example, could (6) yield a spurious causal link between separate effects **A** and **B** of a common cause **C**? Could we coherently set out to do **B** in order to make it probable that **A**? If so, this will depend on the existence of the correlations between **B** and **C** and between **C** and **A**. However, in the circumstances these correlations assume we cannot choose to do **B**, except by doing **C** (otherwise there would be a conflict with the assumption that **C** is a cause of **B**, as interpreted by (6)). Of course in other circumstances we might do **B** directly. But then we would have no reason to regard **A** (or **C**) as more probable as a result.

In general then, (6) prevents us from discovering spurious causes in statistical correlations between effects and either their causes or other effects of those causes. To extract a causal claim from (6) we have to treat the supposed cause as a free action, and this prevents us from drawing on any statistical correlation with the usual causes of events of that kind.

7. DOES IT MATTER THAT AGENCY IS A CAUSAL NOTION?

Now to the question I deferred at the beginning: does it matter that a probabilistic analysis of causation should have to rely on the notion of agency? For a start, we see that the metaphysical realist can no longer argue that we need a realist notion of causation in order to make sense of our behaviour as agents in cases such as the medical Newcomb problems. We have seen that that can all be explained probabilistically, given the constraints of agency. However, the realist might counterattack in two ways. She might say that in invoking agency we simply cut ourselves off from the project of a realist (and naturalistic) account of causation, for agency is clearly not a particularly fundamental feature of the natural world. And she might say that the offered analysis is in any case circular, since agency is a causal notion. I want to take these objections in reverse order.

We have seen that '**A** causes **B**' can plausibly be identified with '**A** raises the probability of **B**', provided that we suitably specify the background against which the probability in question is to be assessed. It is to be assessed under the assumption that **A** is (a product of) a contemplated action of the assessor concerned. Why does the identification go through under this restriction? Because the specification brings with it a particular causal story about the origins of **A**. This serves to exclude spurious causes. Those probabilistic correlations that survive the assumption that **A** is a product of a free action have a good case to be taken as grounded in genuine effects of **A**.

Could we have run the same trick with some other privileged account of the origins of **A**? On the face of it, yes. We could have restricted our probabilities to those that would obtain under the assumption that **A** occurs by divine intervention, or as a result of a one-off collision with another universe, or perhaps in virtue of a quantum fluctuation of infinitesimal probability. It seems that under any of these restrictions the surviving probabilistic correlations will be those that we want to count as genuinely causal. However, it also seems that if we hope to analyse causation in terms of probability then none of these restriction are admissable. For they don't enable talk of causes to drop out in favour of talk about probabilities. Rather, they enable talk of causes to drop out in favour of talk of probabilities <u>in certain causally-specified circumstances</u>. We have shifted the reference to causes, but we haven't eliminated it. It is something like proposing to analyse the property of being red as that of giving off the same part of the electromagnetic spectrum as a stop-light. We don't solve the problem of the physical constitution of redness; we simply reduce it to the problem of the constitution of the colour properties of stop-lights.

Is the above appeal to agency of this kind (as the second of the realist's objections suggests)? I think the difference turns on the fact that the agent's perspective is something we all have - something that may thus be considered prior to the analytic task of understanding causation. From this point there are a number of ways of going on. One would be to think of causation by analogy with the secondary qualities. We might say that causal relations are constituted by probabilistic relations, but that the relevant latter relations are mind (or more particularly agent) dependent, just as secondary qualities are sensory agent dependent. Such a view might seem an attractive elaboration of a Humean projectivism about causation. Causal judgements would be viewed as equivalent to certain sorts of probabilistic judgements, these themselves being expressions of credences on the usual lines.

Another approach, more likely to appeal to a metaphysical realist, would be to take the account of causation in terms of agent probability as a characterization rather than an analysis. The line would run something like this.[11] There are objective causal relations in the world. As agents in the world, we are capable of exploiting these relations to further our ends. Indeed, our knowledge of causal relations derives from this fact. We discover and characterize causal relations in virtue of their relevance to the decisions we face as agents. Thus it is true that for **A** to cause **B** is for **A** to make **B** more probable than otherwise from an agent's perspective. This is not what constitutes causation, for agency is not a fundamental constituent of the world; but it is what makes causation accessible and important for agents, given that there are some.

Both stories make sense of the fact that agency plays a privileged rôle in the suggested account of cause in terms of probability. Both allow that appealing to agency is not in the same boat as appealing to miracles, divine intervention and the like. What puts it in a different boat is the integral rôle of agency in our experience of our place in the world.

We all face the world in two ways: as players as well as spectators, participants as well as observers. Given that we all have both perspectives, they may interact. The two accounts just sketched illustrate two ways this interaction might go. In the first we project onto (what we take to be) the observed world certain products of our rôle as participants. In the second our epistemological access to certain things in the observed world depends on our ability to be participants in the world; we know causes by knowing what it is like to be an agent.

Thus the ubiquity of the agent perspective provides an answer to the second of the realist objections with which we began this section. Explicating causality in terms of agency is not circular, because we don't need an explication of agency in terms of causality. Agency is something of which we all have direct experience.

The first objection does a little better. It is true, I think, that an account of causation in terms of agent probability cannot be a realist analysis - not at any rate unless we are prepared to concede that agency is much more fundamental or causation much less fundamental than most of us are inclined to assume. However, to abandon analysis is not necessarily to abandon the project of an illuminating account of causation in terms of probability. We sketched one sort of realist alternative, the view that agent probabilities provide our mode of access to real causal relations. I conclude that even for the realist it is a significant result that by appealing to agency, the connection between causing and making probable can be made a good deal simpler than most people have thought.

To sum up: Cartwright argues (i) that causal laws cannot be reduced to laws of association, because of the problem of spurious causes; and (ii) that causal laws cannot be eliminated, because they are needed to ground the distinction between effective and ineffective strategies in Newcomb problems. In refuting (ii) we have found the means to refute (i). Agency screens off the spurious associations of a contemplated action. This means not only that there is no need for a distinctively causal decision theory, but also that we may characterize causal regularities as associative regularities that continue to hold from the free agent's distinctive point of view.

8. APPENDIX: SINGLE-CASE PROBLEMS FOR PROBABILISTIC CAUSALITY.

We have shown that an account of causation in terms of agent probability avoids two of the major problems that have plagued earlier probabilistic theories of causality (as indeed their Humean ancestors). It guarantees the asymmetry of the cause/effect relation, and it is not troubled by spurious causation. However, these are not the only cases in which probabilistic judgement and causal judgement have seemed to come apart. For one thing, probabilistic theories have been charged with erring on the other side: with sometimes missing causal connections, or worse still with finding a negative cause where there is actually a positive cause. That is, it has been claimed that there are cases in which we ought to say that an event **A** causes an event **B**, even though it is not the case that P(**B**/**A**)>P(**B**/~**A**) - even though P(**B**/**A**)<P(**B**/~**A**), in some cases. Another problem is that of pre-emptive causation: surely we may raise the probability of an event, only to find that it occurs as an effect of something other than our action.

I have little to say about these cases, for I think that an emphasis on agent probability adds little to existing ways of dealing with such objections. However, I think that they do have some bearing on the question of the legitimacy of appealing to the notion of agency in explicating causation. For they show us that apparently problematic features of judgements about agency are also features of judgements about causation. To save time I shall concentrate on the kind of case that has been judged most problematic for the probabilistic approach, in virtue of involving intrinsically indeterministic physical processes.

Imagine we are trying to rescue an unfortunate cat, whose life is endangered in one of the notorious experiments of the evil Professor Sch... (you know who). In this particular version the poor animal is locked in a lead box with a geiger counter and a radioactive source, and will be electrocuted when the geiger counter fires. Rushing into the laboratory, we open the box to let the cat out. Regrettably, however, in opening the lid we admit to the box a stray cosmic ray. The geiger counter triggers and the cat is snatched from the jaws of life. It seems that we caused the cat's death by opening the box, even though his chances of survival were obviously very much better if we did open the box than if we didn't.

I think the apparent conflict stems from the fact that there are many ways of characterizing a given action, only some of which will be available to an agent at any given time. In cases such as these our causal and probabilistic intuitions rely on different characterizations, and hence seem to conflict. The causal intuition relies on what we have supposed that we know after the event: that we opened the box in such a way as to allow a cosmic ray to enter the geiger counter. But of course this wasn't how we thought of the action in advance. We thought of it simply as opening the box. The two characterizations give rise to quite different assessments of the cat's chances of survival, conditional on our performing the action in question. Simply opening the box raises the probability of survival (compared to not doing so); but opening the box in such a way as to ... lowers the probability of survival (again compared to not doing so). In making causal judgements after the fact in such a case, we allow ourself to appeal to information that was not and perhaps could not have been available to an agent beforehand. If we are less inclined to do this with straightforwardly probabilistic judgements it may be because we are often interested in assessing an agent's reasons for acting in a certain way. We cannot criticize someone for ignoring what they could not reasonably have known. On occasion, however, we do assess probabilities in this retrospective way. Thus it would be natural to say that given that in fact there was a cosmic ray shower in progress at the time, opening the box reduced the cat's chances of survival. Conversely, I think we are sometimes reluctant to base causal judgements on the retrospective perspective. Accused by Professor Sch... of causing the death of his beloved Katze, we are inclined to deny responsibity. What we did, we protest, was to immeasurably improve the poor cat's prospects. If he died in spite of that, we cannot be made to take the blame.

These conflicting intuitions are forced in extreme cases, in which the relevant probabilities arise entirely from an indeterministic physical process. Suppose now for example that we cannot open the box in time to prevent this experiment going ahead. All we can do is change the setting of the control that determines what hourly click counts from the geiger counter will prove deadly for the cat. Knowing the expected rate of decay of the radioactive source, we can choose the most unlikely setting from those available. We choose the interval 0-10, knowing that there is a better than 99% chance that the source will produce more than ten clicks in a given hour. Unhappily the unlikely happens, and the cat dies.

Did we cause the cat's death? Certainly we are inclined to blame ourselves, to say that if only we had chosen differently the cat would still be alive. But on the other hand we feel that we did as much as possible to ensure that the cat would live - that Fate, if anything, should be held responsible for the fact that we failed. These conflicting intuitions again seem to correlate with alternative ways of describing our action. Do we need to force the issue, to insist that the conflict be resolved in one way or the other? A realist may feel bound to say so. Even for the realist, however, there seems little reason to insist on resolving the conflict in favour of the view that we did cause the cat's death.

As I said, I think an emphasis on agent probability adds little to existing ways of dealing with these sorts of problems for probabilistic theories of causality. The sorts of solutions here sketched are available to other probabilistic approaches as well. The present relevance of these cases lies in the fact that they draw attention to a feature of causal judgement that might otherwise be found objectionable in an account in terms of agent probability: namely the dependence of single case causal judgements on our modes of description of the events in question. As a feature of causal judgement however construed, such dependence is no weakness of an account in terms of agent probability.[12]

NOTES

[1] Noûs, **13**(1979). Reprinted as chapter 1 in How the Laws of Physics Lie, Oxford: Oxford University Press, 1983. Page references here are to the latter version.

[2] 'Causal laws and effective strategies', p. 22.

[3] My defence of the orthodox theory, on which the version below draws, appears in 'Against causal decision theory', Synthese, **67**(1986), 195-212. For a sympathetic survey of the other main defences see Paul Horwich, Asymmetries in Time, Cambridge, Mass.; MIT Press, 1987, chapter 11.

[4] A Probabilisitic Theory of Causality, Amsterdam: North-Holland, 1970. I draw here on the less formal exposition in Probabilistic Metaphysics, Oxford: Basil Blackwell, 1984. I have altered Suppes' notation slightly, switching the variables **A** and **B** to conform to my usage elsewhere in this paper.

[5] The same argument counts against those who have taken the decision problems concerned to require objective non-epistemic probabilities, propensities and the like.

[6] David Lewis, 'Causal decision theory', Australasian Journal of Philosophy, **59**(1981), 5-30; at p. 5.

[7] The following is a version of the argument offered in 'Against causal decision theory', Synthese, **67**(1986), 195-212. In that paper I argue also that in so far as it needs to, this defence of the evidential theory extends to standard Newcomb problems and to the Prisoner's Dilemma.

[8] As it happens this is not an unrealistic requirement. We are not crediting Coco with exceptional powers of insight or inference. But even if the required powers were exceptional, that wouldn't invalidate the present argument. We are not after a reconstruction of a common piece of reasoning, but rather a refutation of a philosophical claim about how ordinary reasoning ought rationally to proceed. We want to show that it follows from assumptions that the causal decision theorist accepts that it would in fact be irrational to reason in that manner.

[9] 'Bringing about the past', Philososophical Review, **73**(1964), 338-59.

[10] The same implausible complexity has seemed an objection to other defences of evidential decision theory. For example, David Lewis objects to the Tickle Defence that although it 'does establish that a Newcomb problem cannot arise for a fully rational agent, ... decision theory should not be limited to apply only to the fully rational agent.' (1981, p. 10) I am suggesting that in virtue of the conceptual connections between causality and agent probability, ordinary agents don't need any such defence, and therefore don't need to be rational enough to use it. In knowing the causal story they already know the relevant evidential probabilities. The Tickle Defence, like my alternative argument, should be considered a philosophical solution to a philosophical problem.

[11] I am extrapolating here from a suggestion made in correspondence by Hugh Mellor. Mellor offers a view of much this kind in 'On raising the chances of effects', James H. Fetzer, ed., Probability and Causality, Dortrecht: Reidel, 1988, 229-39.

[12] I am grateful to David Braddon-Mitchell, Barbara Davidson, Andre Fuhrmann, Frank Jackson, Fred Kroon, Hugh Mellor, Peter Menzies and Philip Pettit for comments and conversations on this material.