S. L. HURLEY

NEWCOMB'S PROBLEM, PRISONERS' DILEMMA, AND COLLECTIVE ACTION*

ABSTRACT. Among various cases that equally admit of evidentialist reasoning, the supposedly evidentialist solution has varying degrees of intuitive attractiveness. I suggest that cooperative reasoning may account for the appeal of apparently evidentialist behavior in the cases in which it is intuitively attractive, while the inapplicability of cooperative reasoning may account for the unattractiveness of evidentialist behaviour in other cases. A collective causal power with respect to agreed outcomes, not evidentialist reasoning, makes cooperation attractive in the Prisoners' Dilemma. And a natural though unwarranted assumption of such a power may account for the intuitive appeal of the one-box response in Newcomb's Problem.

1. NEWCOMB'S PROBLEM AND COMMON CAUSES

In Newcomb's Problem, I am offered the choice of taking either just the opaque box in front of me, or both it and another transparent box, in which I can see $1,000. I am told, and believe, that a predictor of human behaviour has already put either nothing or $1,000,000 into the opaque box. He has done so on the basis of whether he predicts that I will take both boxes, or only one. That is, he has put $1,000,000 in the opaque box if and only if he has predicted that I will take just it, and not the other as well. Concerning 99% (or some very high percentage) of the many other cases of choices made by persons in the same circumstances as myself, the predictor has predicted correctly.[1]

Decision theorists split over whether it is rational to take just the opaque box, or both. Causal decision theory favours taking both boxes, basically because, regardless of what I do, the predictor has already either put the million in the opaque box or he has not, and either way I am better off taking the extra thousand, so taking both boxes dominates taking just the opaque box. Evidential decision theory, at least as it was originally conceived, favoured taking only one box, since, granted the very high probability of correct prediction given any particular choice, the expected value of taking one box is considerably higher than that of taking both.[2] More recently it has been argued, for example by Ellery Eells, that the distinction between the two varieties of decision theory as originally conceived is spurious in certain respects, and that,

Synthese **86**: 173–196, 1991.

properly followed through, both theories recommend taking two boxes. But not all are convinced of the two-box solution; some post-theoretical intuitions hold out firmly for the one-box solution.[3] And the pre-theoretical temptation to take one box is quite strong.

Eells's analysis of the common cause structure of Newcomb's Problem casts considerable light. Here is a paradigmatic case of a common cause structure. Suppose there is a high statistical correlation between cigarette smoking and lung cancer. Suppose, however, it is discovered that lung cancer is not caused by smoking; rather, both lung cancer and smoking are caused by a particular gene. If one has this gene, one is very likely both to get lung cancer and to smoke, and refraining from smoking has no effect on whether one gets lung cancer. In Eells's helpful terminology, the gene is the common cause of both smoking and lung cancer in this case, smoking (or not smoking) is a symptomatic act, and getting lung cancer (or not getting it) is a symptomatic outcome.[4] Although symptomatic acts and symptomatic outcomes are strongly statistically correlated, they are causally independent, in the sense that neither causes the other; rather, both are caused by the common cause. In this case, intuition is fairly clear that, if one has the gene and one enjoys smoking, refraining from smoking because of the statistical correlation with lung cancer would be irrational. Refraining from such symptomatic acts, in a case with this common cause structure, would be irrational. If one finds oneself so refraining, that is good news, because of the statistical correlation of such symptomatic acts with the desired symptomatic outcome, but it would be irrational to refrain from smoking for the "news value" of the fact that one has refrained.

In Newcomb's Problem, taking just one box and thereby refraining from the thousand dollars is a symptomatic act. It is strongly statistically correlated with the desired symptomatic outcome of getting the million dollars. But, two-boxers in effect argue, there is not the right kind of causal connection between this symptomatic act and the desired symptomatic outcome; the act does not bring the outcome about. Rather, given the way the problem is described, one should assume that they have a common cause, which causes, on the one hand, the prediction of the predictor and in turn his decision to put either nothing or a million dollars in the opaque box, and, on the other hand, the attitudes of the agent and in turn his decision to take either one box or two. Eells writes:

> It seems that if the agent is rationally to have enormous confidence in the accuracy of the predictor . . . then the agent must believe that there is a causal explanation for his success, though he may not know what that explanation is, and neither may the predictor. Indeed, it seems presupposed by much of our inductive reasoning that a high statistical correlation has a causal explanation. The only kind of causal explanation of the predictor's success that I can think of that is consistent with the set-up of Newcomb's paradox is one that invokes a common cause. . . . Also, it seems that on any plausible account of any kind of successful prediction, the causal structure must be of this form. A successful predictor must have – consciously or unconsciously – a method, in the sense that the predictions are based on observations, conscious or unconscious. And if we look far enough back in the causal chain culminating in the relevant observations, we must be able to find factors that are causally relevant to the event predicted.[5]

However, if we assume that the common cause structure makes it irrational to refrain from smoking for the sake of its news value, and also that the common cause structure is present in Newcomb's Problem, we may wonder why the irrationality of refraining from smoking is so much less controversial than the irrationality of refraining from the thousand. What makes the one-box intuitions so recalcitrant? Indeed, arguments for the two-box solution have sometimes taken the form of a challenge to distinguish cases. But I am turning the challenge around, for diagnostic purposes: Why, at the level of intuition at least, is there a difference to begin with? If the appeal to news value does not have any tendency to make refraining from smoking appear rational, why should it have any tendency to make refraining from the thousand appear rational? Since simple evidentialist reasoning[6] seems to apply in both cases, something more seems to be needed to explain the special appeal of the supposedly evidentialist one-box solution. The intuitive difference suggests that a full diagnosis of one-box-itis may need to uncover further differentiating structure in the cases. News value and evidentialist reasoning do not get to the heart of the diagnostic problem.

In what follows I shall assume for the sake of argument that Newcomb's Problem does instantiate the common cause structure, and that the two-box solution to it is indeed correct. I shall offer a diagnosis of the appeal of the one-box solution, though by a round-about means, which suggests that evidentialist reasoning is a red herring that does not account for intuitions about the cases. I shall first be concerned to distinguish cases that admit the possibility of what I will call collective action (and certain intrapersonal analogues of it) from pure common cause cases. I have suggested elsewhere[7] that collective action may get a bad name by being wrongly assimilated to evidentialism. The intuitive

attractiveness of participation in collective action in appropriate cases does not depend on evidentialist reasoning, but on a collective causal power, which is sometimes but not always present when evidentialist reasoning is available. My diagnosis of one-box intuitions will involve trying to explain why Newcomb's case may appear (though, I think, wrongly) to be an appropriate case for collective action, while there is no parallel tendency for the smoking case and certain other cases. Here, however, I shall not be defending the rationality of collective action; my purpose is rather at the level of explanatory diagnosis. My explanation of intuitive reactions to various cases could be correct even though collective action were in fact irrational. Of course, the hypothesis I offer is open to rigorous experimental testing; the intuitions I here try to explain have been gathered locally and unsystematically.

A word of warning: while I am concerned to explain various intuitive responses, the explanation I offer in terms of amenability to collective action is not intended to reflect the occurrent attitudes of people presented with these various puzzle cases. I do not depend on any suggestion to the effect that in responding to these cases people are consciously attentive to whether conditions appropriate to collective action are met. It is possible, if my hypothesis best fits the data, that their varying intuitive responses to the cases can be explained in terms of these conditions even if they are not aware of the relevance of the conditions. The data, that is, are the varying intuitive responses to the cases, as opposed to people's own views about what is determining their responses (though admittedly this distinction may not always be perfectly sharp).

2. PRISONERS' DILEMMA MAY BE A NEWCOMB PROBLEM, BUT NEWCOMB'S PROBLEM IS NOT A PRISONERS' DILEMMA

David Lewis writes:

> Several authors have observed that Prisoners' Dilemma and Newcomb's Problem are related. . . . But to call them "related" is an understatement. . . . Prisoners' Dilemma is a Newcomb Problem – or rather, two Newcomb Problems side by side, one per prisoner. Only the inessential trappings are different. Let us make them the same.
>
> You and I, the "prisoners", are separated. Each is offered the choice: to rat or not to rat. . . . Ratting is done as follows: one reaches out and takes a transparent box, which is seen to contain a thousand dollars. A prisoner who rats gets to keep the thousand. . . . If either prisoner declines to rat, he is not at all rewarded; but his partner is presented

with a million dollars, nicely packed in an opaque box. . . . There we have it: a perfectly typical case of Prisoners' Dilemma.[8]

The parallels he sets out are persuasive, and are now familiar and widely accepted. I accept that Lewis has shown that Prisoners' dilemmas have certain features distinctive of Newcomb's Problem. In particular, they admit of evidentialist reasoning and the causalist rebuttal to it from dominance reasoning. As Lewis puts it:

> Some who discuss Prisoners' Dilemma think it is rational not to rat if the two partners are enough alike. Their reason is that they believe, justifiably, that those who do not rat will probably not be ratted on by their like-thinking partners. . . .
> . . . . And some – I, for one – who discuss Prisoners' Dilemmas think it is rational to rat no matter how much alike the two partners may be, and no matter how certain they may be that they will decide alike. Our reason is that one is better off if he rats than he would be if he did not, since he would be ratted on or not regardless of whether he ratted.[9]

But there is also a difference between prisoners' dilemmas and common cause problems such as Newcomb's Problem, which seems to have been overlooked. Prisoners' dilemmas have a further feature not found in Newcomb's Problem.

The difference, in a nutshell, is this: the prisoners together can bring it about that they both get the outcome they agree is second-best rather than the outcome they agree is third-best, even though neither can bring this result about individually, and neither can causally affect what the other will do. (For brevity, I shall refer to pairs of outcomes such as the second and third bet in the Prisoners' Dilemma, with respect to the relation between which there is agreement in the preference rankings of the parties, as *agreed outcomes*.[10]) The set of their acts has a collective causal power with respect to agreed outcomes that neither act has individually, despite the fact that neither act can causally affect the other act. I shall explain.

Neither prisoner knows whether the other is going to rat or not per se; but suppose each believes (1) that the other is rational, and (2) that it can be rational, given some agreed outcomes and even in the context or partly conflicting self-interested goals, to participate in collective action, that is, to cooperate with whoever else is cooperating in together bringing about the best agreed outcome possible for the cooperators to bring about through what they together do, given what any non-cooperators do, or are likely to do. (This is no less "neutral" an assumption about what they believe than the assumption that each

believes that acting individualistically is uniquely rational, i.e., on the basis of the causal consequences of one's individual act, given what the others do, or are likely to do. The dominance reasoning supporting the rationality of ratting illustrates individualistic rationality: given that the other rats, I'm better off ratting, and given that the other doesn't, I'm still better off ratting.) Collective action, in this sense, is different not only from acting individualistically, i.e., doing the act that will have the best consequences, given what others do, or are likely to do, but also from doing what would have best consequences were everyone to do it, regardless of whether others do so (or are likely to) or not. If someone committed to collective action became aware that in fact there were no other cooperators, he should do the act that will individually have the best consequences, given what the others will do, or are likely to do; but this is not the case for someone committed to doing what would be best were everyone to do it simpliciter.[11] In acting collectively, the cooperators first identify one another as cooperators, then determine what they together should do, thus avoiding the regress of mutual interdependent predictions of individual behavior as a basis for the action of each.

On these two assumptions, the prisoners may end up acting collectively to secure their joint second-best outcome rather than acting individualistically. Now these beliefs about rationality may be wrong; the prisoners may be irrational in acting collectively. I emphasize that I am not here assuming they would be correct. Moreover, cooperative collective action is no doubt difficult and problematic; in particular it may be extremely difficult to determine reliably who the other cooperators are. But nevertheless, collective action is not impossible.[12] As an empirical matter, people may in fact bring it off. If the prisoners do so, then, whether they are rational or not, they do, collectively, causally bring it about that the outcome they agree is second-best obtains rather than the outcome they agree is third-best. In larger cases also, such as refraining from overfishing a commons, the group of fishermen can collectively bring it about that the commons is not overfished by refraining; as a group, though not as individuals, they have this causal power. There are also cases in which a group of persons who share the same goal may have a causal power collectively, but not as individuals (though such shared-goal collective action cases are of course not Prisoners' dilemmas); for example, members of a party may, by turning out to vote, collectively bring it about that their candidate is elected.

By contrast, in a problem with what I am calling for the time being 'the pure common cause structure' illustrated by Newcomb's Problem, no collection of symptomatic acts is causally efficacious with respect to the desired symptomatic outcome. (But see the end of this paper for reservations about this terminology.) If many people simultaneously faced Newcomb's problems, they could no more bring it about that almost all of them got rich by acting together in taking only one box than they could bring it about that they got rich individually by each taking one box. Generalizing across time rather than across people, a long run of Newcomb's cases which preserves the common cause structure of the single case considered in isolation no more gives rise to a causal power than does the isolated case. (See the discussion of Skyrms on the long run vs. Mackie on reputation effects in repeated Newcomb's Problems, in Section 3 below.) In pure common cause cases, there are no collective causal powers lurking behind the lack of causal power in acts considered individually. To see this most clearly, consider again the most transparent common cause example: the case in which lung cancer is caused by a gene which also causes people who have it to smoke. Even if we consider the class of all people who may have the gene in question, and the class of all possible symptomatic acts by these people, we do not find causal dependence of the desired outcomes on sets of acts.[13]

In rendering a pair of Newcomb Problems side by side into a Prisoners' Dilemma, Lewis moves from the original characterization of the problem, in which I will get my million if and only if it is predicted that I do not take my thousand, to a new characterization, in which I will get my million if and only if you do not take your thousand. But this move admits a possibility of collective action with one's like-thinking, hence potentially predictive double that was not present in the original characterization of Newcomb's Problem and that is also not present if many people simultaneously face Newcomb's problems as originally characterized. That is, this move admits a collective causal power that does not exist in pure common cause cases: the pair of what were merely symptomatic acts are now together, though not individually, possessed of the causal power to bring about the desired agreed outcome.

However, the previous paragraph needs qualification. There is a sense in which, in the original problem, the predictor and the predictee do have the collective causal power to bring it about that the predictee

takes one box and gets exactly a million rather than takes two boxes and gets exactly a thousand. Just as is the case in the Prisoners' Dilemma, and as is not the case in the smoking case, the states of the world on which the outcomes of what one does depend are in fact the acts of another agent. Nevertheless, this collective causal power is strictly irrelevant because of the lack of even partial agreement with respect to outcomes: the predictee in fact has no reason to believe that the predictor prefers that the predictee takes one box and gets exactly a million rather than that he takes two boxes and gets exactly a thousand. Because there is no basis for such an assumption, it is strictly incorrect to interpret the situation of the predictor and predictee as one that admits of collective action. Some degree of mutually recognized overlap in preferences, with respect to agreed outcomes, is necessary for there to be a possibility of collective action in the sense I intend. But perhaps there is a temptation to read the satisfaction of this condition into Newcomb's Problem, and thereby to create an illusory possibility of cooperation with the predictor.

To see this, consider a different case, in which the predictee is a child and the predictor is a parent who wants the child not to be greedy on a particular occasion. In this case, as is not the case in Newcomb's Problem, we are explicitly given the preferences of both parties. The child simply prefers getting more money to less. To make this stipulation more realistic, perhaps amounts of candy, or some other quantity, should be involved, rather than money; I shall stick to the amounts of money involved in the original Newcomb case merely because they are familiar. The parent doesn't mind about whether his prediction is right or not; what he most prefers is that the child not be greedy on this occasion, that is, that he take one box rather than two; this concern has priority over concern with saving money. But as between two situations in which the child takes the same number of boxes, the parent prefers the one that costs him less money. This stipulated preference ordering for the parent may seem unrealistic for the large amounts of money involved in the original Newcomb's Problem, unless we also stipulate that the parent is very rich; however, the basic pattern of concerns, of a parent who doesn't mind spending money in relation to a child, though he doesn't want to throw it away, is very familiar. Thus, both parent and child prefer the child's taking one box and getting a million to the child's taking two and getting a thousand; these are the agreed outcomes. But the child would most like to take two boxes and

get a million plus a thousand and would least like to take one and get nothing. The parent, on the other hand, would most like for the child to take only one box and get nothing (the child has not been greedy and this result has cost the parent nothing) and would least like the child to take both boxes and get a million plus a thousand.[14] Their preference rankings are as follows:

Parent	Child
Child takes one, gets \$0	Child takes two, gets \$M+T
Child takes one, gets \$M	Child takes one, gets \$M
Child takes two, gets \$T	Child takes two, gets \$T
Child takes two, gets \$M+T	Child takes one, gets \$0

Now the pair are in a Prisoners' Dilemma, and, assuming these preferences and thus the fact that the middle pair of outcomes are agreed are known to each, the sense in which they can collectively bring it about that the predictee takes one box and gets exactly a million rather than takes two boxes and gets exactly a thousand does give rise to a possibility of collective action. The prediction in such a case might be made by identifying the child as another cooperator, then determining what's best for all cooperators to do together. Perhaps there is a temptation to project something like these familiar parental motivations onto the predictor in the original Newcomb's Problem in order to make sense of the game he is playing. The one-box temptation could then be understood in terms of the urge to cooperate in a Prisoners' Dilemma of this parent-child type. (If this type of case is to be indefinitely repeated, in the course of parental character training, the child may be rational repeatedly to take one box even from an individualistic perspective.)

This is the situation one would have in Newcomb's Problem if there were any basis for assuming the predictor has such preferences (motivated, for example, by concerns that the predictee take one box rather than two – or, perhaps, to punish causal maximizing – and, other things equal, to save money), and hence that there are any agreed outcomes. But in Newcomb's Problem there is no information given to support such an assumption, as opposed to various other possible assumptions that would not yield any agreed outcomes, about the predictor's preferences. This is why I regard such an interpretation of Newcomb's Prob-

lem as a misinterpretation: it reads an opportunity for cooperation into a situation in which strictly speaking there is no basis for assuming it.

But here is an objection to the foregoing suggestion. I suggested that the one-box temptation may be explained in terms of a misguided urge to cooperate given information that actually provides an inadequate basis for cooperation but which into which it may be natural to read further conditions which would yield a situation appropriate for cooperation. But if this is the explanation, why is there no parallel temptation in the Jones-ruthlessness case described by Gibbard and Harper?[15] The objection is in effect a challenge to distinguish the original Newcomb's Problem, in which the misguided cooperative urge purportedly arises, from the Jones case, in which it seems not to.

The Jones case is as follows. Jones is in competition with other executives for a lucrative promotion. The boss found the competitors so well matched that he employed a psychologist to break the tie by testing for qualities that will lead to successful performance in the corporate world. The test was given on Thursday; the decision to promote will be made on Monday, on the basis of the test. It is now Friday, and Jones learns from a reliable source that all competitors scored equally well on all factors except ruthlessness and that as a result the promotion will go to whichever one of them scored highest on this factor. On Friday, Jones must decide whether or not to fire an underling, Smith, who has had trouble meeting his sales quota this past month because of the death of his wife. Jones believes that Smith will get over his troubles, that leniency, would, other things being equal, be best for the company, that he could, if given a chance, convince the boss of this and that this would reflect well on his own astuteness. But he has no way of contacting the boss until after the promotion decision is announced on Monday. Jones knows that firing Smith is good evidence that Jones has scored highly on ruthlessness, while leniency is good evidence that he has not. Should Jones fire Smith or not?

Firing Smith in this case is intuitively unappealing and irrational, or at least is considerably less tempting than taking one box is. Why? Why doesn't the same misunderstanding of the problem as having a Prisoners' Dilemma structure, in this case as a situation which lends itself to cooperation with the predictor of ruthlessness, arise here? Don't Jones and the boss (and perhaps the psychologist as well) collectively have the causal power to bring it about that Jones fires Smith and gets

promoted rather than that Jones does not fire Smith and does not get promoted?

Once raised explicitly in this way, these questions almost answer themselves. It was natural to understand the Newcomb predictor as having motivations like the parent with respect to the predictee, as one way of making sense of the game he is playing. This interpretation, though strictly speaking unwarranted, supplies the agreed outcomes needed to make the situation admit of cooperation. There is no similar natural misreading of the motivations of the boss or the psychologist in the Jones case. That is, there is no temptation to suppose that either the boss or the psychologist shares Jones preference that Jones fires Smith and gets promoted rather than that Jones not fire Smith and not get promoted. There is no interpretative point in supposing the boss has the relevant preferences with respect to Jones; he just wants to promote the most ruthless person, and probably doesn't mind much one way or another whether that turns out to be Jones. Nor is there any interpretative point in supposing the psychologist cares at all about these outcomes; it's much more natural to suppose she is completely impartial. So it is really not very hard at all to understand why the urge to cooperate intuitively does not arise in this case.

3. THE ANALYSIS OF AGENCY AND FURTHER CASES

Common cause cases teach us to distinguish the relevant kind of causal independence, which obtains between symptomatic acts and outcomes in these cases, from probabilistic independence, which does not. And causal decision theory tells us that rational action depends on causal, not merely probabilistic, relations between acts and outcomes, on relations of bringing about.[16] But the difference pointed out in the preceding section between cases which admit of collective action and pure common cause cases suggests that it is not enough to distinguish causal and probabilistic independence. We must further distinguish at least three different forms in which issues about independence, whether causal or probabilistic, may arise; here I shall be concerned with varieties of causal independence, which fail just when the bringing-about relation holds, though there are probabilistic analogues of each of the three kinds of independence. The first is the familiar form of issues about whether the individual symptomatic act of taking one box, or

refraining from smoking, is causally, though not probabilistically, independent of the symptomatic outcomes, in the sense that one cannot causally bring about the more desired symptomatic outcome by doing the symptomatic act. Call this *individual act-outcome independence* (IAOI). The second is a form of issue familiar from game theory and consideration of Prisoners' dilemmas and collective action, about whether one act is independent of other acts: here the separate acts may be separate acts by each of two prisoners, or by each of many fishermen, each of many voters, etc. Call this *act-act independence* (AAI). AAI is a special case of the independence of states of the world on which outcomes depend from acts, in which the states of the world in question are determined by what other agents do. Each agent in such cases may be aware that his act cannot bring it about that others will act one way or another. He may also be aware that this act cannot itself bring about the better of the agreed outcomes. He may nevertheless ponder whether to cooperate with whoever else is cooperating by participating in a collective act that would bring about the better of the agreed outcomes, given the (likely) behaviour of non-cooperators. The third form of issue, then, is about whether the symptomatic outcomes are independent of some collection of possible acts, or collective act. Call this *collective act-outcome independence* (CAOI). From the fact that causal independence of the first two forms obtains it does not follow that causal independence of the third form obtains.[17] Given at least limited agreement with respect to outcomes, when IAOI and AAI hold but CAOI does not, collective action is in the offing, in the sense that the relevant collective causal power with respect to the agreed outcomes exists. In the absence of this causal power, as in a many-person version of the smoking case, 'collective action' would be causally pointless. In pure common cause cases, neither IAOI nor CAOI holds.

These forms of causal independence may be applied either interpersonally or intrapersonally. That is, the collection or set of possible acts being considered may be a set of acts by different people, or it may be a set of acts all by one person. Interpersonal versions of these forms of independence are typically at issue in questions about collective action and about the rationality of adhering to some form of rule-utilitarianism or related doctrine.[18] Intrapersonal versions of these forms of independence may be at issue in questions about character formation (and probabilistic analogues of them in questions about what

individual behaviour is made rational by consideration of the long run and laws of large numbers.)[19]

In the original Prisoners' Dilemma, in the one-off parent-child case, and in Lewis's Newcombized version of the Prisoners' Dilemma, there is limited agreement with respect to outcomes, and while IAOI and AAI hold, CAOI does not; so collective action is in the offing. Whether or not participation in collective action is rational, people do seem to be weakly disposed, in varying degrees, to such participation; for some people and some kinds of cases, this disposition seems stronger than the disposition to reason and act individualistically.[20]

By contrast, in the one-person smoking case, while IAOI still holds, CAOI also holds, for the various acts of smoking by one person; so no intrapersonal analogue of collective action is in the offing. (Another difference is that AAI seems not to hold between acts of smoking by one person, assuming that person becomes addicted over time.) And in the many-person version of the smoking case as well, IAOI and CAOI both hold again; so collective action is not in the offing. (Here, AAI may or may not hold between acts of smoking of different persons; one person's smoking may or may not causally influence another to smoke.) In the pure common cause case, no collection of symptomatic acts has the relevant causal power with respect to symptomatic outcomes. And this is transparently the case here. There is no temptation to think in terms of cooperating, or participating in collective action, since it is obvious, *ex hypothesi*, that no set of acts of which one's own act might be a member has the relevant causal power. Nor do the genes that cause lung cancer do so by means of acts of smoking; the latter are in a sense epiphenomenal in relation to lung cancer.

There is, admittedly, an evidentialist argument for cooperation by like-minded persons in the Prisoners' Dilemma; but there is also a collective causal power with respect to agreed outcomes, which is lacking in the smoking cases. (Knowledge that another is like-minded should enable each to infer that the other is as disposed to cooperate with whoever else is cooperating as he himself is, and thus may facilitate collective action and the exercise of the collective causal power by facilitating identification of the class of cooperators.) People's intuitive responses to various cases seem to correspond to differences among the cases with respect to their amenability to collective action. It is far less controversial and more intuitive that refraining from smoking would

be irrational than that cooperating in the Prisoners' Dilemma would be irrational. If evidentialist reasoning accounted for the appeal of cooperative collective action, there would seem to be no reason for it to be less intuitively attractive when collective action is not in the offing, as in the smoking cases.

We've already seen the way in which Lewis relates a pair of intertwined Newcomb's Problems to the Prisoners' Dilemma, and how in doing so he changes them from pure common cause cases to a case that admits of the possibility of collective action with respect to agreed outcomes. And we've seen how the parent-child interpretation also has this effect, by supplying agreed outcomes. To introduce a collective causal power with respect to agreed outcomes is to eliminate CAOI; and without CAOI, the pure common cause structure that seems to be the essence of the original Newcomb's Problem, correctly understood, is lost. There are various ways to generalize the Newcomb situation, whether interpersonally or intrapersonally, some of which preserve and some of which do not preserve CAOI. If we simply take a large group of people and present them simultaneously with Newcomb's problems as originally characterized, CAOI and common cause structure are preserved, and collective action is not in the offing. So this way of generalizing Newcomb's Problem interpersonally retains more of the features of the original problem than Lewis's variation on Newcomb's Problem does.

A version of Lewis's Newcombized Prisoners' Dilemma may be realized intrapersonally rather than interpersonally, to illustrate an intrapersonal analogue of a case that admits of collective action. Consider a dilemma for me now, a sports lover, and me later, a music lover. Each of us has to choose, one now and one later, in conditions of secrecy, between taking the first box and taking both boxes. I know now that I have a terrible memory, and will not later remember what I now choose to do. If, as it turns out after we have both chosen, we have both taken both boxes, then a thousand dollars will be given to each of a sports charity and a music charity at some still later point in time. If we have both taken just the first box, a million will be given to each of them. If I now take just the first and I later take both, then the sports charity will get nothing and the music charity will get a million plus a thousand. If I now take both and I later take just the first, then the sports charity will get a million plus a thousand and the music charity will get nothing. I now and I later do, together, have the causal power to bring it about

that each charity gets a million rather than a thousand, whether or not we would be rational to do so. In this case, CAOI does not obtain, even though IAOI and AAI may be assumed to obtain.

There are other possible intrapersonal analogues of cooperative collective action. Consider Gibbard and Harper's Solomon case. King Solomon wants another man's wife. He has studied psychology and political science, which teach him the following:

> [K]ings have two basic personality types, charismatic and uncharismatic. A king's degree of charisma depends on his genetic make-up and early childhood experiences and cannot be changed in adulthood. Now charismatic kings tend to act justly and uncharismatic kings unjustly. Successful revolts against charismatic kings are rare, whereas successful revolts against uncharismatic kings are frequent. Unjust acts themselves, though, do not cause successful revolts; the reason that uncharismatic kings are prone to successful revolts is that they have a sneaky, ignoble bearing. Solomon does not know whether or not he is charismatic; he does know that it is unjust to send for another man's wife.[21]

As stipulatively described, this is a pure common cause case in which both IAOI and CAOI hold: the common cause is charisma, the symptomatic acts are just and unjust acts, and the symptomatic outcomes are revolt and no revolt. Nevertheless, the stipulations made about the relationship of character to acts may be hard to accept as realistic, in particular the stipulation that character is not affected by acts in adulthood.

What would the relation of character to actions have to be for the pure common cause structure to hold and to resist subversion by an intrapersonal analogue of a collective causal power? Of course one's character would have to be the causal source of the relevant symptomatic outcomes, in the way that the gene is the causal source of lung cancer. And one's character would have to be the cause of one's symptomatic act, and moreover of the set of one's symptomatic acts, in the way that the gene is the cause of one's smoking. Moreover, the causation of outcomes by character could not operate by means of one's symptomatic act (by contrast with reputation effects in which IAOI fails – see the discussion of Mackie below) or by means of the set of symptomatic acts (by contrast with Lewis's Newcombized Prisoners' Dilemma and the intrapersonal analogues of it in which CAOI fails), any more than the causation of lung cancer by the gene operates by means of acts of smoking. Symptomatic acts must be individually and collectively epiphenomenal in relation to symptomatic outcomes.

However, where the common cause of acts and outcome is suppos-

edly one's character, it is very hard to regard acts as epiphenomenal in the relevant way. Some aspects of the way in which we naturally understand the relation of character to actions seem to make it hard to swallow what amount to common cause stipulations involving character as a common cause. We may become courageous, as Aristotle points out, by doing courageous acts, or become just by doing just acts (or become charismatic, if you will, by doing just acts, contrary to the stipulation in the Solomon problem). However, it need not be the case that any individual just act causes someone to have a just character, holding all his other acts constant, or causes him to do other just acts. Even so, collectively his just acts over time may have this causal power. That is, CAOI may fail to hold intrapersonally between the set of just acts and a just character, even though IAOI holds between each individual just act and a just character and AAI holds between individual just acts. And if outcomes follow causally from character, then these relations would hold or fail to hold between acts and outcomes as well.

Of course, the Solomon case stipulates in effect that CAOI does hold between sets of acts and character, and between sets of acts and outcomes. But I am suggesting that these stipulations about the relations of acts to character are unrealistic, so that it is natural to slip into a misreading of the case in which CAOI does not hold. Hence, one might expect some covert urge to intrapersonal "collective" action in the form of a concerted effort at character formation to render the irrationality of acting justly in the Solomon case less transparent to intuition than is the irrationality of refraining from smoking or firing Smith. And this prediction is confirmed, at any rate, by my intuitions about the case.

Consider also two further ways of generalizing Newcomb's Problem intrapersonally. Mackie, for example, suggests an intrapersonal generalization over time reminiscent of certain arguments for the rationality of cooperation in Prisoners' dilemmas repeated an indefinite number of times. He writes:

> Suppose . . . that the player is in fact free to make either choice on each occasion, unrestricted by any established character, but that the psychologist-seer thinks that the player's actions are determined by his character, and suppose that the player knows that the seer believes this. Suppose, besides, that the game is to be played repeatedly by the same player against the same seer. Then it will be sensible for the player to take only the closed box every time, since on these assumptions this will ensure that the seer

regularly puts [the large amount] into it. These assumptions – in particular that of repeated playing of the game – reverse the direction of causation and enable the player's choices to determine the seer's moves. . . .[22]

In these circumstances, as in the indefinitely repeated parent-child case, reputation effects may give individual acts of taking one box the power to bring about desired outcomes, albeit with a delay. Reputation effects operate via the expectations of others about future acts by the same person, since, for example, they may plausibly take the act to be evidence of the state of character that caused it. Without IAOI, the pure common cause structure of the original Newcomb's Problem is again lost. Taking one box is rational here in terms of the causal powers of individual acts, via reputation effects, just as not ratting in a series of indefinitely repeated Prisoners' dilemmas may be rational in purely individualistic terms, via reputation effects, with no need for any appeal to collective action.[23]

But compare another way of generalizing Newcomb's Problem intrapersonally, across time, discussed by Brian Skyrms, which does not give rise to reputation effects, and which preserves IAOI and CAOI and hence the common cause structure of the original Newcomb Problem. Skyrms considers the view, which he regards as fallacious, that evidential decision theory is favoured over causal decision theory by the law of large numbers, i.e., by fact that, in a long run of independent trials of a game of chance, the average winnings per trial will almost certainly converge to the expected winnings on a single trial. He draws a parallel between Newcomb's Problem and a case involving a biased coin, in which the agent has degree of belief of 1/2 in each of two hypotheses: that the chance of heads is 2/3, and that the chance of heads is 1/3. "His degree of belief that heads will come up in a given single toss is 1/2. The relevant form of the law of large numbers shows that he has degree of belief one that in an infinite sequence of independent trials, the relative frequency will converge to either 1/3 or 2/3. He does not think that the relative frequency will converge to the single-case degree of belief of 1/2. . . ." Concerning a generalization of Newcomb's Problem across time he then writes:

We have a $1,000,000 slot machine good for a free play set either to have a chance of 0.999999 of paying the million or to have a chance of 0.000001 of paying out the million, you don't know which. There is a $1,000 bill already sitting in the payoff tray. You can either play and keep the thousand or play and return the thousand to the management. The machine was previously set by a psychologist after giving you, for reasons unknown

to you, a battery of psychological tests, and so forth as before. The management cannot cheat; there is no way to alter the setting of the machine when you decide whether to return the thousand. So the chances are fixed by the state alone. They are not influenced here by the act as they would be if the management could cheat. Then you should believe that in an infinite sequence of objectively independent repetitions of this experiment (whichever experimental setup it is) the average payoff of the strategy of taking the thousand will exceed the average payoff of the strategy of not taking the thousand by one thousand dollars. . . .

When I said that the idea that the long run argument favors evidential decision theory rests on a fallacy, I meant the fallacy of treating uncertainty about chance as if it were certainty about a different chance. If you thought that your *symmetric uncertainty* about the bias of the biased coin was undistinguishable from *knowledge* that the coin was fair, then you might try to apply the law of large numbers for independent and identically distributed trials to conclude that the long-run relative frequency of heads will almost certainly be 1/2. But this is a *mistake*.[24]

In Skyrms's intrapersonal generalization, as I said, reputation effects are ruled out and hence IAOI holds. And there seems to be no reason for CAOI to fail intrapersonally here if it does not fail interpersonally, for a large group of people simultaneously facing Newcomb's problems, which we have already admitted; so CAOI holds here too, and collective action is not in the offing. Thus, the pure common cause structure of the original Newcomb's Problem is preserved. Taking the thousand repeatedly is rational.

4. CONCLUDING REMARKS

Reputation effects in certain indefinitely repeated cases negate IAOI; collective causal powers with respect to agreed outcomes negate CAOI; either is sufficient to subvert what I've been calling the pure common cause structure of the original Newcomb's Problem. There is, however, something misleading about this way of putting the point. The point is not that the various acts that may, given the right interconnecting attitudes, together constitute collective action do not have a common cause. If they did have a common cause, but nevertheless constituted collective action, it would be irrelevant that they had a common cause. It is rather that the collective causal power with respect to agreed outcomes adds something which gives rise to an apparent reason for action not present in mere common cause cases. If the prisoners in the dilemma were identical twins brought up in identical environments, and if as a result the assumption were warranted that the causes of their

acts were common, this would not lessen the intuitive attractiveness of collective action; intuitively, it would be irrelevant, so long as the causal link running from common causes to agreed outcomes operated through the cooperative acts. Parallel comments would apply with respect to common causes of the cooperative acts in the music/sports lover case and the parent-child case. What is intuitively at issue in these cases is not whether or not my act, or any collective act it might be part of, is already causally determined. It is rather whether my act (whether already causally determined or not), might be part of a collective act (whether already causally determined or not) that itself has causal power with respect to agreed outcomes. Even if the predictor in the original Newcomb's Problem were perfectly reliable, one would have no reason to take just one box: not because one's own act is already causally determined, but because neither one's own act, nor any collective act of which it is a part, has the relevant causal powers with respect to agreed outcomes. This case is like a smoking case with 100% reliable determination of smoking (given the availability of tobacco) by the gene. But compare a parent/child case with a perfectly reliable parent predictor such that both the prediction and the behaviour by the predictee are products of a common cause. Here, there is still a collective causal power with respect to agreed outcomes, just as there is in the Prisoners' Dilemma with identical twins identically raised. The common causal determination of the constituent acts of a collective act does not per se undermine the attractiveness of cooperation if the causal link from common cause to agreed outcomes goes through the cooperative acts.

I have suggested that cooperative reasoning may account for the appeal of apparently evidentialist behaviour; a collective causal power with respect to agreed outcomes, not evidentialist reasoning, makes cooperation attractive, whether rightly or wrongly, in the Prisoners' Dilemma. And the natural illusion of such a power may account for the intuitive appeal of certain responses in various other cases, including Newcomb's Problem. Note again that I have not here endorsed or tried to defend the rationality of collective action, but merely to account for different intuitive reactions to various cases. Among various cases that equally admit of evidentialist reasoning, the supposedly evidentialist solution has varying degrees of intuitive attractiveness. The hypothesis I have put forward here is that this variation may be accounted for, at least in part, in terms of differential amenability (actual or apparent)

to collective action. Collective action is intuitively more tempting than evidentialism; whether rightly or wrongly is another question.

NOTES

* This paper was originally submitted to *Synthese* in March 1989. For helpful comments and criticisms of earlier versions I am grateful to Michael Bacharach, John Broome, David Gauthier, Isaac Levi, Adam Morton, Derek Parfit, Howard Sobel, Robert Sugden, Bas van Fraassen, and members of audiences on various occasions on which I have presented this paper. I am also grateful to the Humanities Council of Princeton University for their generous support during the period when this paper was written.

[1] It is of course important to distinguish conditional probabilities of a particular prediction, given the corresponding choice, from conditional probabilities of a particular choice, given the corresponding prediction. As I understand Newcomb's Problem, conditional probabilities of the former type rather than the latter are part of the description of the problem in terms of reliable prediction. A reliable prediction is thus being understood along the lines of reliable tests or reliable witnesses: the relevant conditional probability is that of the test or witness indicating that such and such, given that the world is so and so. I take this stipulation of how the problem is to be understood to be supported by the supposition made in the description of the problem that, concerning some high percentage of the many other cases of choices made under the Newcomb circumstances, the choice was predicted correctly (as opposed to the supposition that, concerning some high percentage of the many other cases of predictions made under Newcomb circumstances, the predicted choice was made). See Nozick 1985, p. 107; Lewis 1985, p. 253.

For a different view of how Newcomb's Problem is correctly to be understood, see Isaac Levi 1975, 1982, 1985; see also Levi 1983. I cannot address Levi's views in this article, but assume for the sake of argument many of the positions of causal decision theory that he argues against.

[2] Given what has been said in note 1 about the high probability of correct prediction given a particular choice, this is not the non sequitur that Levi argues against (see Levi 1982), since the problem is not understood to begin with in the way Levi understands it and the fallacious inference he complains of is not made.

[3] See, for example, Horgan 1985.

[4] See Eells 1982, p. 92ff.

[5] Eells 1982, pp. 210–11.

[6] By this phrase I intend to distinguish simple evidentialist reasoning for the one-box solution as originally conceived from the sophisticated evidentialism (involving the so-called tickle defense and screening off arguments) as developed by Eells 1982, and others. The latter would give the opposite verdict in both cases, but would equally fail to account for the intuitive distinction between them.

[7] See Hurley 1989, chapter 8, sections 1–4, especially the discussion of the diagnostic vs. the cooperative interpretations of Quattrone and Tversky's voting experiment.

[8] Lewis 1985, pp. 251–52.

[9] Lewis 1985, p. 254. See also Gibbard and Harper 1985, pp. 156–57; Horgan 1985, p. 180.

[10] I intend agreed outcomes to be members of a pair of outcomes both ranked in the same way by the relevant parties: for example, *a* and *b* are agreed outcomes when both parties agree that *a* ranks above *b*. So the outcomes of "Both keep silent" and of "Both confess" are agreed outcomes in the Prisoners' Dilemma because the same outcome is preferred to the other by both parties:

A	B
A confesses, B keeps silent	B confesses, A keeps silent
Both keep silent	Both keep silent
Both confess	Both confess
B confesses, A keeps silent	A confesses, B keeps silent

The agreement in question is on the relation between outcomes in a ranking, not on the position – first, second, etc. – of an outcome in a ranking; thus, an outcome is not an agreed outcome in this sense when, for example, it is second in the rankings of each party, but stands in no agreed relation to another outcome. There are complications for the notion of agreed outcomes involving triples of outcomes, which I do not address here: for example, when both parties agree that *a* and *b* rank above *c*, but one puts *a* above *b* and the other *b* above *a*.

[11] I have discussed these matters in Hurley 1989, chapter 8, sections 1–4; and see Regan 1980.

[12] See Regan 1980, for an argument to the effect that the notion of cooperating with whoever else is cooperating does not involve an infinite regress or necessary indeterminacy; see and compare Sobel 1985. See also Howard 1989, on the possibility of a Prisoners' Dilemma-playing computer program that recognizes programs identical to itself and cooperates when it recognizes that its opponent program is identical to itself.

[13] See and compare Eells 1985, on the distinction between what he calls Type A and Type B beliefs, and on why ". . . what is relevant to our agent in his deliberation is the probability of his having the common cause conditional on his performing (or not performing) and asymptomatic act, and not the frequency with which the common cause is present among people who perform (or do not) the symptomatic act", pp. 200–01, 204. See also Eells 1982, chapter 6 and p. 209.

[14] I am indebted to objections made by Howard Sobel, which prompted me to clarify various points about what I call 'parental' motivation and why I regard it as unfounded to interpret the predict in Newcomb's Problem as similarly motivated.

[15] See Gibbard and Harper 1985, pp. 142–43.

[16] On what it is for an act to bring about an outcome, such that the relevant kind of causal independence between acts and outcomes does not obtain, see Gibbard and Harper, 1985, pp. 144–5. Causal decision theory typically employs counterfactuals to capture this notion: an act brings about an outcome if, the act were to be done the outcome would obtain, and it is not the case that if the act weren't to be done the outcome would obtain. The maximization exercise then focusses on the utility of an outcome times the probability that an act will bring it about. Note that it is not sufficient for the truth of the counterfactual "if the act were to be done the outcome would obtain" that the act merely somewhat raises the probability of the outcome, or that the act is a necessary condition of the outcome. Nor is it necessary for the truth of this counterfactual that the act raise the probability of the outcome to 100% (unlikely events in far possible

worlds may keep this probability below 100%), or that in all possible worlds if the act is done the outcome obtains.

[17] There would seem at first blush to be eight possible combinations of IAOI, AAI and CAOI. Actually there may be fewer than eight, however. For example, without either AAI or CAOI it's hard to see how one could have IAOI: if I can cause other people to act, perhaps through training or organization, and we together can bring about the result, then it would seem that I can bring about the result.

[18] Act-utilitarian criticisms of the rule-utilitarianism, for example, often involve pointing out the causal irrelevance to individual action of hypotheses about what would happen were everyone to act in some particular way, given that one's own act has no tendency to cause others to act that way and that one's own failure to act has no tendency to cause others not to act that way. Rule-utilitarianism differs from collective action in my sense in predicating what should be done by one person on what would happen were others to act similarly rather than on what is best for all those actually disposed to cooperate with other cooperators to do, but it does not depart from the assumption of interpersonal act-act causal independence. Again, see Regan 1980, chapters 8–10.

[19] See Hacking 1965, for criticism of the connection Peirce draws between rationality with respect to the long run and a kind of collective action. Peirce suggests that someone who must guess on a life and death issue on the sole basis of frequency information, and whose immediate concern is not long run success but ensuring any run at all, should identify with others and think of himself as making one in the long run of all human beings' guesses. Thus, his one-off guess in accord with frequency information may be justified, not in terms of the long run of his own guesses, but rather in terms of the long run of guessing by human beings. See, e.g., Peirce 1955 pp. 162ff.

Hacking replies, first, that the rationality of this guess is evident even to a misanthrope, and, secondly, that it would be evident even to a nuclear button pusher on whose guess turned the continued existence of human life. See Hacking 1965, p. 47. But Hacking's first point goes wrong to the extent it assumes collective action requires completely shared goals or benevolence (see Hurley 1989, chapters 8.1–4); it does require that there be some agreed outcomes. His second point also seems to me to go wrong, because the fact that bad luck early on might preclude collective action by nuclear button pushers by eliminating the human race does not mean that, given good luck, collective action by nuclear button pushers is not possible. Even here, where collective action presupposes survival, Peirce's plausible idea would be that it is possible for the class of nuclear button pushers collectively to maximize their chances of survival.

The intuitive idea behind various laws of large numbers is that, if we repeat an experiment involving a random variable very many times, the average of the empirical values the variable take will approach the underlying expected or statistical mean value. The result of each repetition of the experiment is assumed to be independent, both probabilistically and causally, of the results of other repetitions.

[20] See my discussion of Quattrone and Tversky's voting experiment, in Hurley 1989, chapter 8.3; and their report of it in Quattrone and Tversky 1986. See also Sen 1982 and 1985.

[21] See Gibbard and Harper 1985, p. 141.

[22] Mackie 1985.

[23] On cooperation in repeated Prisoners' Dilemmas, see various articles in Campbell and Sowden 1985, part V.

[24] Skyrms 1984, p. 88.

REFERENCES

Campbell, Richmond and Lanning Sowden: 1985, *Paradoxes of Rationality and Cooperation*, University of British Columbia Press, Vancouver.

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

Eells, E.: 1985, 'Causality, Decision and Newcomb's Problem', in Richmond Campbell and Lanning Sowden, (eds.), *Paradoxes of Rationality and Cooperation*, University of British Columbia Press, Vancouver.

Gibbard, A. and W. L. Harper: 1985, 'Counterfactuals and Two Kinds of Expected Utility', in Richmond Campbell and Lanning Sowden, (eds.), *Paradoxes of Rationality and Cooperation*, University of British Columbia Press, Vancouver.

Hacking, Ian: 1965, *Logic of Statistical Inference*, Cambridge University Press, Cambridge.

Horgan, Terence: 1985, 'Counterfactuals and Newcomb's Problem' and 'Newcomb's Problem: A Stalemate', both in Richmond Campbell and Lanning Sowden (eds.), *Paradoxes of Rationality and Cooperation*, University of British Columbia Press, Vancouver.

Howard, J. V.: 1989, 'Cooperation in the Prisoners' Dilemma' (typescript, London School of Economics).

Hurley, S. L.: 1989, *Natural Reasons: Personality and Polity*, Oxford University Press, New York.

Levi, Isaac: 1975, 'Newcomb's Many Problems', *Theory and Decision* **VI**, 161–75.

Levi, Isaac: 1982, 'A Note on Newcombmania', *Journal of Philosophy* **LXXIX**, 337–42.

Levi, Isaac: 1983, 'The Wrong Box', *Journal of Philosophy* **LXXX**, 534–42.

Levi, Isaac: 1985, 'Common Causes, Smoking, and Lung Cancer', in Richmond Campbell and Lanning Sowden (eds.), *Paradoxes of Rationality and Cooperation*, University of British Columbia Press, Vancouver.

Lewis, David: 1985, 'Prisoners' Dilemma is a Newcomb Problem', in Richmond Campbell and Lanning Sowden (eds.), *Paradoxes of Rationality and Cooperation*, University of British Columbia Press, Vancouver.

Mackie, J. L.: 1985, 'Newcomb's Problem and the Direction of Causation', in his *Logic and Knowledge, Selected Papers*, Vol. I, Clarendon Press, Oxford.

Nozick, F.: 1985, 'Newcomb's Problem and Two Principles of Choice', in Richmond Campbell and Lanning Sowden (eds.), *Paradoxes of Rationality and Cooperation*, University of British Columbia Press, Vancouver.

Peirce, C. S.: 1955, 'On the Doctrine of Chances, with Later Reflections', in Justus Buchler (ed.), *Philosophical Writings of Peirce*, Dover Publications, New York.

Quattrone, G. A., and Amos Tversky: 1986, 'Self-Deception and the Voter's Illusion', in Jon Elster (ed.), *The Multiple Self*, Cambridge University Press, Cambridge.

Regan, Donald: 1980, *Utilitarianism and Co-operation*, Oxford University Press, Oxford.

Sen, A. K.: 1982, 'Preference and the Concept of Choice', in his *Choice, Welfare and Measurement*, Blackwell, Oxford.

Sen, A. K.: 1985, 'Goals, Commitment, and Identity', *Journal of Law, Economics and Organization* **1**, 341–55.

Skyrms, Brian: 1984, *Pragmatics and Empiricism*, Yale University Press, New Haven.

Sobel, Jordan Howard: 1985, 'Utilitarianism and Cooperation', *Dialogue* **XXIV**, 137–52.

St. Edmund Hall
Oxford University
Oxford OX1 4AR
England

EDITORIAL NOTE

Synthese editors are committed to seeing to it that their contributors receive proper credit for their ideas. In the case of Dr. Hurley's paper and the paper 'Some Versions of Newcomb's Problem are Prisoners' Dilemmas' by Professor Sobel (below), their early history is intertwined. In particular the idea that the original Newcomb's Problem may be interpreted as a Prisoners' Dilemma by supplementing the description of the case with appropriate preference ranking occurs both in Hurley's paper and Sobel's paper. It may therefore be in order to register the fact that the idea was recorded by Hurley in the March 1989 version of her paper, i.e., four months before Sobel wrote his paper.