

NEWCOMB'S PROBLEM AS A THOUGHT EXPERIMENT

Tamara Horowitz

SOME philosophers believe that Newcomb's problem is a useful source of intuitions against which to test theories of rational choice. And some philosophers believe that Newcomb's problem is outrageous enough that intuitions about what to do in the face of it ought to have no bearing on the enterprise of codifying the principles of practical reason. I shall argue that if Newcomb's problem is considered properly, both of these opinions are right. Newcomb's problem as ordinarily presented conceals two very different kinds of philosophical thought experiment. One of them is a clear illustration of the importance of indirect strategies. The other is a thought experiment with interesting implications for metaphysics, but no relevance at all to the theory of rational choice.

I

An important part of the methodology of decision theory is testing principles of rational choice by performing thought experiments. Performing a thought experiment in this field consists in constructing a hypothetical situation in which someone is faced with a decision, seeing what decision intuitively seems rational, and comparing this decision to the one yielded by the principles of decision endorsed by some theory. This methodology, like any other scientific methodology, raises certain questions. What conditions, if any, must a hypothetical situation meet in order to be relevant as a test case for a theory of rational choice? Are there systematic errors one should be aware of when testing theories in this way? By focusing on a case where this methodology has not resulted in smooth sailing I hope to shed some light on these methodological issues.

The central point to notice in attacking these two questions is that it is impossible to describe the hypothetical decision situation without building into the description a great many normative facts. Otherwise the people described will be mere shadows of people. The normative facts that are relevant here are epistemic. For instance, agents may be described as assigning certain subjective probabilities to certain propositions, or as

having (or lacking) good reason for certain beliefs, or as knowing certain propositions and not others.

The normative facts can be included in two different ways. We can describe the people in great detail; we can describe their epistemic standards, their psychology, and all their evidence. Good novelists can do this well. Philosophers are typically less skilled at it, and, more importantly, philosophers must bring out very sharply the normative facts which are especially relevant to the purpose of the thought experiment. The other way to include the normative facts is to stipulate them.

When the normative facts of a hypothetical situation are stipulated to some extent I will call the thought experiment a "schema." In one commonly used example one stipulates that people in a certain hypothetical situation *know* that there is a gene which both inclines people to smoke cigarettes and disposes them to contract lung cancer. A "realization" of this schema is the result of filling out the description of the hypothetical people in a way that makes one or more stipulated normative facts become a plausible interpretation of the people's evaluative position in the hypothetical situation. For instance, one could describe the sort of evidence people have for an hypothesis about genes, smoking and cancer, and describe the scientific sophistication of the people, thereby making it plausible to interpret them as knowing "there is a gene which...." A realization of a thought experiment is any such "filling in" of a thought experiment schema. Obviously a schema can be more or less filled in; thus the term "realization" is convenient shorthand for more precise but unnecessary relational terminology. Usually there will be different realizations of a given schema.

When a thought experiment schema is realized, we may discover that the particular realization requires us to drop some stipulated normative fact other than the one which we have "realized away." We may even discover, upon making a number of tries at realization, that we cannot realize the schematic thought experiment at all without invalidating some of the stipulated normative facts. This would happen if we started with a thought experiment schema concerning a "completely moral ax-murderer" in a culture essentially like ours. Nobody would be taken in by such a case. But in more interesting cases the risk of making normative stipulations which will not survive realization is real. And in interesting cases there is always the risk that staying at the level of stipulation of a given thought experiment schema will blind us to the variety of realizations. This has happened with Newcomb's problem.

"Newcomb's problem" asks what we should do if the following events took place: A mysterious extra-terrestrial being descends and astounds us for a while with various exhibitions of inhuman intellectual prowess. As a curtain-closing act, the Being leaves a pair of boxes for each adult human, one transparent and containing a thousand dollars and one opaque. The Being announces that it has predicted what each of us will do when, as has

been scheduled, we choose to take either the contents of both boxes or the contents only of the opaque box. On the basis of these predictions, the Being has placed a million dollars in someone's opaque box if the person will choose just that box, and nothing in the opaque box if the person will choose both boxes. As time goes by, many thousands of people make choices, and the Being's rate of successful prediction is very high, around 90 percent (about nine of ten one-boxers have found a million, and nine of ten two-boxers have found only the obvious thousand). "Our" time has come to choose.

The Newcomb problem stipulates that (I) we know that a superior Being has made predictions about our choices far into the future and arranged money in boxes accordingly. It also stipulates that (II) we know a fair amount about the outcome of people making their choices but (III) we do not understand the causal processes that lead to these outcomes. I will argue that there is no way to realize the three stipulations of this thought experiment schema while at the same time having a thought experiment which is relevant to the theory of rational choice.

II

I shall use the term "causal decision theory" for the following variation on classical Bayesian decision theory.¹

Suppose an agent must choose among alternative actions A_1, \dots, A_n . And suppose that the agent has a way of partitioning the world into a set of logically exhaustive and logically mutually exclusive possible states S_1, \dots, S_m . For each alternative action A_i and possible state S_j , the action will result in an "outcome" O_{ij} . Assume the agent has, and recognizes, preferences among these outcomes in a way that can be codified by assigning a numerical utility value to each outcome. Finally, the agent knows his subjective probabilities for each conditional of the form "If I perform A_i , then S_j will obtain." These conditionals have as truth conditions the requirement that "If I perform A , then B will obtain" is true if and only if either my performing A brings about B , or B would obtain whether or not I performed A . Then the agent should proceed to assign an expected utility to each alternative action A_i by multiplying the probability of "If I perform A_i , then S_j will obtain" by the utility value of O_{ij} , and adding these results for each j .² If some action A_i receives the highest expected utility, the agent should perform it.

"Bayesian Decision theory" replaces the probability of the conditional "If I perform A_i , then S_j will occur" by the conditional probability of S_j on A_i in this rule.

I shall use the term "the dominance principle" for the decision rule which says that an agent should perform an action which weakly dominates its alternatives. Action A_i weakly dominates the rest of A_1, \dots, A_n if and only if

for some set of possible states $S_1 \dots S_m$, logically exhaustive, logically mutually exclusive and causally independent of $A_1 \dots A_n$, the outcome of A_i in some possible state S_j is preferred by the agent to the outcomes in S_j of all other actions in the set $A_1 \dots A_n$, and in no possible state is the outcome of any other action preferred to the outcome of A_i in that possible state.

Some philosophers believe that in the circumstances described by Newcomb's problem, we would find the dominance principle and Bayesian decision theory at odds in their recommendation, while causal decision theory would make the same recommendation as the dominance principle (which, moreover, is intuitively right).³ I will argue that this would constitute partial grounds for accepting causal decision theory *only* if the situation of Newcomb's problem is a choice situation we could in principle confront.

I shall argue later that there is a realization of Newcomb's problem which seems to be the problem these philosophers have in mind, but which falls short of doing what it is supposed to do since it is not a choice situation which we could confront. First I shall consider another realization of Newcomb's problem. Realized in this way it is a thought experiment which describes a situation which we could in principle confront, and it provides intuitive data useful in an investigation of decision theory. It does not, however, support causal decision theory against Bayesian decision theory. (Of course other thought experiments may do so.)

III

Consider the following realization of Newcomb's problem. The question is this: Suppose we have good reason to think an extra-terrestrial being has visited us for a while. (I stipulate as a constraint on the hypothetical situation that we have good reason, of whatever kind, for believing that this was an extra-terrestrial). Moreover we have some evidence that the being performed remarkable feats of prediction. For instance the being seems to have announced days or weeks ahead of time what certain people would do or say on some occasion. Most of these announcements turned out in retrospect to have been true, so far as "we" can tell. Perhaps the results have been reported in the news media, and confirmed in one or two cases by observers "we" know personally. Further, we have good reason to believe the being announced before leaving that boxes were going to be left for each of us, with money distributed in the boxes on the basis of predictions made by the being (details as usual). Finally, we have good reason to believe that nine out of ten one-boxers so far have found a million, while nine out of ten two-boxers have found empty opaque boxes. What should we do?

Presumably all attempts to explain how the being could manage such incredible predictions have been unconvincing. We have no idea how anyone or anything could do it. But we all know there are hustlers. It is

far more reasonable to believe in an extra-terrestrial hustler than to believe in an extra-terrestrial seer capable of predicting years or even months ahead the outcome of individual human decisions (an exception: if nearly everyone had been, say, a one-boxer we would have some reason to believe that this reflects a surprising regularity in human behavior which might somehow have been antecedently known). The being probably together with whatever humans are involved in the enterprise of placing and guarding boxes, supervising choices, and so on, cheats. Either the distribution of money is triggered by the physical act of selecting boxes, or it is triggered somewhat earlier by tip-off behavior exhibited by people about to choose.

Of course we will know that many people seem to believe in the prediction myth, but people can be taken in rather easily by simple, human conjuring or mind-reading trickery. If we are reasonable people, and we want to know in detail why the being's supposed "predictions" have been so "accurate," we should hope that the matter will be investigated by people competent to uncover conspiracies, detect the use of conjurers' apparatus, recognize mentalists' techniques, or contribute in some other way to uncovering large-scale fraud.

No reasonable person would think otherwise in this situation. And no additional details of the case will affect this conclusion. (For instance suppose we learn that the box-tenders were not selected by the being. They are randomly chosen West Point cadets. Reply: It is far more reasonable to believe West Point cadets can be bribed than to believe that some creature can predict free human behavior years in advance). Of course Newcomb's problem as ordinarily presented simply stipulates that we know the being has left money according to its predictions. But in the quite well-defined hypothetical situation I have described we would not know any such thing. Stipulation (I) cannot be sustained given the epistemic position we would actually find ourselves in.

Deciding what to do in this realization of the Newcomb problem would amount to deciding whether there seemed to be a way to beat the system. We can imagine variations on the problem where there is. Investigation by qualified people might show that the distribution of money is triggered by the actual behavior of box selection. Perhaps selections are announced in the hall outside the box room, whereupon giggling box-tenders go in and "bring out the contents of" the selected boxes. It would be clear that they reward one-boxers most of the time so we should take one box. This is not an interesting subcase.

The following is an interesting subcase: Physical selection of boxes is done by going alone into a room in which the boxes rest in plain view on an open table. The selection process is filmed by what seems to be an ordinary automatic movie camera, supposedly to prevent cheating by the person making the selection. (All of this is reported in the media and also

by several generally reliable friends who have made their selections). Before going in to make a selection, the candidate is interviewed for an hour by box-tenders.

Here it would be reasonable to think the box-tenders attempt to discover during the interview what the candidate will do, and succeed about nine times out of ten. They plant the money accordingly. There would be some probability that the physical act of selection triggers some very sophisticated apparatus which very rapidly affects an unobserved distribution of money. But there would be a much higher probability that during the pre-choice interview the interviewers decide what the candidate will do, and arrange the money before the physical act of box-selection occurs. Further details of the standard interview might or might not give us clues as to what the tip-offs are that the interviewers look for. If the interviewers typically stuck to asking general personal questions while studiously watching the candidate, it would seem they used some kind of "cold-reading" technique, the disciplined practical psychology that has long been an art form among professional fortune tellers. In that case we would be hard put to guess which behaviors of ours would be salient. On the other hand, the interviewers might be more transparent. For instance, they might frequently and unexpectedly ask, "How many boxes will you take," looking at stopwatches to measure the response time. In this case we could make speculative conjectures about their diagnostic method, but surely not with much confidence.

I shall assume these details can be filled in so as to make it plausible that we would be in the following epistemic position: (a) The probability is very high that behavior people exhibit during the interview tips off the interviewers what choice will be made, and the money is suitably arranged sometime before the physical act of box-selection. (b) The probability is very low that the trick is done by some mechanism which injects or removes money as a causal result of physical box-selection. (c) No hypothesis about the diagnostic methods of the interviewers is reasonable enough and detailed enough to justify the risk of trying to beat them at their game by faking just a few selected tip-off behaviors during the interview.

Our problem would be to behave as nearly as we can the way "typical" one-boxers behave during the interview. We might come up with a clever trick enabling us to do this, but actually take both boxes. Perhaps we could have ourselves hypnotized and given a complicated suggestion requiring us to "believe" we were going to take one box up to the last second, and then switch. We would need to have a great deal of confidence in this approach to justify the risk, and I shall consider only the case where we do not have such confidence.

The reasonable strategy for us to adopt would be an "indirect strategy." In my terminology, an indirect strategy for achieving an end involves setting oneself, or developing a firm intention, to perform an action *A* not

because one thinks it probable that *A* itself will cause the desired end, but because one thinks it probable that something or other one will do (perhaps unknowingly) as a result of having the intention to perform *A* will cause the desired end. When these conditions are satisfied I say that it is probable that action *A* will "bring about" the desired end.

Indirect strategies of a slightly different kind are available when the action is a long-term practice. It can be probable that one will achieve a desired end by engaging in long-term practice *A* because it is probable that something or other one does in the course of this practice causes the end. For instance, it might be a good indirect strategy, in this sense, to form oneself, or one's child, as an altruist so as to maximize one's own possible sources of pleasure by enabling oneself to take pleasure in the well-being of others.

A simpler example would be a situation where we know there is a strong statistical correlation between cigarette smoking and lung cancer, but where we also have expert scientific opinion that nothing in the smoke of cigarettes is carcinogenic. Various hypotheses would be available, including the hypothesis that the scientific opinion is wrong, and the hypothesis that some causal factors lead to a coincidence of cigarette smoking and a carcinogenic environment (or even a coincidence of cigarette smoking and a carcinogenic life-style). For instance, commercial advertising and social customs which encourage smoking also encourage smoking in bars. Drinking alcohol might directly cause a cancer liability, or, more treacherously, something about bars might be carcinogenic (think of it as a carcinogenic fungus found largely in bars). In ignorance of the correct hypothesis, but strongly preferring no smoking and no cancer to smoking and cancer, a good indirect strategy would be to stop smoking.

In the realization of Newcomb's problem I have described, an indirect strategy of the first kind is in order. In fact the problem is a clear illustration of the rationality of such strategies. The reasonable thing to do is form the firm intention of taking one box and stick to it. It is foolish to attempt an end run around accomplished tricksters who know their trickery as well as three-card monte hustlers know theirs. The suggestion might be made that we mimic the effect of the hypnotic suggestion scenario by sheer cleverness, first pretending to intend taking just one box, but finally snatching both. If there were evidence that others had tried such a ploy successfully it might be reasonable. But in the absence of such evidence, we should consider the usual fate of marks like us who try outfoxing a skilled mentalist and the powerful evidence that these box-tenders are skilled mentalists although they do not say so. We want to make them think we are one-boxers, but we do not know how to do this except by being one-boxers. There is another possibility, of course. It might be that at the interview the box-tenders psychologically manipulate candidates into making a particular choice (in conjurers' jargon, they "force" a given

choice). We cannot do anything about that, and must ignore this possibility in deliberation about a supposedly free choice.

Whenever we try to use an indirect strategy to achieve some end, we should realize that the point of the strategy is to insure that we perform some action or other which will be efficacious in securing the end, although we may not be able to conceptually individuate the action in a way that would enable us to directly undertake to perform it. In the case at hand, we can say that we must "appear to be prospective one-boxers to the interviewers" but we do not know what act-types at other levels of description will make us so appear. Our total evidence makes it very probable that if we steadfastly set ourselves to take one box, some pattern of behavior we engage in will result ultimately in there being a million dollars in the opaque box. We can call that pattern of behavior "action X ," in which case what we want to do is to perform action X . The action "taking one box" is explicitly cited in our formulation of the indirect strategy, whereas action X is not. But this should not make us think that we are deciding to perform the action "taking one box" *simpliciter*. We are deciding to perform action X by means of a strategy which requires us, among other things, to take one box.

Looked at this way our decision turns on whether to perform action X by means of the strategy of firmly and constantly intending to take one box. Let A_1 , be taking one box, A_2 be taking two boxes, m be the state of there being a million in the opaque box, $-m$ be the state of there being nothing in the opaque box, XA_1 be performing action X by firmly and constantly intending to perform A_1 , and let A_3 be some live alternative, such as using hypnosis to enable us to appear as one-boxers but switch at the last moment (nothing is affected by simplifying our position to one where this is the only live alternative). Then Bayesian decision theory requires us to assess the conditional probability of m and $-m$ on XA_1 and A_3 . The subjective probabilities of m on XA_1 and $-m$ on A_3 are high, whereas the probabilities of $-m$ on XA_1 and m on A_3 are low. Bayesian decision theory recommends XA_1 against A_3 . (I assume the reader can supply enough details of the case to justify numbers that work).

Causal decision theory requires us to assess the subjective probability of the conditionals:

- (i) If I perform XA_1 then m will occur.
- (ii) If I perform XA_1 then $-m$ will occur.
- (iii) If I perform A_3 then m will occur.
- (iv) If I perform A_3 then $-m$ will occur.

(i) and (iv) are very probable, while (ii) and (iii) are improbable. Causal decision theory recommends XA_1 against A_3 .

The dominance principle cannot be applied, since the condition that m and $-m$ be causally independent of XA_1 and A_3 is not known by us to be

satisfied. It is very probable that it is not satisfied, but we do not know this for certain. So in this realization of Newcomb's problem both Bayesian and causal decision theory recommend the same decision, and neither is inconsistent with dominance reasoning.

It might be argued that if we try to compare the acts A_1 and A_2 , rather than XA_1 and A_3 , causal decision theory recommends the wrong action. The argument goes like this: A_1 is the correct choice over against A_2 , since doing A_1 is required by a correct indirect strategy. But it is extremely likely in this hypothetical situation that the money will be distributed before either A_1 or A_2 is actually performed. So causal decision theory recommends performing A_2 . The modification this argument forces in causal decision theory is minor. We need only understand "brings about" in the truth conditions for the conditionals in the way I defined it in Section III, rather than as equivalent to "causes."

IV

The realization of Newcomb's problem discussed in the last section fills in enough details concerning the epistemic position of the people who appear in the thought experiment to justify interpreting them as satisfying stipulations (II) and (III). But this realization requires dropping stipulation (I). I have suggested that there is some modest philosophical payoff from considering this realization. Since it is a realization in which the people are epistemically like us it reflects what we would actually think and do if we were confronted with the claims that constitute Newcomb's problem together with the sort of evidence for those claims usually envisaged by writers on the subject. I tailored the final details so as to make up a useful thought experiment that illustrates the rationality of indirect strategies (of one kind) and that lets us see how Bayesian and causal decision theory should be brought to bear in such cases. Clearly our theories of rational choice can be fairly tested by this thought experiment. It concerns a situation which we could confront, and a theory of rational choice should recommend the correct action in any such situation where it applies at all.

It might be thought that this realization of Newcomb's problem is not the most interesting one. It is stipulation (I) that has especially gripped philosophers studying the problem. Even if the realization with people most like us forces us to abandon stipulation (I) we must allow some other realization which preserves it. We can describe a second realization in which (I) is preserved, but at great cost to its utility as a thought experiment in decision theory. Suppose we want to preserve the stipulation that we know the being has predicted what we will do and long ago distributed money accordingly. This requires that "we" be very different from the way we are culturally. It is better to imagine a civilization of people for whom it would be rational to accept, even after careful discussion and reflection, that the being had a way to make months-or-years-ahead predictions of

human decisions, rather than rejecting this in favor of the hypothesis of an elaborate deception. Call them the Gullibles.

The Gullibles must be unlike us in deep ways. One of us who endorsed the predicting-being hypothesis would do so in the face of (ignored) epistemically relevant experience of human nature: experience of human deceit and conspiracy on the one hand, and experience of the unpredictability of human behavior over long stretches of time on the other hand. One of *us* who did that would not be a Gullible. He would be an epistemically inconsistent person who assigned probabilities out of line with available evidence. We want to imagine true Gullibles, people who can assign a high probability to the predicting-being hypothesis consistently with their other assignments of subjective probability.

One way to do this is to suppose that the Gullibles don't have the concepts of fraud and deception, and are thus unable to frame the hypothesis that there is trickery and fraudulent misrepresentation in the Newcomb problem. I prefer another picture of them, one that maximizes similarities between them and at least some of us: Assume the Gullibles believe people sometimes have magical powers enabling them to see the future. The Gullibles also have concepts of fraud and trickery, but these concepts are narrower than ours. In the Gullible world-view a person can "trick" another by seducing him into doing something, for example, by making a false promise. But the Gullibles do not believe people can substantially manipulate the apparent causal order of nature. Certain itinerant gypsies among them perform card tricks and work elaborate confidence games with deceptive props, but Gullible conventional wisdom has it that these gypsies are magical seers. Gullible scientific psychology has evolved several imaginative theories of prescience, although hard-nosed critics among them question whether these theories are robustly testable.

Suppose someone goes up to a Gullible and claims to be able to predict by looking into the future what card the Gullible will randomly select from a deck. He writes something on a piece of paper, seals it in an envelope, and hands the Gullible a deck of fifty-two cards each of which is a queen of spades. The Gullible chooses a card without looking at the faces. The paper turns out to say "queen of spades." Presumably the Gullible will take this as good, although perhaps not absolutely conclusive, evidence that the other person predicted an essentially random event by seeing the future. But what is more important the Gullible will do this without further inquiry and without supposing that further inquiry might be in order. He would not bother to turn over the cards because he will not have a likely enough alternative hypothesis (except "it was luck," and that could be taken care of by further trials).

Gullibles would not, as we would, assign a low probability to prescience in card magic pending further investigation for fakery. This is a principled difference in the way they make epistemic evaluations, and it would affect

a great many judgements. When people claim to know that something will happen later, it is uncommon for us to investigate for fraud and deception. But this is either because little hangs on it, or because we have "trust" of a certain kind. Gullibles would not investigate for fraud and deception of this kind no matter what hung on it, and they would have no need for our concept of trust, recognizing nothing of significant practical importance to contrast it with. Thus the Gullibles will respond to Newcomb-problem evidence by endorsing the predicting-being hypothesis. They don't think at all likely, or even see, possibilities that we see and that we consider significant. They accept as live possibilities things that we rule out as absurd. These differences make for the differences in the assignment of prior probabilities brought by the Gullibles into the Newcomb problem.

A true Gullible is not like you and me, however gullible we may sometimes be through epistemic inconsistency. We would not and could not become Gullibles as a result of certain new experiences, not even as a result of watching a large silver sphere land in the back yard, followed by weeks of hearing and reading about so-called "astounding predictions." No such experience could blot out the experience of the world, including people, we have already accumulated. We could "become Gullibles" by gradual cultural evolution and we could have been raised in a Gullible culture in the first place. But we cannot become Gullibles as a result of confronting a novel situation and gathering information about it.

Let us consider, then, a realization of the schematic Newcomb problem in which "we" may be plausibly interpreted as rationally accepting the predicting-being hypothesis, so that if it is true we could argue that we know it, as stipulated in the schema (I)-(III). This realization (hereafter "the second") requires that we be Gullibles. Therefore, *we* cannot confront such a situation, in the sense of "cannot" described above. I suggest that theories of rational decision need only stand testing in hypothetical situations which *we* can confront, in this sense of "can." Otherwise we shall have to demand that such theories reflect the "rational" decision that would, intuitively, be made by people whose cultural epistemic standards are substantially unlike ours. This would be like demanding that moral theories survive testing in thought experiments about cultures where unhappiness is prized or social discord positively sought.

A natural reply is that if we consider sufficiently high-level epistemic standards we would see that the Gullibles need not differ from us. For instance, they, like us, are guided by the rule: before accepting an explanation, investigate whether alternative explanations better fit the facts. It is just that in some cases they don't see serious alternatives where we do. But all sorts of epistemic standards, high-level and low-level, will come into play in a given decision situation, and epistemic standards at this level of generality are compatible with all sorts of outlandish lower-level standards, for instance, taking statistical data which can be interpreted as

showing retro-causation as very likely indicating retro-causation. We would not expect a good theory of rational choice to imply and legitimate the decisions deemed rational by people who thought like this. Analogously, people who prize unhappiness might still coincide with us in moral standards at a high level of generality, such as "seek to maximize what is prized." Moral theories should not be tested against the moral judgements of such people as these.

The second realization of the Newcomb problem is intelligible. Unlike the first it cashes in stipulation (I) in the schema of the Newcomb problem. And also unlike the first it is a thought experiment irrelevant to decision theory.

A final remark: If no realization of a given thought experiment schema is evidence for one philosophical theory against another, then the thought experiment schema does not decide between these theories. Plausibly, the realizations of Newcomb's problem I have described exhaust the interesting possibilities. So, plausibly, Newcomb's problem provides no evidence for causal decision theory *vis-a-vis* Bayesian decision theory.⁴

NOTES

1. Versions of causal decision theory can be found in David Lewis, "Causal Decision Theory," *Australasian Journal of Philosophy*, vol. 59 (1981); Allan Gibbard and William Harper, "Counterfactuals and Two Kinds of Expected Utility," in C. A. Hooker, J. J. Leach, and E. F. McClennan (eds.), *Foundations and Applications of Decision Theory*, Volume 1 (Dordrecht, Holland: D. Reidel, 1978); Brian Skyrms, "The Role of Causal Factors in Rational Decision," in his *Causal Necessity* (New Haven: Yale University Press, 1980). The version I describe is essentially that of Gibbard and Harper.

2. The semantics for these conditionals is stipulated. The reader can translate them into the subjunctive if that seems preferable.

3. For example see Gibbard and Harper, p. 153.

4. I am indebted to Joseph Camp, John Forge, Richard Jeffrey, Gerald Massey, and Nicholas Rescher for helpful criticisms of earlier versions of this paper.