Divine Foreknowledge and Newcomb's Paradox

William Lane Craig
Used by permission of *The Journal of Philosophy* 85 (1988): 135-150.

SUMMARY

Newcomb's Paradox provides an illuminating non-theological illustration of the problem of divine foreknowledge and human freedom. We are to imagine a being with great predictive powers and to suppose we are confronted with two boxes, B1 and B2. B1 contains $1,000; B2 contains either $1,000,000 or nothing. We may choose either B2 alone or B1 and B2 together. If the being predicts that you choose both boxes, he does not put anything in B2; if he predicts that you choose B2 only, he puts $1,000,000 in B2. What should you choose? A proper construction of the pay-off matrix for the decision vindicates the one-box choice. If this is correct, then those who claim that God's knowledge is counterfactually dependent on future contingents foreknown by Him are likewise vindicated.

DIVINE FOREKNOWLEDGE AND NEWCOMB'S PARADOX

Undoubtedly the most provocative and elucidating illustration of the problem of theological fatalism is what has come to be known as Newcomb's Paradox. Originally the brainchild of William Newcomb of the University of California's Lawrence Livermore Laboratory, this puzzle was passed on to the philosophical public by Robert Nozick in 1969 and has generated such debate that one recent disputant speaks of current philosophy's "Newcombmania." [1]

The Puzzle Conditions

According to Nozick's account, we are to imagine a being in whose predictive powers we have enormous confidence; indeed, this being has never made an incorrect prediction of one's choices. Suppose then that we are confronted with two boxes, B1 and B2. B1 contains $1,000; B2 may contain either $1,000,000 or nothing at all. We are given the option of taking the contents either of B2 alone or of Bl and B2 together. Suppose, furthermore, that the following are true:

1. If the being predicts that you will take what is in B1 and B2, he does not put the $1,000,000 in B2.

2. If the being predicts that you will take only what is in B2, he puts the $1,000,000 in B2.

Nozick further stipulates that if one randomizes his choice, then the being does not put the

$1,000,000 in B2.

Now what is one to do? There are two "plausible looking and highly intuitive arguments" which require different decisions. [2]

According to the first argument, one reasons: if I take what is in both boxes, the being will almost certainly have predicted this and left B2 empty. On the other hand, if I take B2 alone, he will have put the $1,000,000 in it. So I shall take B2 alone. According to the second argument, one reasons: The $1,000,000 is already sitting in B2 or it is not, and which situation obtains is already fixed and determined. If the being has already put $1,000,000 in B2 and I choose both, then I get $1,001,000. If he has not, then I get $1,000. Either way I get $1,000 more than by taking B2 alone.

Nozick seeks to augment the force of each argument by means of the following further stipulations: With regard to the first argument, suppose that all previous people who chose B2 alone got the $1,000,000. All the "shrewdies" who followed the recommendation of the second argument wound up with only $1,000. It would be rational for a third person to bet, giving high odds, that if you take both boxes, you will get only $1,000. In fact, if the award of the money were delayed, even you ought to offer such a bet! With regard to the second argument, suppose that B1 is transparent so that you can see the $1,000 sitting there. The $1,000,000 is already either in B2 or not. "Are you going to take only what is in B2?" asks Nozick. Suppose, furthermore, that B2 has a transparent side facing a third person, who can therefore plainly see whether B2 is empty or not. The money is not going to appear or disappear. "Are you going to take what is only in the second box, passing up the additional $1,000 which you can plainly see?" Nozick demands. Moreover, whatever the state of B2, this third person is hoping that you will take both boxes, and you know that he must be so hoping. "Are you going to take only what is in the second box," asks Nozick incredulously, "passing up the additional $1,000 which you can plainly see and ignoring my interally given hope that you take both?" [3] In the face of these two arguments, what should one do?

Theological Implications

Nozick originally presented the paradox as a dilemma within the realm of decision theory, but it is of obvious interest for the metaphysician and philosopher of religion as well. For it is almost irresistable to identify Nozick's "being" with an omniscient God and to construe Newcomb's Paradox as an illustration of the problem of theological fatalism. In a later piece Nozick himself approves of the identification of the Being (now capitalized) with God. [4] Bar-Hillel and Margalit make the connection with fatalism when they assert that if such a being existed, then he would contribute just the kind of evidence that would disprove one's illusion that he can choose arbitrarily between the boxes: ". . . the facts really imply that there is no free choice, but the illusion of free

choice remains, and one has to behave as if free choice exists." [5] Similarly Don Locke: ". . .once the Predictor has made his prediction, that prediction becomes fixed and unalterable: having made the one prediction, it is no longer possible for him to make the other. So given that the Predictor is absolutely infallible, it is at the time of choosing equally impossible, and in just the same sense, for the Chooser to make any choice other than that predicted." [6] According to Locke, the fact that the Predictor will no doubt have correctly predicted my choice as he has all others "gives me every reason to think that I have no choice in the matter at all, or that if I do have any freedom, it is a freedom I am unlikely to exercise." [7] Schlesinger, on the other hand, thinks that the fatalistic implications of Newcomb's Paradox succeed in showing that an infallible and omniscient Predictor cannot exist. [8] Similarly, an exultant Isaac Asimov proclaims:

I would, without hesitation, take both boxes . . . I am myself a determinist, but it is perfectly clear to me that any human being worthy of being considered a human being (including most certainly myself) would prefer free will, if such a thing could exist. . . Now, then, suppose you take both boxes and it turns out (as it almost certainly will) that God has foreseen this and placed nothing in the second box. You will then, at least, have expressed your willingness to gamble on his nonomniscience and on your own free will and will have willingly given up a million dollars for the sake of that willingness-itself a snap of the finger in the face of the Almighty and a vote, however futile, for free will. . . And, of course, if God has muffed and left a million dollars in the box, then not only will you have gained that million, but *far more imponant* you will have demonstrated God's nonomniscience. [9]

Unwilling to abandon either divine foreknowledge or human freedom, Dennis Ahern concludes from his analysis of Newcomb's Paradox that the problem of foreknowledge and freedom remains an unresolved paradox. For it is equally implausible to believe either

3. One has control over God's past beliefs without recourse to the objectionable notion of backward causation

or

4. An action otherwise free becomes not free simply because it is foreknown or predicted.

But the falsity of (3) implies the truth of (4) and the falsity of (4) implies the truth of (3). Thus, if infallible foreknowledge existed, ". . .we should have sound reasons for believing it would not have a bearing on whether an action was performed freely *and* there would be no freedom of action." [10]

What may be said to this purported challenge of Newcomb's Paradox to divine foreknowledge or

human freedom? To begin with, it seems that we can safely dismiss Ahern's middle way between the dilemma's horns. For what Ahern has left us with is not a paradox, but an antinomy. If correct, his reasoning has demonstrated that the assumption of divine foreknowledge entails contradictory propositions concerning the freedom of foreknown actions. Therefore, the initial assumption which generated the antinomy must be rejected. Accordingly, Ahern should side with Schlesinger and Asimov in rejecting divine omniscience.

The alleged alternatives, then, with which Newcomb's Paradox confronts us are a denial of divine foreknowledge or a denial of human freedom. The incompatibility of these two assumptions is thought to be demonstrated by the fatalism implicit in the Newcomb game. The issue, therefore, is whether Newcomb's Paradox entails fatalism.

Nozick's Dilemma

Perhaps the best way to get at this issue is to return to the original dilemma posed by Nozick for decision theory. According to the *Expected Utility Principle,* among those actions available to a person, he should perform that action with maximal expected utility. According to the *Dominance Principle,* if there is a partition of states of the world such that relative to it action *a* weakly dominates action *b*, then *a* should be performed rather than *b*. Now these two principles seem to come into conflict in Newcomb's Paradox. We may construct the following pay-off matrix for the Expected Utility Principle:

		Being	
		A	B
		predicts agent will take B2 alone	predicts agent will take B1 & B2
Agent	i. takes B2 alone	$1,000,000	$0
	ii. takes B1 & B2	$1,001,000	$1,000

According to this principle, we may calculate the expected utility of the agent's respective actions

by muliplying each of its mutually exclusive outcomes by the probability of each state's obtaining and adding these products together. Given a probability of .9 for the Being's prediction's being accurate, the expected utility of action (i) is (.9 x $1,000,000) + (.1 x $0) = $900,000. The expected utility of action (ii) is (.1 x $1,001,000) + (.9 x $1,000) = $101,000. On this principle, one should choose to do (i). But according to the Dominance Principle, if the world is divided into various states and some action *a* is best in one state and at least equal in all the others, one should choose to perform *a*. But in this case we have such a partition of the world into states *A* and *B*, determined by the Being's predictions. Here action (ii) is strongly dominant, for in either case one acquires $1,000 more than he would by performing action (i). So one ought to take both boxes.

Now it is often pointed out, for example by Cargile, Olin, and others, that for the Dominance Principle to be valid, the states of the world must be causally and probabilistically independent of the actions to be taken. [11] That is to say, if performing action (ii) in some way brings about or renders more probable state *B,* for example, then the principle no longer applies. States *A* and *B* are probabilistically independent of actions (i) and (ii) if the probability of *A* given that (i) is taken is the same as the probability of *A* given that (ii) is taken, and likewise for *B*. In the Newcomb situation, however, the probability of *A* or *B*'s obtaining is not independent of whether the agent chooses (i) or (ii).Therefore, the dominance argument fails.

But Nozick is ready with a response. [12] He furnishes the following example of a situation in which the states are not probabilistically independent of the actions and yet the Principle of Dominance clearly applies. Suppose person *P* knows that either person *S* or *T* was his father. *S* had a fatal hereditary disease, but *T* did not. If *S* was *P*'s father, then *P* will also die of this disease; if *T*, then he will not. Now this disease makes one intellectually inclined. *P* is deciding whether to go on to graduate school or become a baseball player, and he slightly prefers the academic life. Let *w* = *P* is briefly an academic and then dies; *x* = *P* is an academic and *Z* = *P* is briefly an athlete and then dies; =*P* is an athelete and normal. Accordingly we can construct the following matrix, assigning preference values to *w, x, y, z*.

		Father	
		A	B
		S is *P*'s father	*T* is *P*'s father

Son	i. goes to grad school	*w* (-20)	*x* (100)
	ii. plays baseball	*y* (-25)	*z* (95)

The Dominance Principle tells P to choose (i). But in that case, be probably has the disease. So the Principle of Expected Utility would advise him to choose (ii). But this latter recommendation, says Nozick, is "perfectly wild." The probabilities favor (ii), but which state obtains is already fixed and determined and does not depend on *P*'s action. By choosing (ii), *P* does not make it less likely that *S* is his father nor make it less likely that he will die of the disease. Thus, ". . . in situations in which the states, though not probabilistically independent of the actions, are already fixed and determined, where actions do not affect whether or not the states obtain, then it *seems* that is legitimate to use the dominance principle. . . " [13] Yet even then it is not so much the fact that the states are fixed and determined that is critical, he adds, but whether one's actions affect which one is actual. For in the Newcomb situation, the prediction could be made and the choice taken and only then the money placed in the boxes on the basis of the prediction. "This suggests that the crucial fact is *not* whether the states are already fixed and determined, but whether the actions *influence* or *affect* which state obtains". [14] Where such influence exists, one should always maximize utility.

Divine Foreknowledge and the One-Box Strategy

Now in the conditions originally laid down in Newcomb's Paradox, no such influence exists. That is to say, contrary to the impression given by several writers, the being of Newcomb's Paradox did not make his predictions on the basis of precognition. On Nozick's formulation, Newcomb's Paradox is analogous to the situation described in the case of *P*'s deciding to study or play sport. The decision is wholly independent of the state which obtains. But once the Being is identified with God, the picture changes radically: for God's prediction is based on precognition of the decision, or in the language of theology, foreknowledge. In this case the actions and the states are not independent, for God predicts what He knows one will do. Hence, Nozick admits, ". . . if one believes that the way the predictor works is by looking into the future; he, in some sense, sees what you are doing, and hence is no more likely to be wrong about what you do than someone else who is standing there at the time and watching you, and would normally see you, say, open only one box, then there is no problem. You take only what is in the second box." [15] In fact, as Plantinga observes, [16] in the case of divine foreknowledge there is a logically demonstrative argument for the one-box strategy of the form A $\Box\!\!\rightarrow$B; (A & B) $\Box\!\!\rightarrow$C; therefore A $\Box\!\!\rightarrow$C:

5. If one were to take B1 and B2, then God would have believed that one would take B1 and B2.

6. If one were to take B1 and B2 and God believed that one would take B1 and B2, then God would have put nothing in B2.

7. If one were to take B1 and B2, then God would have put nothing in B2.

A parallel argument proves that if one were to choose B2 alone, God would have put $1,000,000 in B2. Thus, given the puzzle conditions, the only rational choice is to choose B2 alone.

Objections to the One-Box Strategy

Backward Causation

Now several philosophers, such as Mackie and others, have Objected that such an account of the Being's predictive ability entails the dubious thesis of backward causation. [17] According to Mackie, taking only one box would be justified if there occurs an extreme form of backward causation according to which the causal lines are drawn backward in time from the choice to the prediction and then forward from the prediction to the placing of the contents in the box. This analysis, however, seems to rest upon a misunderstanding in which the causal relation between an event or thing and its effect is conflated with the semantic relation between a true proposition and its corresponding state of affairs. For if at *tn* I choose B2 alone, then the proposition "*W* chooses B2 alone" is true at *tn* because of the semantic relation which obtains between a true proposition and the corresponding state of affairs which makes it true; by the same token " *W* will choose B2 alone" is true prior to t*n*. "*W* chose B2 alone" is true subsequent to *tn*, and " *W chooses* B2 alone at *tn*" is omnitemporally true. The relation obtaining between a true proposition and its corresponding state of affairs is semantic, not causal. Now God, knowing all true propositions, therefore knows the true future contingent proposition concerning my choice of the boxes. Again no causal relation obtains here. Hence, the charge of backward causation seems entirely misconceived: we have simply the semantic relation between true propositions and their corresponding states of affairs and the divine property of knowing all true propositions. Nozick remarks that he employed terms such as ''influence,'' ''affect,'' and so forth, without paying much attention to technical precision. [18] Now we can see more clearly that in the case of divine foreknowledge the "influence" exercised by the agent's choice over the Being's predictions is not a retro-causal influence, but rather the supplying of the truth conditions for some of the future contingent propositions known by God. Since the Being's predictions are made on the basis of his knowledge of such future contingent propositions, states *A* and *B* are not independent of actions (i) and (ii) and therefore the Principle of Dominance is in this case invalid.

Backtracking Counterfactuals

Objections to Backtracking Counterfactuals

It may still be objected that such an analysis is counterintuitive and paradoxical. It is incredible that something one does now could affect what God believed in the past such that were one to act differently God would have believed differently and that given that God did believe that one will do something one is nonetheless free to do something else. The problem here lies with (5) and its parallel

8. If one were to take B2 alone, then God would have believed that one would take B2 alone.

Ahern regards this as paradoxical because in choosing B2 alone one is giving up, from the perspective of past facts, a sure $1,000. For in choosing B2 alone, one *knows* that there is in fact $1,001,000 in the two boxes. Choosing B2 alone is the right strategy, but one must live with the "uncomfortable knowledge" that at the time of choosing B2 alone God's belief is "unalterably tucked away in the past" and there is really $1,001,000 in the boxes. [19] After choosing B2 alone one must be prepared to say, "If I had chosen both boxes, I would not have gotten the $1,001,000. " But an opponent might retort, "Of course you would have, since it was there! Therefore, you must not have been free to choose both." This is in fact precisely the reaction of Schlesinger, who claims that the one box strategy is self-contradictory. [20] He reiterates Nozick's argument concerning the well-wisher who can see the contents of the boxes and sincerely hopes that one will choose both. If the one box stragtegy is correct, it is not in my best interests to follow the advice of a sufficiently intelligent and well-informed well-wisher. But if a well-wisher is someone who invariably advises me to do what is in my best interests, then this amounts to saying that it is not in my best interests to do what is in my best interests, which is self-contradictory. Moreover, one may argue that the choice of both boxes is a better choice because the Predictor himself, having sealed the contents inside, knows the choice of both boxes is superior. [21] He knows that the chooser cannot place himself in a less favorable position by choosing both. If asked, "Would the chooser lose anything should he attempt to choose both?" the Predictor would have to say, no. He may believe, however, that choosing both "is not open" to the chooser and assert correctly that "If the agent were to choose both, he would be better off."

Backtracking Counterfactuals and an Inerrant Predictor

Now if we assume that God's precognitive beliefs are merely actually infallible, that is, inerrant in the actual world, then the adjudication of this issue will depend on whether we follow David Lewis in insisting on a standard resolution of vagueness in comparing the possible worlds in which the various counterfactuals involved in Newcomb's Paradox are true, or whether we will allow so-

called "backtracking" counterfactnals in our resolution of vagueness. According to Lewis's point of view, the standard method of resolving vagueness in assessing similarity between possible worlds involves preserving as intact as possible the same past history in the respective worlds; thus there is a temporal asymmetry in counterfactual dependence: if the past were different, present or future events might be otherwise in the closest possible world, but if the present or future were different, we cannot say that the closest worlds are ones in which past events would be otherwise. [22] Lewis acknowledges that some contexts may require a special resolution of vagueness, but he elsewhere makes clear that the Newcomb situation is not one of them. [23] In that situation backtracking counterfactuals are not allowed; accordingly it is true that

9. If I took only one box, I would be poorer by $1,000 than I will be after taking both.

According to Lewis, the "essential element" here is the fact that whether or not I get the $1,000,000 is causally independent of what I do now. [24]

Horgan, on the other hand, argues that the Newcomb situation is precisely one in which a special resolution of vagueness employing backtracking counterfactuals should be employed. [25] The one box solution gives top priority to maintaining the Being's accuracy in the nearest possible world. The closest world in which I take both boxes instead of one will be a world in which the being correctly predicted this and therefore left B2 empty. This means that the past history of that world will be slightly different from that of the actual world, in which I choose B2 alone; but it is more important to preserve the Being's accuracy than a perfect historical match in specifying the closest possible world. Under the special resolution of vagueness, (9) is false; on the contrary (5) and (8) are true.

Horgan attempts to break the deadlock between these two competing resolutions of vagueness by arguing that only the special resolution is pragmatically appropriate in this situation. Given my overwhelming conviction of the being's predictive accuracy, I am virtually certain that the actual world is a world in which the being has accurately predicted what I shall do. Hence, worlds in which the being errs ought to be regarded as irrelevant for the purposes of decision-making. Thus, the special resolution is pragmatically appropriate because the closest world in which I do action (i) is one in which *A* obtains and the closest world in which I do action (ii) is one in which *B* obtains. No corresponding meta-level argument exists for the standard resolution. All the defender of the standard resolution can do is to appeal again to the intuition that

10.*Either* I would get $1,001,000 if I chose both boxes and I would get $1,000,000 if I chose B2 alone, *Or* I would get $1,000 if I chose both boxes and I would get $0 if I chose B2 alone.

But (10) is true only if one already accepts the standard resolution. By contrast, the defender of the

special resolution has an independent justification for adopting backtracking counterfactuals, namely, I am virtually certain, independent of any beliefs I have concerning whether I shall do (i) or (ii), that a world in which the being errs is not actual. Horgan's defense of backtracking counterfactuals in this connection would seem all the more conclusive when the being is God. For now are absolutely certain that the prediction is not in error.

Isaac Levi has, however, objected to Horgan's reasoning, [26] charging that Horgan fallaciously concludes from

11.The probability is high that the agent will choose both boxes if the being will so predict

to

12.The probability is high that if the agent will choose both boxes, then the being will so predict.

Levi grants that we should choose B2 alone if the probability is high that if the agent will pick both boxes, then the being will predict this. But in the original Newcomb's Paradox, one is not warranted in assuming (12). Hence, Levi has been characterized as a "no-boxer, " since on his view the initial conditions laid down in the Newcomb Problem are underdetermined in not specifying whether both sets of conditional probabilities are high, so that neither choice can be judged to be rationally preferable. [27]

In a recent reply to Levi, [28] Horgan concedes that according to the usual formulation of the paradox it is only laid down that most of the being's two-box predictions have been correct, as have most of his one-box predictions, and that the agent knows this; but that this only shows the probability of a two-box choice is high on a two-box prediction and the probability of a one-box choice is high on a one-box prediction. Levi is correct that these probabilities can be high even if the converse probabilities are not both high. But Horgan asserts that he construes the Newcomb situation to involve implicitly some further conditions: (i) that almost all of those who have chosen both boxes in the past have received $1,000; (ii) that almost all of those who have chosen only the second box have received $1,000,000; and (iii) that the agent knows these facts. In other words, Horgan takes it to be *built into* Newcomb's Paradox that for the agent the probability is high that if he chooses B2 the being will have predicted this and the probability is high that if he chooses B2 and B2 the being will have predicted that. This is the reasonable and natural way to construe the problem because only then do the paradoxical conflicts arise. In any event, he concludes, "I suppose there is no prior fact of the matter as to whether the implicit conditions just mentioned are part of Newcomb's Problem or not. Very well, I hereby *stipulate* that the conditions are included, as I used the term 'Newcomb's Problem'." [29]

Campbell complains that if one makes Horgan's stipulations, then Newcomb's Paradox cannot be used to *test* one's decision principle; one simply relies on it. The original underdetermined problem is too indeterminate to argue for either decision principle, and if one makes additional stipulations to remove this indeterminacy, he imposes so much structure on the problem that it can no longer serve as an intuitive confirmation of the principle which one favors. [30] But Campbell's dilemma seems dubious to me. In the first place, even if one makes Horgan's stipulations, the success of the one-box argument is going to depend on the cogency of Horgan's meta-level arguments concerning the permissibility of a special resolution of vagueness, and, as we shall see, Horgan himself seems to think there is plenty of room for debate there. (In any case, Campbell's point would not affect the importance of Newcomb's Paradox for the philosopher of religion, as opposed to the decision theorist, for our interest in the problem concerns its implications for theological fatalism.) But, secondly, is it in fact the case that these stipulations were not included in Nozik's original formulation of the problem? A good case can be made that they were. As for conditions (i) and (ii), Nozick explicitly states that the being has never made an incorrect prediction of one's choices. He himself stipulates that all previous people who chose B2 alone got the $1,000,000 and that all the "shrewdies" who chose B1 and B2 wound up with only $1,000. And as for condition (iii), the very puzzle arises because the agent is aware of the being's enormously successful previous track record. Hence, Nozick asserts that it would be rational for the agent himself to offer a bet, giving high odds, that if he takes both boxes he will get only $1,000. Thus, it would seem that the Newcomb Problem is not underdetermined after all. Of course, no-boxers may find the underdetermined version of the paradox more intriguing (though finally inconclusive), and that is a philosopher's privilege; but he ought not then to claim that he is discussing the genuine Newcomb Problem, for his version would seem to be an attenuation of the original.

Now even given these conditions, the success of the one-box strategy is going to depend on the admissibility of a special resolution of vagueness; for invariant two-boxers like Lewis and Gibbard and Harper insist that the rational choice is to choose both boxes even if one knows that in so choosing he will get only $1,000, since it is also true that if one were to choose only one box, he would be $1,000 poorer than he shall be after choosing both. But Horgan claims to have offered a meta-level argument for preferring a non-standard resolution of vagueness so that the two-boxer's counterfactual claim is false. Eells has, however, charged that Horgan's argument for a one box choice is as circular as the two-boxer's appeal to (10). [31] For in stating that I am virtually certain, independent of any beliefs I have concerning whether I shall do action (i) or action (ii), that a world in which the being errs is not actual, I presuppose the backtracking resolution of vagueness. For the independence spoken of here must mean that the above outcome is counterfactually independent of whether (i) or (ii) is performed, and I can have such certainly only if a backtracking resolution is presupposed. Hence, the argument begs the question. But Horgan responds that

Eells has misconstrued the independence spoken of here. [32] Horgan is not saying that my certainty of getting either $1,000,000 or $1,000 is counterfactually independent of how I choose, but that it is independent of any beliefs I have about how I shall choose; that is to say, the agent in the Newcomb situation has a set of premisses which implies that it is highly probable that a world in which one receives $1,000,000 or a world in which one receives $1,000 will become actual, and this set of premisses includes no propositions about the probability of one's choosing (i) or the probability of one's choosing (ii). This notion of independence involves no counterfactuals, and so the argument is not circular.

Eells attempts to rehabilitate the two-box argument as well, proposing a new *C*-resolution of vagueness according to which all the differences between a closest world in which one chooses (ii) and the actual world must be causal results of the occurrence of (ii) in the closest (ii)-type world. Under such a resolution, a one-box strategy would require backward causation. So if we give high priority to avoiding backward causation, the two-box choice is always preferred. [33] But surely now it is Eells who is making question-begging stipulations. Why should we adopt a *C*-resolution? Why cannot the closest world include those with some difference due to a non-causal counterfactual dependence upon an action? Why should we construe counterfactual dependence as causal? Why regard a possible world as the closest (ii)-type world only if I would actualize it (in the causal sense) by choosing (ii), rather than regarding a world as the closest (ii)-type world only if it would be actual were I to choose (ii)? As Horgan notes, Eells's argument is not really a meta-level argument at all, but just another ground level proposal without higher justification. [34]

Nonetheless, Horgan now reluctantly admits that the debate between one-boxers and two boxers is a "hopeless stalemate." [35] For the two-boxer can consistently refuse to seek a meta-level defense of the standard resolution which does not itself appeal to counterfactuals. The two-boxer need not accept the normative principle that one ought to adopt a meta-level defense which avoids reference to counterfactuals. He can simply cite (10) in support of the standard resolution, concede that his meta-level normative premiss is equivalent to his ground level premise that one ought to choose both boxes, and then say that he simply regards both these premisses as true.

Now it seems to me that Horgan concedes too much. For he allows the two-boxer to reject the meta-meta-level claim that

13. For purposes of choosing a vagueness-resolution to adopt in practical decision making, one ought to act on the basis of a meta-level normative premiss that makes no appeal to counterfactuals; for the question of how to resolve the vagueness of counterfactuals is precisely what is at issue.

But why let the two-boxer get away with this? It seems entirely reasonable and plausible to accept (13), so why should the two-boxer be exempt? Indeed, Horgan himself provides a striking practical incentive for adopting (13) in envisaging a Newcomb situation in which a two-box choice leads to one's death, so that the two-boxer's refusal to accept (13) results in the adoption of a decision principle which proves personally disastrous. Surely this result suggests that (13) is correct, since refusal to accept it as normative may result in adopting a personally injurious decision principle which has no justification beyond itself. If (13) is correct, then the two-boxer's argument is circular.

But even if Horgan is correct in conceding that the justification of the two-box strategy is not viciously circular, that does not therefore mean that the debate is stalemated. For the normative premisses used to justify the two-box choice could be simply false, if not circular. Given the cogency of the meta-level argument for the one-box strategy, the normative premisses of the two-box argument must be false. And Horgan's reasoning in defense of one-box choice does seem compelling if we reconstruct the payoff matrix used to determine one's choice. For Horgan's analysis closely resembles that of Ferejohn, who argues that in a decision-theoretic context, the payoff matrix ought to be formulated, not in terms of the being's predicting this or that choice, but in terms of the Being's predictions' being correct or incorrect:

		State of Nature	
		A	B
		Being predicts correctly	Being predicts incorrectly
Agent	i. takes B2 alone	$1,000,000	$0
	ii. takes B1 & B2	$1,000	$1,001,000

Here there is no dominant choice for the agent; therefore, he must maximize expected utility. Given one's overwhelming conviction of the being's correctness, the proper choice is to take B2 alone. Brams points out that this representation of Newcomb's Paradox depends on the assumption that the being has no control over whether *A* or *B* obtains. This is not the same as his

being able to correctly predict one's choice, for he almost surely can. Rather (if I understand Brams correctly) it is a matter of whether the being can control when he is correct; perhaps he just is correct most of the time, but not by his design. It just happens that most of his guesses come out right. In such a case, Ferejohn's matrix is the one to use. On the other hand, if the being is able to control whether *A* or *B* obtains, then one is not playing against a passive state of nature; therefore, Nozick's matrix is correct, with its conflict between the Dominance and Expected Utility Principles, though it is incomplete because it assigns no preferences for *A* or *B* on the being's part. Observing that there is nothing in Nozick's original statement of the paradox which suggests that the Being has control over the correctness of his predictions-that is, his predictions are not based on what the agent will do-Brams asserts that Ferejohn's matrix is appropriate. [36] Horgan's emphasis on preserving the Being's correctness would therefore be justified and the one-box strategy vindicated. This defense of the one-box strategy does not run afoul of Levi's or Lewis's objection because what the being predicts does not enter into the matrix. Therefore, the two-box strategy must be rejected.

Now if the Being is God, Ferejohn's matrix would be appropriate if we take the predictions to represent God's true beliefs, for God presumably entertains solely true beliefs by nature, not by choice. On the other hand, in a game situation God could deliberately give false predictions to make things more interesting. In that case, Nozick's original matrix ought to be used. But then surely we would be justified in assuming that the being in Nozick's original paradox was not trying to give false predictions; his preference was to be correct on every try. If this is the case, then in preferring to give correct predictions and being able to control when he does so, God will predict *A* only when the agent chooses (i) and will predict *B* only when the agent chooses (ii). Hence, the one-box strategy is once again vindicated. Whether we use Ferejohn's matrix or Nozick's, then, a special resolution of vagueness is warranted.

In any case, when the predictor is God, the two-box strategy is plainly the wrong answer, since the agent's choice and God's prediction are not unrelated, as in the original Newcomb Problem, but are related by precognition. The predictions are based on foreknowledge of the choices, and so even invariant two-boxers concerning the original Newcomb's Paradox must concede that since, when the predictor is God, the predictions are determined by the choices, a special resolution of vagueness is in order and the rational choice is to choose one box, even though the contents of both boxes are fixed and determined at the time of choosing.

Applying this analysis to Schlesinger's objections, it becomes apparent that his well-wisher was presupposing a standard resolution of vagueness. Had he been sufficiently well-informed, he would have wished that the agent choose B2 alone. Or rather, seeing the money in B2 he would rejoice that his friend is going to choose B2 alone; or seeing no money in B2 he would regret that

his friend is about to blunder by choosing both boxes. In a sense, wishing, except in the sense of regret, is inappropriate for the well-wisher, since a moment's glance informs him what the future will be, and therefore hoping that one will do something has no place. Schlesinger's Predictor, too, presupposes the standard resolution of vagueness. Otherwise, in answer to the query as to whether the agent would lose something by choosing both boxes, he would reply, "Yes, he *would;* but he will not choose both and therefore I have sealed up the $1,000,000. If he *were* to choose both, he would be worse off because I would not have placed $1,000,000 in B2. But happily he will not. " In fact, a sufficiently well-informed chooser, were the contents of the boxes exposed also to his view prior to his choice, would realize what his choice will be. Had he resolved to take only one box, he would not upon seeing the contents of both boxes before him suddenly change his mind, tempting as that might be, for he would know that were he to choose both boxes, it would turn out that the million he had seen was, after all, hallucinatory or in some way unreal.

Backtracking Counterfactuals and an Essentially Infallible Predictor

If we hold that the predictor is not merely inerrant, but infallible, then in fact no appeal to a special resolution need be made. For most theists hold that God's foreknowledge is not merely inerrant but essentially infallible. Therefore, worlds in which God's prediction errs are not even possible. On this basis the standard resolution alone suffices to ensure a one-box choice, for the only possible worlds in which I choose two boxes are worlds in which I get only $1,000. No worlds in which I choose two boxes exist in which the past history of the actual world, in which I choose one box, remains intact. In all worlds in which I choose both boxes, God predicts this and leaves B2 empty. Thus, (5) and (8) are entirely vindicated.

Newcomb's Paradox and Freedom

But does that mean that in the actual world I am not free to choose otherwise, as Ahern alleges? Are we left with the theological fatalism which prompted our inquiry? By now the answer should be clear. It is I by my freely chosen actions who supply the truth conditions for the future contingent propositions known by God. The semantic relation between a true proposition and the corresponding state of affairs is not only non-causal, but asymmetric, The proposition depends for its truth on which state of affairs obtains, not *vice versa.* Were I to choose otherwise than I shall, different propositions would have been true than are, and God's knowledge would have been different than it is. Given that God foreknows what I shall choose, it only follows that I shall not choose otherwise, not that I could not. The fact that I cannot actualize worlds in which God's prediction errs is no infringement on my freedom, since all this means is that I am not free to actualize worlds in which I both perform some action *a* and do not perform *a.* The Newcomb Paradox provides no reason for thinking that from

14. There is $1,000,000 in B2 because I am going to choose B2

and

15. Were I going to choose B1 and B2, the $1,000,000 would not be in B2,

it follows that

16. I am not free to choose B1 and B2.

As Cargile puts it, "The player is free-he just cannot escape being 'seen' making his free choice." [37] Admittedly one may feel uncomfortable about the fact that in choosing B2 alone one commits oneself to the existence of $1,001,000 in the boxes. In this sense, a feeling of strangeness remains. But discomfort is not paradox, nor does a feeling of strangeness warrant a fallacious inference to fatalism.

Conclusion

Newcomb's Paradox thus serves as an illustrative vindication of the compatibility of divine foreknowledge and human freedom. A proper understanding of the counterfactual conditionals involved enables us to see that the pastness of God's knowledge serves neither to make God's beliefs counterfactually closed nor to rob us of genuine freedom. It is evident that our decisions determine God's past beliefs about those decisions and do so without invoking an objectionable backward causation. It is also clear that in the context of foreknowledge, backtracking counterfactuals are entirely appropriate and that no alteration of the past occurs. With the justification of the one box strategy, the death of theological fatalism seems ensured.

Footnotes

[1]

Isaac levi, "A Note on Newcombmania," *Journal of Philosophy* 79 (1982): 337-42. Further indication of the philosophical interest in this puzzle is the very fine anthology edited by Richmond Campbell and Lanning Sowden. *Paradoxes of Rationality and Cooperation: Prisoners' Dilemma and Newcomb's Problem* (Vancouver: University of British Columbia Press, 1985). See especially their comprehensive bibliography.

[2]

Robert Nozick, "Newcomb's Problem and Two principles of Choice," in *Essays in Honor of Carl G. Hempl,* ed. Nicholas Rescher, Synthese Library (Dordrecht, Holland: D. Reidel, 1969), p 115.

[3]

Ibid., p. 116.

[4]

Robert Nozick, cited in Martin Gardiner, "Mathematical Games," *Scientific American,* March 1974, p. 102.

[5]

Maya Bar-Hillel and Avishai Margalit, "Newcomb's Paradox Revisited," *British Journal for the Philosophy of Science* 23(1972): 301.

[6]

Don Locke, "How to Make a Newcomb Choice," *Analysis* 38(1978): 21.

[7]

Ibid.,p. 23. Cf. Don Locke, "Causation, Compatibilism and Newcomb's Paradox," *Analysis* 39 (1979): 210-11.

[8]

George Schlesinger, *Aspects of Time* (Indianapolis: Hackett, 1980), pp. 79, 144.

[9]

Isaac Asimov, cited in Gardiner, "Games," p. 104.

[10]

Dennis M. Ahern, "Foreknowledge: Nelson Pike and Newcomb's Problem," *Religious Studies* 75 (1979): 489.

[11]

James Cargile, "Newcomb's Paradox," *British Journal for the Philosophy of Scirnce* 26 (1975):

235-6; Doris Olin, "Newcomb's Problem: Further Investigations," *American Philosophical Quarterly* 13 (1976): 130-1; Bar-Hillel and Margalit, "Newcomb's Paradox," p. 297.

[12]

Cf. Alan Gibbard and William L. Harper, "Counterfactuals and Two Kinds of Expected Utility," in *Foundations and Applications* of *Decision Theory,* ed. C. A. Hooker, J. I. Leach, and E. F. McClennen, 2 vols., vol. 1: *Theoretical Foundations,* The Universtiy of Western Ontario Series in Philosophy of Science 13 (Dordrecht, Holland: D. Reidel, 1978), pp. 129-52. They imagine the case of Solomon, who learns that charismatic kings are not prone to revolts of the people, whereas uncharismatic kings are. Moreover, being charismatic or not is purely genetic. Gibbard and Harper contend that it would be irrational for Solomon to refrain from adultery on the grounds that this would be evidence that he is uncharismatic, even though he would welcome the news that he is about to refrain because that would be evidence that he is indeed charismatic. Refraining from adultery would be evidence that he is charismatic, but to refrain for this reason would be irrational, since refraining does nothing to bring it about that he is charismatic. "The 'utility' of an act should be its genuine expected efficacy in bringing about states of affairs the agent wants, not the degree to which news of the act ought to cheer the agent."(*Ibid*., p. 140.) The Newcomb problem, however, has the same structure as the case of Solomon. Hence, they conclude, one ought to take both boxes. According to David Lewis, one boxers are convinced by indicative conditionals: if I take one box, I shall be a millionaire; but if I take both boxes I shall not. Two boxers readily admit the truth of these indicative conditionals, but insist that even if the being is infallible, such that one knows that in taking two boxes he will receive only $1,000, still the rational course is to take both boxes. They take this stand because they are convinced by counterfactual conditionals: if I took only one box, I would be poorer by $1,000 than I shall be after taking both. Since the prediction and placement of the money is not conditioned by my choice, one cannot legitimately employ a backtracking counterfactual instead of the foregoing normal counterfactual. When confronted with the taunt, "If you're so smart, why ain'cha rich?" Lewis retorts that two boxers are not rich because riches are reserved for the irrational. (David Lewis, "'Why Ain'cha Rich?'," *Nous* 15 [1981]: 377-80; so also Locke, "Newcomb Choice," p. 23.) Cf. Doris Olin, "Newcomb's Problem, Dominance, and Expected Utility," in *Theoretical Foundations,* pp. 385-98; Daniel Hunter and Reed Richter, "Counterfactuals and Newcomb's Paradox," *Synthese* 39 (1978): 256-8.

[13]

Nozick,"Newcomb's Problem," p. 127.

[14]

Ibid., p. 132. On the irrelevancy of the prediction prior to the choice, see Robert E. Grandy, "What the Well-Wisher Didn't Know," *Australasian Journal of Philosophy* 55 (1977): 82-90; Andre Gallois, "How not to Make a Newcomb Choice, *Analysis* 39 (1979): 49-53; David Lewis, "Prisoners' Dilemma is a Newcomb Problem, *Philosophy and Public Affairs* 8 (1979): 236-7. According to Lewis, it is "agreed all around" that what really matters is not the prediction's being made in advance, but its being causally independent of one's choice; this is especially evident in the Prisoner's Dilemma, which is a type of Newcomb Problem, for the prisoners' choices are merely independent, not temporally ordered. This insight seems very relevant to theological debates over the temporal necessity of divine foreknowledge, for this necesssity would seem to amount only to the independence of God's foreknowledge and future free choices.

[15]

Nozick, "Newcomb's Problem," p 134.

[16]

Alvin Plantinga, "Ockham's Way Out," Faith and Philosophy 3(1986): 256.

[17]

J. L. Mackie, "Newcomb's Paradox and the Direction of Causation, " *Canadian Journal of Philosophy* 7 (1977): 214, 223; Gregory S. Kavia, "What is Newcomb's Problem About?" *American Philosophical Quarterly* 17 (1980): 278; Bar-Hillel and Margalit, "Newcomb's Paradox," p. 299; Cargile, "Newcomb's Paradox," p. 237; Schlesinger, *Time,* p. 76.

[18]

Nozick, "Newcomb's Problem," p. 146.

[19]

Ahern, "Foreknowledge," p. 484.

[20]

G. Schlesinger, "The Unpredictability of Free Choices," *British Journal for the Philosophy of Science* 25 (1974): 209-21.

[21]

Schlesinger, *Time,* pp. 78-83.

[22]

David Lewis, "Counterfactual Dependence and Time's Arrow, " *Nous* 13 (1979): 456-7.

[23]

Lewis, "Rich," p. 377 Lewis, " Prisoners' Dilemma," pp. 236-7. Like Nozick, Lewis uses the notion of causal influence very broadly. Since in the original paradox the being did not make his predictions based on precognition, Lewis points out that nothing I do now will have any effect on whether I get my million or not. Therefore, a backtracking counterfactual is impermissable. If we suppose that God's foreknowledge is determined by one's choice, however, Lewis's objection would no longer be relevant.

[24]

Lewis, "Prisoners' Dilemma," pp. 236-7. Like Nozick, Lewis uses the notion of casual influence very broadly. Since in the original paradox the being did not make his predictions based on precognition, Lewis points out that nothing I do now will have any effect on whether I get my million or not. Therefore, a backtracking counterfactual is impermissible. If we suppose that God's foreknowledge is determined by one's choice, however, Lewis' objection would no longer be relevant.

[25]

Terence Horgan, "Counterfactuals and Newcomb's Problem," *Journal of Philosopy* 78 (1981): 331-56.

[26]

Levi, "Newcombmania," p. 337.

[27]

See Richmond Campbell, "Introduction," in *Paradoxes,* p. 24. Levi does agree, however, that if the predictor is inerrant, then the one-box strategy is correct.

[28]

Terence Horgan, "Newcomb's Problem: A Stalemate," in *Paradoxes.* p. 224.