

Canadian Journal of Philosophy

Review: [untitled]

Author(s): John Martin Fischer

Source: *Canadian Journal of Philosophy*, Vol. 31, No. 3 (Sep., 2001), pp. 427-444

Published by: [Canadian Journal of Philosophy](#)

Stable URL: <http://www.jstor.org/stable/40232124>

Accessed: 24/02/2011 13:21

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=cjp>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Canadian Journal of Philosophy is collaborating with JSTOR to digitize, preserve and extend access to *Canadian Journal of Philosophy*.

<http://www.jstor.org>

Critical Notice

JORDAN HOWARD SOBEL, *Puzzles for the Will*. Toronto: University of Toronto Press 1998. Pp. xiii + 212.

I Introduction

In this subtle and rigorous book Jordan Howard Sobel answers the question of whether causal determinism is compatible with free will with a resounding, 'Yes and No.' Does this mean that he accepts the doctrine of semicompatibilism: that causal determinism is compatible with moral responsibility, even though causal determinism rules out the sort of free will that involves 'genuine' alternative possibilities? Regrettably, no. In fact, he presents a critique of semicompatibilism. Rather, Sobel's view is that it is crucial to distinguish different versions of the doctrine of causal determinism. Certain versions are compatible with free will, whereas other versions are incompatible with free will. Perhaps the most valuable feature of this book is the careful delineation of the different kinds of causal determinism, and the associated demonstration that the compatibility question hinges on the particular version of causal determinism in question. I shall begin by laying out (briefly) the overall argument of the book (and, I confess, making a few evaluative comments along the way), after which I shall offer a (sketchy) defense of some of my views criticized by Sobel (including semicompatibilism).

II Overall Development of the Book and Some Reflections

1. Chapter One

Sobel's Chapter One is a critical discussion of various versions of 'logical fatalism.' This is the view that 'trivial' or 'tautological' premises can (and do) yield the conclusion that human freedom is illusory — that, despite the prevalent and perhaps natural appearance to the contrary, we do not make genuine choices and cannot choose to do other than we actually choose to do. Logical fatalism issues in the same conclusion as (say) hard determinism or theological incompatibilism (conjoined with the premise

that God exists), but it employs what are alleged to be trivial and uncontentious premises to get to this result. Sobel's complaint about logical fatalism is that it tries (in ways that can be seen to be futile) to get something from nothing. Sobel distinguishes various versions of logical fatalism, and he shows that they all fail because of some sort of ambiguity (or, in his favored term, 'amphiboly'). When regimented carefully, according to Sobel, one can see that the relevant arguments are either invalid or unsound, and that they involve ambiguities (or amphibolies) of scope or mood.

I found Sobel's analysis of logical fatalism to be penetrating, but I would like to register some doubts about what I take to be an unnecessarily extreme conclusion Sobel derives from his ruminations about logical fatalism:

There is, I contend, a contradiction in the view that something that could not conceivably be otherwise even so *matters* and is a proper object of dismay or joy, or a proper source of solace. The logical fatalist's exclamations regarding their conclusions, whether they be of horror ("How awful!") or of appreciation ("How comforting!"), invite the querulous challenge, "But these propositions are not awful or comforting in contrast with their negations, *if*, as you insist, their negations are not merely not so in this poor world of ours, but are not so in any possible world. You say that your fatalistic conclusions are necessary and that their negations are absolutely impossible. But how can these conclusions of yours be horrible or wonderful if their negations are impossible?" (43)

Although I agree with Sobel's analysis of the (fatal) flaws of logical fatalism, I wonder whether it is also true that it is *never* of interest to us (in the way indicated) whether a necessary truth (or, if it is different, something that could not conceivably be otherwise) is in fact true (or the case). Consider, for example, the proposition that God exists. On some (not implausible) views, this proposition is, if true, necessarily true; similarly, if the state of affairs in question obtains, then it could not conceivably have been otherwise. And yet (some) people care deeply about whether God exists, and would, for example, react with great joy to a conclusive demonstration that God exists (or with horror to such a demonstration that God does not exist). Further, it seems that the people who care about the propositions in question genuinely do care about (for example) the proposition that God exists, and not the (admittedly merely) contingent proposition that the sentence, 'God exists,' expresses the necessary proposition.

2. Chapter Two

In Chapter Two Sobel discusses various puzzling contexts in which it appears that an agent's choice can be predicted in advance. More specifically, in these scenarios an agent has good reason for thinking that his

conduct has been truly predicted in advance. 'He is not sure which action of several that seem open to him has been predicted, and he will not be sure of this until he acts. But he is sure, or nearly sure, in these cases that whatever he will do, it has already been predicted that he will do it' (52). In some such contexts (for example, the context of 'The Samarra Problem,' 55-6), Sobel argues that there is no rational 'solution' or choice, insofar as no choice could be ratified upon reflection (59-61). In other situations, however, including the classic 'Newcomb's Problem' (52-4), Sobel argues that there is a rational solution; in the case of Newcomb's Problem (which I shall present more explicitly below), Sobel argues on behalf of the Two Box strategy (62-9).

3. Chapter Three

In many respects, Chapter Three, 'Free Will and Determinism,' is the heart of the book. In it Sobel distinguishes among 'forms,' 'modes,' and 'varieties' of causal determinism, and presents arguments that whereas certain varieties of causal determinism are incompatible with free will, others are compatible with free will. On Sobel's view, all 'forms' of causal determinism are 'elaborations' of the 'causal principle' that 'every event has an antecedent cause' (84). Consider, for example, 'State-to-State Causal Determinism — CD(SS): Every state has at some antecedent time a cause that is a state.' Note that this elaboration of the causal principle is consistent with the world's having a beginning in time in the sense of a 'latest early bound' (88). On this sort of view, one might be able to trace back from any event to previous events which caused it, but since there are an infinite number of times, one need never get back to the latest early bound:

For example, CD(SS) is consistent with the world's having begun five minutes ago so that though nothing existed and the world was in *no* state at any time up to then, things did exist and the world was in some state at every subsequent time. Suppose that the world *did* begin five minutes ago in that sense. Was there *time* for all the causes that CD(SS) requires for all actual states of the world? Yes. Take any state of the world. By hypothesis it takes place at some time subsequent to five minutes ago. So there is a time, indeed there are infinitely many times, between this state's time and five minutes ago, and this no matter how near to that beginning this state's time is. (88)

It is important to note that CD(SS) does not entail the doctrine of 'Ancient Causes: every event has, for every time t , a cause that is an event that takes place at a time antecedent to t ' (93). Sobel offers a form of causal determinism which does in fact entail Ancient Causes, and which is put (as with CD(SS)) in terms of 'states': 'CD(PSS) [Perpetual State-to-State Causal Determinism]: Every state has at every antecedent time a cause that is a state' (98). The distinction between these two forms of determi-

nism will play an important role in the distinction between those varieties of causal determinism that are incompatible with free will and those that are compatible with free will. Additionally, Sobel carefully delineates a range of other forms of causal determinism.

Sobel characterizes the 'modes' and 'varieties' of causal determinism as follows:

These modes consist of several possible views concerning determinism itself, and concerning laws of nature and the past as these figure in forms of determinism. The views to be considered concern how determinisms, and the laws and the past of which they speak, are or are not, *counterfactually conditional* (*variable or unfixed*) on possible states, processes, and choices. I cast these views as "modes" of determinisms that when combined with the "forms" of determinisms produce "varieties." (114)

For example, Sobel characterizes 'Fixed-Laws Determinism' as follows:

Every event in history is determined, *and* no matter what were to happen at a time — no matter what choice, state, or process that is internally-consistent-with-the-laws were then to obtain or begin — *the laws would remain the same.* (116)

Similarly, Sobel characterizes 'Fixed-Past Determinism' as follows:

Every event in history is determined, *and* for any time no matter what state, choice, or process were to take place or begin at this time, *the past to this time would remain the same.* (118)

Thus Sobel builds into the mode — and thus variety — of causal determinism various counterfactual claims that typically are not considered part of determinism *per se*, but rather the relationship between deterministic causation, human choices and behavior, the laws of nature, and the past. I do not, however, think (nor do I interpret Sobel as arguing) that anything substantive depends on whether such claims are considered part of determinism — as specifying a variety of determinism — or are further claims connecting (or expressing counterfactual relations between) deterministic causation to (and) notions such as freedom, choice, the past, and the laws.

Sobel holds that 'choices proper' are 'acts of will' — 'organization or focusing of wills' — of agents (107). He continues:

For an agent to choose to do some action would be for this agent to make up his mind to do that action, or to set his mind to do this action. Choices, as understood here, that is, these acts of wills, would not themselves be states of the world or possible subjects of laws of nature: apart from nature, they would have neither natural causes nor natural effects. But choices as here understood, if there are any, are things that can in *some* sense "be responsible" for and "be reflected" in the states of the world. They would, I assume, typically be acausally responsible for and reflected in certain sets of minds of agents that can lead causally to actions. (107-8)

Sobel, however, contends that his arguments do not require a belief in such 'extra-worldly' choices. He seems to think that he can make do with 'sets of minds' to do things, which are 'certain states of the world for which choices proper, if there are any, are typically responsible, and in which they are typically reflected' (108). Sobel calls these sets of minds 'natural choices.' Free Will, Sobel says, is 'best understood as an agent's capacity to choose to do things and consequent to his choosing to do these things, where the things chosen are different from and incompatible with things he will in fact do' (111). Thus free will, as Sobel understands it, has 'alternative possibilities' built into it: if free will corresponds to a sort of control, it would be — or at least require — (in my term) 'regulative control.'¹

Sobel focuses his attention on the relationship between four varieties of causal determinism and free will. He first argues that a certain mode of State-to-State Perpetual Causal Determinism, a mode in which it is assumed that the laws and the past are 'fixed' in a certain specific way, is incompatible with free will. It is particularly important, in my view, to note (as Sobel does) that his proof makes absolutely no use of any sort of modal transfer principle (or transfer of necessity principle). As Sobel puts it, 'The argument I have given for the incompatibility of [causal determinism of a specific variety] and free will is not so much a *modal* argument as a *conditional* argument' (127). In my view, Sobel's rigorous argument provides a decisive refutation of Van Inwagen's contention that any valid argument for incompatibilism depends (at some level) on a suitable transfer of necessity principle.²

I have the utmost admiration for the rigorous development of the conditional argument for incompatibilism presented by Sobel. It should be pointed out however that Sobel joins the chorus of other distinguished philosophers in exhibiting 'surprise' that I was willing in *The Metaphysics of Free Will* to sketch a conditional argument which I claimed was a good

1 For discussions of two sorts of control — regulative control and guidance control, see: John Martin Fischer, *The Metaphysics of Free Will* (Oxford: Blackwell 1994), especially 131-89; and John Martin Fischer and Mark Ravizza, S.J., *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press 1998).

2 For Van Inwagen's claim, see Peter Van Inwagen, 'When is the Will Free?' in J. Tomberlin, ed., *Philosophical Perspectives* Vol. 14, *Action Theory and Philosophy of Mind*. (Atascadero, CA: Ridgeview Publishing Company 1990) 399-422; and 'When the Will Is Not Free,' *Philosophical Studies* 75 (1994) 95-113. There is a reply in John Martin Fischer and Mark Ravizza, 'Free Will and the Modal Principle,' *Philosophical Studies* 83 (1996) 213-30.

argument, although not formally valid.⁴ Sobel points out that I do 'not claim that the Conditional Version of the Argument for Incompatibilism is "formally valid"' (196).⁵ Sobel goes on to say, 'Fischer is right not to claim that the argument of his Conditional Version is valid' (197). Note that the claim attributed to me has gone from one concerning 'formal validity' to one concerning 'validity.'

But whereas I conceded that my Conditional Argument is not *formally valid*, I nevertheless claimed that it is 'valid' in the sense that anyone who accepts the premises should also accept the conclusion, given the form of the argument *and the content of the premises and conclusion*. Based on the distinction and logical expertise of my opponents here, perhaps I should simply concede. The concession wouldn't matter much, since Sobel thinks that a valid conditional argument can be constructed, and Van Inwagen (and Warfield) think that other valid arguments for incompatibilism can be constructed. But I still would suggest that there are good arguments that are not formally valid, and some of these are 'valid' even though they are not formally valid: they are valid in part in virtue of their *content*. Further, I would still claim that the Conditional Argument for Incompatibilism I presented in *The Metaphysics of Free Will* is valid in this sense.⁶

Having argued that a certain class of varieties of causal determinism is incompatible with free will, Sobel next argues that another class of varieties of causal determinism is entirely compatible with free will (146-64). It is crucial here that the varieties in question are modes (even stringent modes) of mere State-to-State Causal Determinism (as opposed

4 For various levels of irritation and alarm, bordering in some cases on apoplexy, see Van Inwagen, 'When Is the Will Free?' and 'When the Will Is Not Free'; and Ted Warfield, 'Review of John Fischer's *The Metaphysics of Free Will*,' *Faith and Philosophy* 14 (1997) 261-5, and 'Causal Determinism and Human Freedom are Incompatible: A New Argument for Incompatibilism,' forthcoming in Tomberlin, ed., *Philosophical Perspectives*. There are replies in John Martin Fischer and Mark Ravizza, 'When the Will is Free,' in J. Tomberlin, ed., *Philosophical Perspectives* Vol. 6, *Ethics* (Atascadero, CA: Ridgeview 1992) 423-51; and 'Free Will and the Modal Principle'; and John Martin Fischer, 'A New Compatibilism,' *Philosophical Topics* 24 (1996) 49-66, reprinted in Laura Waddell Ekstrom, ed., *Agency and Responsibility: Essays on the Metaphysics of Freedom* (Boulder, CO: Westview Press 2001) 38-56.

5 Fischer, *The Metaphysics of Free Will*, 228n.43

6 For a more detailed development of this argument in response to Van Inwagen ('When Is the Will Free?' and 'When the Will Is Not Free') and Warfield ('Review of John Fischer's *The Metaphysics of Free Will*' and 'Causal Determinism and Human Freedom are Incompatible: A New Argument for Incompatibilism'), see Fischer, 'A New Compatibilism.'

to Perpetual State-to-State Causal Determinism). As such, the varieties of determinism do not entail the doctrine of Ancient Causes, and it seems to me that it is precisely this feature of them that allows for the 'wobble room' required for free will.

CD(SS) is consistent with an action's not having ancient causes. Rather, the action may have been produced by a series of antecedent causes that has no beginning, but which does not ever reach a certain antecedent temporal 'bound.' Sobel calls such a series a 'fast-starting, beginningless' series of states (89). The consistency of CD(SS) with the existence of such fast-starting, beginningless series of states renders even stringent modes of CD(SS) compatible with free will.

Sobel's compatibility argument is ingenious and intriguing. It should be noted, however, that although Sobel describes the form of causal determinism in question (CD(SS)) as 'a venerable and relatively demanding form of determinism' (146), this form (not embodying Ancient Causes) plays almost no role in contemporary and (as far as I know) historical discussions of the compatibility question. The doctrine of causal determinism which worries such incompatibilists as Carl Ginet and Peter Van Inwagen is a variety of causal determinism that does entail the doctrine of Ancient Causes.

Why might it be thought that the deeper threat to free will comes from those varieties of causal determinism that do indeed entail Ancient Causes? Whereas it is perhaps not easy to specify the reason, the basic intuition appears to be that we would not have the relevant sort of freedom or control, if causal sequences that began prior to our birth issue in our choices and actions. Or, alternatively, one might worry that we would not have the relevant sort of control (involving alternative possibilities) if our choices and actions were mere intermediate links in causal chains that one could trace back in time to a point at which it is indisputable that we had no part in them. Indeed, in a different context (the context of criticizing my semicompatibilistic contention that causal determinism does not rule out moral responsibility), Sobel approvingly quotes Derk Pereboom, 'The incompatibilist's most fundamental claim is that moral responsibility requires that one's choice and action not result from a deterministic causal process that traces back to factors beyond one's control' (145).⁷ If *this* is the source of the incompatibilist worry, then exhibiting the compatibility of CD(SS) with free will does not make much progress.

7 Derk Pereboom, 'Determinism *al Dente*,' *Nous* 29 (1995) 21-45, esp. 27

To make the point a bit more explicitly, suppose CD(SS) obtains and a 'natural choice' is associated with a certain fast-starting, beginningless causal sequence. Now, as I understand Sobel's view, a natural choice otherwise would *not* require the past to be different from what it actually was, precisely because of the nature of a fast-starting, beginningless series; after all, it could be different indefinitely far back, without ever getting to the 'temporal bound' (148). This gives the wiggle room for free will, but it also diminishes the dialectical appeal of the compatibility result, because it is typically precisely the worry that determinism would entail that a choice otherwise would require the past (or the laws) to be different that bothers the incompatibilist.

4. Chapter Four

The final chapter, Chapter Four, is a slightly revised version of an extended critical notice of my book, *The Metaphysics of Free Will*.⁸ In this chapter, Sobel provides useful formalizations of various principles of practical reasoning I developed, as well as rigorous and formal versions of the arguments for the incompatibility of causal determinism (in the sense that entails Ancient Causes) and free will (in the sense that involves alternative possibilities) that I sketched in this book. I particularly welcome the elaborate demonstration that certain of these arguments (as with Sobel's related argument discussed above) do *not* depend on any transfer of necessity principle (contrary to Van Inwagen's assertion).

In *The Metaphysics of Free Will* I argued that a central idea that governs our practical reasoning can be employed to give a new perspective on Newcomb's Problem. I argued that accepting this idea leads to a good argument for adopting the Two Box solution in the case of a merely inerrant (never actually wrong) predictor and the One Box solution in the case of an *infallible* (necessarily correct) predictor. Whereas Sobel agrees with the Two-Box approach in the context of inerrancy, he disagrees with the One-Box approach in the context of infallibility. Further, he contends that my own principle should have led me to conclude that *whatever* choice is made in the context of infallibility is *by default* rational. I shall explain Sobel's criticism, and sketch a reply, in the following section.

Sobel also criticizes my argument that God's foreknowledge is incompatible with free will (176-86). Interestingly, God's foreknowledge would seem to imply precisely the doctrine of Ancient Causes (or a

8 Jordan Howard Sobel, 'Critical Notice of John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control*,' *Canadian Journal of Philosophy* 28 (1998) 95-117

relevantly similar doctrine pertaining to non-causally linked prior states). But Sobel contends that God's prior states of mind (that constitute His belief-states) need not have a structure that is like the physical structures that help us to represent the world (and that constitute our belief-states). Rather, the idea is that God has some sort of 'direct apprehension' of the future that is not mediated by representational states. Thus, human free will (in the sense that involves alternative possibilities) need not require the power so to act that the past would have been different from what it actually was. On Sobel's view, then, God's foreknowledge need not pose the same sort of threat as the varieties of causal determinism that entail Ancient Causes.

Additionally, Sobel mounts (in this and the previous chapter) a critique of my doctrine of 'semicompatibilism' — the view that causal determinism need not rule out moral responsibility, even though it does rule out free will (in the sense that involves alternative possibilities). Here Sobel invokes, as he did in the previous chapter, Pereboom and Kant. Recall that Pereboom claims that if an action results from a deterministic causal process that traces back to factors beyond the control of the agent, he is not morally responsible for the action. For the second separate time in the book, Sobel reminds us that in the *Critique of Practical Reason* Kant 'was scornful of the "wretched subterfuge ... [and] petty word-jugglery" of the idea that "although the actions of man are necessarily determined by causes which precede in time, we [may] yet call them free"' (145-6 and 191).

I pointed out in *The Metaphysics of Free Will* that agency has a forward-looking component (deliberation) as well as a backward looking component (moral responsibility). I expressed optimism that, just as the lack of genuinely available alternative possibilities is compatible with moral responsibility, so the lack of such alternative possibilities could be seen to be compatible with deliberation and practical reasoning. For Sobel, however, the glass is half empty:

I view this project with extreme skepticism. Until what for me is the totally unexpected is at least *begun* so that I have a glimmer of how it just might be accomplished, I will be confident that *deliberative agency* entails regulative control [and the associated alternative possibilities], and is thus incompatible with any metaphysics that excludes regulative control. (193)

III Some Critical Reflections: A Defense of Some Central Doctrines of *The Metaphysics of Free Will*

1. Newcomb's Problem

Here is 'Newcomb's Problem,' originally proposed by Robert Nozick:

Suppose a being in whose power to predict your choices you have enormous confidence.... There are two boxes, (B1) and (B2). (B1) contains \$1000. (B2) contains either \$1,000,000 (\$M), or nothing.... You have a choice between two actions: (1) taking what is in both boxes, (2) taking only what is in the second box. Furthermore, and you know this.... (I) If the being predicts you will take what is in both boxes, he does not put the \$M in the second box. (II) If the being predicts you will take only what is in the second box, he does put the \$M in the second box.... First, the being makes its prediction. Then you make your choice. What do you do?⁹ (169)

In *The Metaphysics of Free Will* I pointed out that one's view about the appropriate or rational strategy in Newcomb's problem is naturally thought to depend one's views about the counterfactuals in question. As David Lewis has emphasized, counterfactuals are 'infected with vagueness.' Standardly we resolve this vagueness in such a way that we hold fixed the actual features of the past (or perhaps what we take to be those features), but in certain special contexts we resolve the vagueness in such a way as to allow for backtracking counterfactuals.

If Newcomb's Problem is a context in which the standard resolution of vagueness is appropriate (as Lewis suggests), then I could reason as follows. Either the predictor already has predicted that I would take two boxes or one box. I don't know which prediction the predictor made (and thus what the actual past is in this respect). I can, however, suppose, first, that the predictor predicted that I would take both boxes. Now, employing the standard resolution of vagueness, I can assume that if I were to take both boxes, I would get \$1000, and if I were to take one box, I would get nothing. So, on the supposition that the predictor predicted that I would be a 'two-boxer,' there is a dominance strategy — take the two boxes.

Now I can suppose that the predictor already has predicted that I would take one box. Again employing the standard resolution of vagueness, I can assume that if I were to take both boxes, I would get \$M + \$1000, and if I were to take the one box, I would get \$M. On the supposition that the predictor predicted that I would be a 'one-boxer,'

9 Robert Nozick, 'Newcomb's Problem and Two Principles of Choice,' in R. Campbell and L. Sowden, eds., *Rationality, Cooperation, and Paradox: The Prisoner's Dilemma and Newcomb's Problem* (Vancouver: University of British Columbia Press 1985) 107-33, esp. 114-15. (Abridged from N. Rescher, et al., eds., *Essays in Honor of Carl G. Hempel* (Dordrecht: Reidel Publishing Company 1969) 114-46.)

there is again a dominance strategy, and it is again that I should take the two boxes. Thus it seems that I can conclude that the two-box strategy is correct.

But the above sort of argumentation has been challenged in a fascinating article by Terence Horgan.¹⁰ Horgan maintains, as opposed to Lewis, that Newcomb's Problem is a context in which the standard resolution of the vagueness of counterfactuals is inappropriate. Resolving the vagueness in the way appropriate to the context, it turns out that the 'backtracking counterfactuals' are true, and that the predictor's predictions are not counterfactually independent of my choices. On this approach, if I were to choose the two boxes, then the predictor would have predicted this, and I would only get \$1000; but if I were to choose the one box, then the predictor would have predicted this, and I would get \$M. Thus, I should pick the one box.

This dialectical context seems to be a standoff. How could either Lewis or Horgan argue decisively for his approach to the resolution of the vagueness of the relevant counterfactuals? In *The Metaphysics of Free Will* I proposed a way around this standoff. That is, I proposed an argument for the two-box solution (in the context of an inerrant predictor) that is neutral as regards the counterfactuals in question. Rather, it employs an ingredient I used in what I called the 'Basic Version' of the argument for the incompatibility of causal determinism and free will (in the sense that involves alternative possibilities). This ingredient entails that an agent can perform only those actions that are extensions of the actual past, and thus that the reasons relevant to an individual's deliberations are those which obtain in the worlds that are extensions of the actual past.

Given this extremely plausible constraint on our abilities, and the associated connection with practical reasoning and the relevance of reasons, I presented the following simple argument for the two-box solution. Either the predictor put the \$M in Box (B2) or not. If he did, then the only possible worlds I can now actualize have in their past that this is so, and I get \$1000 more in the one of these in which I take both boxes. If he did not, then the only possible worlds I can now actualize (and thus relevant to my practical reasoning) have in their past that this is so, and I at least get \$1000 in the one of these in which I take both boxes. Thus although I do not know this feature of the actual past, there is a dominance strategy in the sense that on *either* assumption about the past, I should take the two boxes. The dominance strategy here is parallel to the one sketched above, given the standard resolution of the vagueness of

10 'Counterfactuals and Newcomb's Problem,' *Journal of Philosophy* 78 (1981) 331-56

counterfactuals; it is different insofar as it does not rely on counterfactuals at all, but only facts about accessibility to possible worlds.

An important feature of my approach to Newcomb's Problem is that it entails an asymmetry between the mere inerrancy of the predictor and his infallibility. A merely inerrant predictor is always actually correct, but could be incorrect (is incorrect in some possible worlds). An infallible predictor is necessarily always correct — is incorrect in no possible world. On my approach, if the predictor is assumed to be infallible, this changes the situation considerably, and under such an assumption it seems to me rational to select the one box:

The simple point I wish to make is that move from inerrancy to infallibility is *significant*.... When one affirms ... pertinent can-claims in the Newcomb situation with a merely inerrant predictor, one implicitly assumes that either choosing one box or two boxes could be extensions of the actual past. So suppose the predictor predicted that I would choose one box. Nevertheless it is imagined that I can choose two boxes; corresponding to this thought is a possible world in which the predictor is wrong.... But there are *no* such possible worlds, if the predictor is infallible. Thus it seems that, if the predictor is genuinely infallible, the puzzle conditions are not coherent: it cannot be blithely assumed that I can either take one box to two. After all, *one* of the two predictions is part of the actual past, and given the infallibility of the predictor, there can only be *one* relevant extension of the past.... Imagine a revised Newcomb's Problem in which [the predictor is infallible].... Although I can in fact only choose one of these options, I do not know in advance of making my decision which option I can take.... Under such circumstances, it seems rational to choose the one box.¹¹

Sobel criticizes my claim that in the context of an infallible predictor it is rational to choose the one box. He calls my view 'doubly strange' (172). He says, 'First, it is very strange to speak of *choice* — of what it is *rational to choose* — in situations in which, by hypothesis, *one has no choice*' (172). Sobel says:

An agent in such a situation, an agent who was sure he was in such a situation, could not think that he had a decision-problem, where that would be a problem what to do. He might wonder *what he was going to do*, and as a consequence was going to get. He might very well be *curious*. But he could not, consistent with his opinion that he has no choice and that only one action is for him something he can do, wonder *what to do*. Suppose I am sure that of two doors, behind which there are "prizes" ... exactly one is sealed shut. I cannot, with that certainty in mind, deliberate by which door to leave. All I can do is wonder by which door I will leave, and hope that it opens to the best "prize." (172)

11 Fischer, *The Metaphysics of Free Will*, 105-6; Sobel cites this passage on 171-2 of *Puzzles for the Will*.

Sobel's second point is that on my own principles I should have said that in the context of an infallible predictor whatever I choose is *by default* rational, since whatever I do is the only thing I could do. Whatever I choose is the only possible extension of the actual past, and thus is by default rational. Thus, according to Sobel, my own principles preclude my conclusion that the one-box strategy is uniquely rational. (It should be noted that even this conclusion would entail an interesting asymmetry between the context of mere inerrancy and genuine infallibility; it would, however, be a different asymmetry from the one I suggested.)

Although admittedly the circumstances envisaged above are in some ways 'strange,' I do not think that my way of sorting through them is either strange or 'doubly strange.' Or perhaps it would be better to say that my approach may be 'strange' in the sense of a departure from commonsense (insofar as it allows for deliberation even in the absence of the genuine existence of alternative possibilities, or even the belief in such alternative possibilities), but to insist that this sort of strangeness is not a decisive objection to it. After all, I have admitted from the beginning that commonsense presupposes alternative possibilities (for responsibility and deliberation), but I have maintained that we can *reconstruct* our ordinary views in such a way as to preserve what we really care about (without preserving the assumption of alternative possibilities).

I think that one can *make* a choice without 'having a choice.' Having a choice, it seems to me, entails having alternative possibilities; but making a choice does not. Perhaps Sobel is correct that in his case I cannot deliberate by which door to leave; but certainly I could deliberate about which door to choose, and I could consider reasons for choosing each of the doors. I could certainly ask myself which door it would be *rational to choose*; that is, I could reflect on my reasons for choosing each of the doors and come to a judgment about where the best reasons lie. Further, I do not see why one could not engage in such reflection, even if one knows that there is just one *choice* one can make. (Note that it is not nearly so jarring to say this — that I can deliberate about what choice to make even though I know that there is just one choice available to me — as it is to say that I am able to deliberate about by which door to leave in Sobel's example.) I will have more to say about this set of issues below when I address the possibility of deliberation in a world in which there are no genuinely available alternative possibilities.¹²

12 I have discussed these issues in John Martin Fischer, 'The Metaphysics of Free Will: A Reply to My Critics,' *Journal of Social Philosophy* 29 (1998) 157-67, esp. 157-60, where I am replying to Carl Ginet, 'Comments on John Fischer, *The Metaphysics of Free Will*,' *Journal of Social Philosophy* 29 (1998) 126-34. For a particularly helpful defense of the

As to the second criticism, I would suggest the following way of thinking about how my approach structures one's deliberation in the context of an infallible predictor. Recall that in the case of a merely inerrant predictor, there was a *dominance argument* to the conclusion that I ought to take the two boxes; on either assumption about the past, taking the two boxes is best. But there is no such dominance argument in the context of an infallible predictor. On my approach, I should reason as follows. I don't know whether the predictor predicted that I would choose two boxes or one. Suppose, first, that he predicted that I would choose two; then it follows that I can only choose two, and therefore the world in which I choose two is the best of all the worlds accessible to me. Suppose, next, that he predicted that I would choose the one box; then it follows that I can only choose the one box, and therefore the world in which I choose one is the best of all the worlds accessible to me. So there is *no* dominance argument for a particular strategy, in contrast to the situation in the context of mere inerrancy.

Given this, and given that I do not know which choice I in fact will make, I think it is appropriate to ask under what circumstances I would be better: in the world in which I choose the one box, or the world in which I choose the two boxes. Clearly, I am considerably better off in the world in which I choose one box. Given this, the fact that I don't know that choosing one box is something I cannot do, and the fact that there is no dominance argument against choosing the one box, it seems to me that it is rational for me to choose the one box. That is, choosing the one box is recommended by the reasons available to me — it is the rational choice to make, among the options that are not ruled out, given my knowledge. Whether I can actually make this choice or not is not something I know; but nevertheless I can know (as in Sobel's case of the locked doors) what it is *rational for me to choose*. And presumably Newcomb's Problem is precisely about what it is rational to choose.¹³

view that one can usefully deliberate and engage in practical reasoning, even in a context in which one knows that there is just one alternative genuinely available to one, see Hilary Bok, *Freedom and Responsibility* (Princeton: Princeton University Press 1998), esp. 110-14. There are also useful discussions of the possibility of deliberation in the absence of alternative possibilities in John Martin Fischer, Robert Kane, and Ishtiyaque Haji, 'Free Will and Moral Responsibility: Three Recent Views,' *Proceedings of the 1999 Greater Philadelphia Philosophy Consortium at Bryn Mawr College*, Michael McKenna, ed., *Journal of Ethics* (2000) 1-91.

13 In *The Metaphysics of Free Will*, I claimed that the asymmetry between mere inerrancy and infallibility was suggested (although not defended) by Robert Nozick. Sobel points out that Nozick never explicitly addressed the relationship between these two sorts of predictors, and thus did not endorse the asymmetry in question (172-3).

2. God and free will

In *The Metaphysics of Free Will* I suggested that if God's beliefs are like human beliefs in having a representational content, then there is a strong argument that God's foreknowledge rules out human free will. Indeed, I argued that (on this supposition) the arguments for the incompatibility of causal determinism with human free will and the incompatibility of God's foreknowledge and human free will are on a par. I contended that insofar as God's belief-states are representational, they should have some configuration that presents the world as being a certain way. Slightly (but not much) more carefully, I argued that (insofar as the states in question are representational), one and the same mental state (even of God's mind) cannot count as one belief, if one proposition about the future obtains, and another belief (or not a belief at all) if a different proposition about the future obtains.¹⁴ My argument was that if God's beliefs have this sort of configuration, they are relevantly similar to human beliefs, and thus they should be considered fixed features of the past (as are human beliefs).

Sobel raises the question of whether God's beliefs should be construed as representational (in this way), and I agree that it is not evident that God's beliefs should be so interpreted. He raises the possibility that we should understand God's belief states as involving direct apprehension of facts — past, present, and future (181-2). I certainly agree that this is one possible interpretation of God's knowledge. What makes me somewhat uncomfortable about this approach are two considerations. First, it renders it mysterious how God can behave differently in different possible scenarios (based on different information allegedly available to Him). Additionally, it entails widespread backward causation (of a sort different from any posited by relativistic physics). That is, every macroscopic event would initiate a backward-flowing causal chain which would issue in a state of direct apprehension in God's mind at all points in the past. I suppose one must accept a certain amount of mystery in this terrain, but embracing this sort of backward causation is a high price to pay, in my view.

In an important paper written many years ago, Nelson Pike pointed out that in his famous argument for the incompatibility of God's foreknowledge and human freedom, he was assuming that God's beliefs are

I appreciate the correction, in no small measure because it has revealed that I had previously considerably underestimated my originality.

14 Fischer, *The Metaphysics of Free Will*, esp. 120-5

relevantly similar to human beliefs.¹⁵ He contended that his argument presupposed that the *concept* of belief is the same as applied to human beings and God. He further suggested that if essential omniscience changes the concept of belief, then it would not be surprising that one could then defend a kind of theological compatibilism; under such circumstances however Pike contended that one should not say that one had reconciled God's prior *beliefs* with human freedom.

I do not know exactly what is essential to the concept of 'belief.' I am inclined to think that if belief is analytically tied to truth, as in the case of an essentially omniscient agent, then this would *not* in itself entail a change in the concept of belief. Perhaps, however, my argument in *The Metaphysics of Free Will* can point us to something that is 'central' to human belief and which any belief must have in order for it to be *relevantly similar* to human beliefs, for the purposes of the argument for incompatibilism. That is, my argument is that human belief-states have an internal configuration that represents the world as being a certain way, and that would continue to represent the world as being that way, even given certain actual changes in the world. If God's beliefs lack this sort of structure, perhaps they do not share a sufficient similarity to human beliefs to be properly called 'beliefs.' Or perhaps my argument simply gives a more specific structure to Pike's suggestion that his argument for incompatibilism applies only to those individuals who have states that are *relevantly similar* to human beliefs. In any case, I concede that if one interprets God's mental states as radically different in this specific way (as regards representational content), one can avoid the thrust of the incompatibilist's argument. The question then becomes, 'At what price?'

3. Semicompatibilism forever!

As I pointed out above, Sobel on two separate occasions in *Puzzles for the Will* reminds us that we have it on Kant's authority that it is a 'wretched subterfuge ... [and] petty world-jugglery' to suppose that 'although the actions of man are necessarily determined by causes which precede in time, we [may] yet call them free.' Sobel employs this quotation from the *Critique of Practical Reason* to buttress his view that causation in the actual sequence of events leading to human choice and action would rule out moral responsibility, quite apart from considerations pertaining to alternative possibilities. Thus Sobel impugns my doctrine of semicompatibi-

15 Nelson Pike, 'Of God and Freedom: A Rejoinder,' *Philosophical Review* 75 (1966) 369-79

lism, according to which causal determination in itself need not threaten moral responsibility, even if it does in fact rule out alternative possibilities.

Sobel's critique here resonates with complicated dialectical issues that I have discussed at some length elsewhere.¹⁶ Here I primarily wish to point out that in the passage quoted by Sobel, Kant does not explicitly distinguish between two kinds of freedom: one that involves alternative possibilities ('freedom to choose and do otherwise' or, in my terms, 'regulative control') and one that need not ('choosing and acting freely' or, in my terms, 'guidance control'). I am inclined to endorse the claim that it is a kind of 'subterfuge,' or at least rather implausible, to suppose that causal determination (in a sense that entails Ancient Causes) is compatible with freedom of the first kind (that involves alternative possibilities). This is because the arguments for incompatibilism (of the sort discussed above) apply no matter what the specific factors are which causally determine the choice and action; the arguments don't allow for a distinction between 'internal' and 'external' causal determination. But it does *not* follow that it is a subterfuge, a terminological trick, or otherwise a confusion to suppose that there is a genuine and legitimate form of freedom (guidance control) that can be exhibited even in a causally deterministic world and can ground moral responsibility. I would additionally suggest (although I cannot make the argument here) that it does not follow from *anything* that there is *any* sort of subterfuge, trick, or confusion involved in semicompatibilism.¹⁷

4. *Deliberation in a world without alternative possibilities*

As I noted in my reply above to Sobel's critique of my approach to Newcomb's Problem, I believe that there can still be a point to deliberation and practical reasoning, even when the agent knows that there is just one alternative that is genuinely open to him. This is because such an agent might not know which particular alternative is the only one he can actually choose (and take); given this epistemic indeterminacy, the agent should ask himself which action he has most reason to perform. This process of seeking to figure out what he has most reason to choose is completely compatible with believing that he has just one real choice available to him, and that whatever he ends up choosing (and doing)

16 See, especially, John Martin Fischer, 'Frankfurt-Style Compatibilism,' forthcoming in S. Buss and L. Overton, eds., *The Contours of Agency: Essays in Honor of Harry Frankfurt* (Cambridge, MA: The MIT Press 2002).

17 For a more detailed defense, see my 'Frankfurt-Style Compatibilism.'

will be the only option he could have chosen (and taken).¹⁸ The *point* of practical reasoning and deliberation is to seek to figure out what one has the strongest or best reason to do, all-things-considered; this rationale is perfectly compatible with both causal determinism and the agent's knowing that he has just one available path into the future. What is crucial is that the agent not know which path this is.

In *The Metaphysics of Free Will*, I said that we can have moral responsibility, even if the future is not a garden of forking paths. I speculated that we could also have practical reasoning and deliberation. All that is required, in my view, is that the garden of forking paths correspond to *epistemic alternatives*, and not genuinely available *metaphysical alternatives*. In such a context, deliberation and practical reasoning make perfect sense. Neither of the twin aspects of agency — moral responsibility and practical reasoning — requires the existence of a garden of metaphysically open branching paths into the future.

Is my view strange, or, as Sobel might put it, 'doubly strange'? Yes, if 'strangeness' means simply a departure from common sense. My semicompatibilism is indeed a departure from deeply ingrained features of common sense. That's what makes it a challenging and disturbing doctrine. But I would point out that *no* view about causal determinism, free will, and moral responsibility seems to be able adequately to capture *all* aspects of common sense. It is important to note that the semicompatibilist can be *confident* about our status as persons and morally responsible agents. In contrast, the incompatibilist (about causal determinism and moral responsibility) must say that our very status as persons, and as morally responsible agents, depends on the abstruse ruminations of the theoretical physicists. Isn't it 'strange' to suppose that we could be thus held hostage?¹⁹

Received: January, 2001

Revised: March, 2001

JOHN MARTIN FISCHER
University of California, Riverside
Riverside, CA 92521-0201
USA

18 As I pointed out above, there is a detailed and thoughtful defense of this thesis in Bok, *Freedom and Responsibility*.

19 For an additional defense of my approach to Newcomb's Problem, see my 'Newcomb's Problem: A Reply to Carlson,' *Analysis*, forthcoming. I and my co-author offer a further development and defense of semicompatibilism in *Responsibility and Control*. I am grateful to Jordan Howard Sobel for his helpful and gracious comments.