
The Philosophical Review, Vol. XCVII, No. 3 (July 1988)

AN ESSAY ON FREE WILL. By PETER VAN INWAGEN. New York, N.Y., Oxford University Press, 1983. Pp. vii, 248.

van Inwagen divides his essay into two parts. In the first part, he explores the question of whether free will is compatible with causal determinism. In the second part, he addresses such questions as whether we have free will, whether determinism obtains, and whether we are morally responsible for our behavior.

Before van Inwagen turns to the relationship between causal determinism and free will, he discusses fatalism: the doctrine that mere facts of logic and semantics alone, apart from any substantive metaphysical assumption, entail that we lack free will. (For van Inwagen, that we have free will implies that some of us at least sometimes are able to—"have it in our power" to—do other than we actually do.) van Inwagen presents illuminating and convincing refutations of various arguments for fatalism.

As van Inwagen shows, the arguments for the incompatibility of causal determinism and free will are considerably stronger than the arguments for fatalism. For example, if the pertinent arguments employ a "fixity-of-the-past" premise, then the argument for incompatibilism is stronger than the argument for fatalism; the argument for the incompatibility of causal determinism and free will employs a premise that posits the fixity of indisputably "genuine," temporally non-relational (or "hard") facts about the past, whereas the corresponding premise of the argument for fatalism (implausibly) contends that "non-genuine," temporally relational (or "soft") facts about the past are fixed. In the course of his discussions of fatalism and incompatibilism, van Inwagen persuasively establishes that the incompatibilist need not commit any sort of straightforward blunder associated with certain fatalistic arguments, and he need not commit a modal fallacy (sometimes attributed to the incompatibilist).

van Inwagen carefully develops and defends three different versions of a basic argument for incompatibilism, which he calls the "Consequence Argument." This basic argument exploits the fact that, if causal determinism obtains, then every action is the consequence of the past and the natural laws. Roughly, the argument is as follows. If causal determinism obtains and I am free to do other than I actually do, then either I am free so to act that the past would have been other than it actually was or I am free so to act that some natural law which does obtain would *not* obtain.

But since the past and the natural laws are "fixed"—out of my control now—it follows that if causal determinism obtains, then I am not now free to do other than I actually do (and thus I lack free will).

The three arguments for incompatibilism that van Inwagen presents are different ways of developing the Consequence Argument. That is, all three arguments employ some sort of fixity-of-the-past premise and some sort of fixity-of-the-laws premise to generate the incompatibilistic conclusion. The first argument explicitly employs the technical notion of "being able to render a proposition false," but otherwise employs the resources of ordinary English. The second argument employs the apparatus of possible-worlds semantics. The third argument is a "modal argument"; it uses the modal principle which van Inwagen calls principle (β). Principle (β) is a principle of "transfer of powerlessness." It says that if p obtains and you have no choice about p's obtaining, and if p implies q and you have no choice about p's implying q, then q obtains and you have no choice about q's obtaining. (β) is a kind of "closure" principle for powerlessness, structurally parallel to the principle of closure of knowledge under known implication. van Inwagen argues that the three versions of the Consequence Argument provide a convincing case for incompatibilism.

van Inwagen next turns to three arguments for compatibilism: the "Paradigm Case Argument," the "Conditional Analysis Argument," and the "*Mind* Argument." He provides instructive and persuasive reasons to reject the Paradigm Case Argument and the Conditional Analysis Argument. The *Mind* Argument is the argument (associated with Hobart) that free will is incompatible with the absence of causal determinism. Of course, this is not, strictly speaking, an argument for compatibilism, unless it is assumed that free will is possible. van Inwagen's intriguing discussion of the *Mind* Argument focuses on the question of whether free acts can issue from indeterministic sequences. van Inwagen claims that the *Mind* Argument does not conclusively establish that indeterminism is inconsistent with free will, but he concedes that he does not have an account of how free will is consistent with indeterminism. van Inwagen prefers the admittedly "puzzling" conclusion that free will is consistent with indeterminism to the "inconceivable" propositions that free will is impossible or that free will is consistent with causal determinism.

van Inwagen finally considers the "traditional" questions of whether we actually have free will and whether causal determinism obtains. He claims that it is intuitively clear that we (or at least many of us) are morally responsible for at least some of our behavior. Further, he argues (against such philosophers as Harry Frankfurt) that moral responsibility requires free will. Thus, he concludes that we actually have free will. And since he has argued that free will is incompatible with causal determinism, he concludes that causal determinism does not obtain. For van Inwagen, our

reasons for thinking that we are morally responsible are reasons to conclude that the world is indeterministic.

There are two claims of van Inwagen on which I wish to focus here: (1) Free will is incompatible with causal determinism, and (2) Moral responsibility requires free will. Let's begin with the first claim. I believe that van Inwagen's arguments for the first claim are, although powerful, inconclusive. All three versions of the Consequence Argument employ *some* fixity-of-the-past-premise and *some* fixity-of-the-laws premise. But both sorts of premises are ambiguous. Consider first the claim that the past is fixed. One might distinguish two principles:

(FP1) If *e1* occurred at *t1*, then no agent can at any time later than *t1* initiate a causal sequence issuing in *e1*'s not occurring at *t1*.

(FP2) If *e1* occurred at *t1*, then no agent can at any time later than *t1* perform an action such that if he were to perform it, *e1* would not have occurred at *t1*.

Now whereas it is evident that (FP1) is acceptable, it is perhaps unclear that (FP2) is valid. And one sort of compatibilist might both reject (FP2) and remind us that van Inwagen's arguments *need*, not just (FP1), but also (FP2).

Consider next the claim that the natural laws are fixed. One might distinguish two principles:

(FL1) No agent can ever perform an act that itself would be or would cause a law-breaking event.

(FL2) No agent can ever so act that a law-breaking event would (at some point) have occurred.

Now whereas it is evident that (FL1) is acceptable, it is perhaps unclear that (FL2) is valid. And another sort of compatibilist might both reject (FL2) and remind us that van Inwagen's arguments *need* not just (FL1), but also (FL2).

Of course, the three arguments employ different language. Thus, the formulations of the fixity-of-the-past and fixity-of-the-laws premises will be different. But I believe that there will be appropriately formulated analogues of (FP1), (FP2), (FL1), and (FL2) pertinent to each argument. Whereas it may appear that compatibilism is committed to an unpalatable claim about the past or the natural laws, the compatibilist insists that the unpalatable claim is not entailed by his theory, and the claim entailed by his theory is not unpalatable.

Given the disambiguation of the relevant premises, it becomes clear that

van Inwagen's arguments are not decisive. But of course it does not follow that they are not powerful and challenging arguments or that any of the compatibilist's moves is obviously felicitous. It is, after all, rather difficult to produce a *decisive* argument in metaphysics. One way of viewing van Inwagen's arguments is to construe them as offering the following sorts of challenges to the compatibilist. Given that you accept (FP1), why not also accept (FP2)? And given that you accept (FL1), why not also accept (FL2)? Also, van Inwagen can be understood as contending that (FP2) and (FL2) can be seen "directly" to be intuitively plausible. I believe that some progress would be made in this dispute if one could present "direct" arguments for or against such principles as (FP2) and (FL2), or if the compatibilist could produce a principled way of distinguishing between his acceptance of such claims as (FP1) and (FL1) and his rejection of such claims as (FP2) and (FL2).

I now turn to van Inwagen's claim that moral responsibility requires free will. van Inwagen presents a fascinating argument that, despite examples produced by Harry Frankfurt, moral responsibility requires free will. More precisely, the claim is that if any human agent is responsible for *anything*—for some act, omission, event, or state of affairs—then someone could have performed some act he did not in fact perform, or someone could have prevented some event that in fact occurred, or someone could have prevented some state of affairs that in fact obtained. van Inwagen derives this claim—that the existence of moral responsibility requires the free will thesis—from three principles related to but different from the "Principle of Alternate Possibilities":

(PAP) A person is morally responsible for what he has done only if he could have done otherwise.

van Inwagen's three principles are:

(PPA) A person is morally responsible for failing to perform a given act only if he could have performed that act.

(PPP1) A person is morally responsible for a certain event-particular only if he could have prevented it.

(PPP2) A person is morally responsible for a certain state of affairs only if (that state of affairs obtains and) he could have prevented it from obtaining.

van Inwagen claims that: (1) (PPA), (PPP1), and (PPP2) are part of a valid argument which is independent of (PAP) to the conclusion that moral responsibility requires free will, and (2) none of the trio of principles can be shown to be false by the sorts of examples adduced by Frankfurt to im-

pugn (PAP). van Inwagen concludes that moral responsibility requires free will, even if (PAP) is false.

In order to evaluate van Inwagen's position, it will be useful to have before us his version of a "Frankfurt counter-example" to (PAP):

> Suppose there is a man called Gunnar who has decided to shoot his colleague Ridley. Suppose a third man, Cosser, very much desires that Gunnar shoot Ridley. Cosser is naturally delighted with Gunnar's present intention to shoot Ridley, but he realizes that people sometimes change their minds. Accordingly, he devises the following plan: if Gunnar should change his mind about shooting Ridley, Cosser will *cause* Gunnar to shoot Ridley. We may suppose that Cosser is able directly to manipulate Gunnar's nervous system, and is thus able, in the fullest and strongest sense of the word, to *cause* Gunnar to act according to his wishes. Let us suppose, moreover, that there is nothing Gunnar can do about Cosser's intentions or about the power Cosser has over his acts. It would seem therefore, that Gunnar has no choice about whether he shoots Ridley (pp. 162–163).

van Inwagen argues that no Frankfurt-type case can show any of his trio of principles false. The argumentation is intricate and ingenious, and the details will have to be omitted from this review. I wish here to attend to certain aspects of van Inwagen's defense of (PPP1) and (PPP2). Basically, van Inwagen's strategy is to distinguish between the pertinent event-particulars and event-universals. With regard to the event-particular, the agent can be responsible for it, but he can also prevent it. And with regard to the event-universal, the agent cannot prevent it from obtaining, but he cannot be held responsible for it. So according to van Inwagen, once we distinguish between two kinds of events—particulars and universals—we can see that there is not *one* sort of thing of which it is true *both* that it is unavoidable and that the agent is morally responsible for it. For example, Gunnar is morally responsible for the event-particular denoted by "Ridley's death," but this particular was avoidable: it is *not* the case that the same particular would have occurred, had Cosser caused Gunnar to shoot Ridley, since the event which would have been denoted by "Ridley's death" would have had a different causal genesis from that of the event actually denoted by "Ridley's death." (Of course, this point depends on the claim that causal origins are essential features of event-particulars.) And the event-universal, *that Ridley dies (at some time or other),* is unavoidable, but Gunnar is *not* responsible for it; rather, he is responsible for the fact that Ridley dies in a particular way (and, of course, *this* is avoidable).

I believe that van Inwagen's argument here is unconvincing. Consider (PPP1). van Inwagen points out that Frankfurt-type examples do not show (PPP1) to be unacceptable in virtue of presenting cases in which it is both true that an agent is responsible for some event *e* and *e*—the very same event-particular—would also have occurred in the alternate sequence. The strategy which van Inwagen attributes to Frankfurt would certainly be *one* way of showing (PPP1) to be untenable, but I do not believe that it is

the *only* way to do so. Another way of at least casting doubt upon (PPP1) would be to argue that (PPP1) owes its plausibility to some principle (PPP1*) and that Frankfurt-type examples are counterexamples to (PPP1*). If the only plausible reason to accept (PPP1) is that one accepts (PPP1*), and there are counterexamples to (PPP1*), then considerable doubt is cast upon (PPP1). I would suggest—and this is only an undefended suggestion here—that (PPP1) owes its plausibility to some principle such as:

(PPP1*) A person is morally responsible for a certain event-particular only if he could have freely brought about a different event-particular.

Why would one wish to defend (PPP1) unless one accepted some principle such as (PPP1*)? But clearly Frankfurt-type examples provide counterexamples to (PPP1*). So I believe that Frankfurt-type examples do provide reason to be uncomfortable about (PPP1), albeit in a somewhat less direct fashion than envisaged by van Inwagen.

Consider now (PPP2). When Gunnar shoots Ridley, it might seem that Gunnar is morally responsible for the obtaining of the state of affairs *that Ridley dies*, although he could not have prevented it. But van Inwagen claims that, whereas Gunnar could not have prevented the state of affairs *that Ridley dies*, he is not morally responsible for it. And van Inwagen points out that one cannot argue that Gunnar is morally responsible for the obtaining of the state of affairs *that Ridley dies* by noting that Gunnar did something sufficient for the obtaining of this state of affairs. After all, if one could use this sort of argument, one could conclude that Gunnar is morally responsible for the obtaining of the state of affairs *that Ridley is mortal*; but for this Gunnar is manifestly *not* morally responsible.

I agree that it is not the case that whenever one (knowingly and intentionally) does something which is a sufficient condition for the obtaining of a state of affairs, one is morally responsible for that state of affairs. (If this principle were valid, Gunnar would also be morally responsible for the state of affairs *that two plus two equals four*.) But just as with (PPP1), we should not conclude from the inadequacy of *one* argument for the rejection of (PPP2) that there is *no* good argument (based on Frankfurt-type examples) for rejecting it. Perhaps Gunnar is morally responsible for the obtaining of the state of affairs *that Ridley dies* in virtue of freely doing something that at least in part explains the obtaining of the state of affairs.[1] When Gunnar pulls the trigger, he freely does something that in

[1]I am indebted here to: Robert Heinaman, "Incompatibilism Without the Principle of Alternative Possibilities," *Australasian Journal of Philosophy* 64 (1986), pp. 266–276, esp. p. 271.

part explains why the state of affairs *that Ridley dies* obtains. But his pulling the trigger does not in any way explain why Ridley is mortal. It is important to see that one can do something that is such that one's doing it entails that a certain state of affairs obtains without doing something that at least in part explains why the state of affairs obtains. When one breaks a glass, one does something that entails that the state of affairs *that the glass is breakable* obtains; but one's action in no way explains *why* the glass is breakable.

van Inwagen might respond by insisting that the states of affairs *that Ridley dies* and *that Ridley is mortal* are identical. Of course, this claim will help van Inwagen to respond to the suggestion that Gunnar's pulling the trigger at least in part explains why *that Ridley dies* obtains but not why *that Ridley is mortal* obtains and thus that Gunnar is morally responsible for the state of affairs *that Ridley dies* but *not* the state of affairs *that Ridley is mortal* only if explanation is "extensional," which is dubious. But worse still for van Inwagen (who does apparently believe that the two states of affairs are identical: p. 173), it seems to me that the states of affairs *that Ridley is mortal* and *that Ridley dies* are not identical. Note, first, that the claim that they are identical is plausible only on van Inwagen's assumption that the sentence "Ridley dies" (in the specification of the state of affairs *that Ridley dies)* is tenseless. On this interpretation, the state of affairs *that Ridley dies* is the state of affairs *that Ridley dies at some time or other.* But even this tenseless state of affairs is not identical to the state of affairs *that Ridley is mortal.* The following example is due to Robert Heinaman.[2] Suppose that a man is subject to death—and thus mortal—for his first sixty years, which he survives. But imagine that he then takes a miracle drug that makes him invulnerable to death; he would *no longer* be mortal after sixty. Here it is true that the state of affairs *that Ridley is mortal* obtained at some times, but *that Ridley dies* never obtains. So they are not the same state of affairs, and it seems to me that van Inwagen has not provided any reason to deny that Gunnar can be held morally responsible for the state of affairs *that Ridley dies* even though he could not prevent it.

Further, it seems to me natural to suppose that we hold agents morally responsible for either tensed states of affairs (such as *that Ridley dies now*) or temporally specific states of affairs (such as *that Ridley dies at T*), rather than tenseless, temporally unspecific states of affairs such as *that Ridley dies.* Consider the state of affairs *that Ridley dies at T.* Now it appears to be true that Gunnar did something freely—pulled the trigger—that at least in part explains why Ridley dies at T. But the contrast with the state of affairs *that Ridley is mortal* is clear: Gunnar's pulling the trigger in no way explains why Ridley is mortal. I conclude that there might be cases in

[2]Ibid., p. 271.

which Gunnar can be held morally responsible for the obtaining of the state of affairs *that Ridley dies at T* even though he cannot prevent this state of affairs from obtaining. Evidently, there is reason to reject (PPP2).

If one rejects (PPP1) and (PPP2), as I am inclined to do, then one can resist the argument that moral responsibility requires free will. Moral responsibility may require the exercise of *some* sort of freedom—it may require that the agent act freely—but I believe that it need not require free will (in van Inwagen's sense). A rejection of (PPP1) and (PPP2) makes it possible to embrace semi-compatibilism, the doctrine that moral responsibility is compatible with causal determinism, even if causal determinism is incompatible with free will. One advantage of semi-compatibilism is that a semi-compatibilist need not accept the conclusion (with which van Inwagen himself is admittedly somewhat unhappy) that our reasons to think ourselves morally responsible are reasons to think that the world is indeterministic.

This book is not a balanced book. Rather, it is relentlessly (and some might say tendentiously) incompatibilistic. But I believe that it has considerable merit. It is an extremely intelligent, resourceful, and rigorous book. It is filled with subtle, sophisticated, imaginative, and often ingenious argumentation. van Inwagen systematically puts his finger on the operative intuitions of incompatibilism, and he presents incompatibilism as forcefully as has ever been done. Even if one wishes to resist his conclusions and maintain some sort of compatibilism, van Inwagen makes one see, perhaps more clearly than ever before, the *price* of one's compatibilism.[3]

JOHN MARTIN FISCHER

University of California, Riverside

[3]In preparing this review, I have used material that has appeared in: John Martin Fischer, "Responsibility and Control," *Journal of Philosophy* 89 (1982), pp. 24–40; "Van Inwagen on Free Will," *Philosophical Quarterly* 36 (1986), pp. 252–260; and "Introduction: Responsibility and Freedom," in: John Martin Fischer, ed., *Moral Responsibility* (Ithaca, N.Y.: Cornell University Press, 1986), pp. 9–61.

The Philosophical Review, Vol. XCVII, No. 3 (July 1988)

SELF-KNOWLEDGE IN PLATO'S PHAEDRUS. By CHARLES L. GRISWOLD, JR. New Haven, Conn., Yale University Press, 1986. Pp. xii, 315. $29.50.

This book has value both as an interpretation of Plato's *Phaedrus* and as