JOHN MARTIN FISCHER and MARK RAVIZZA

FREE WILL AND THE MODAL PRINCIPLE

INTRODUCTION

In his stimulating paper "When Is the Will Free?", Peter van Inwagen explores the relationship between a very plausible modal principle and free will.[1] In our paper, "When the Will Is Free," we challenged some of van Inwagen's theses about this relationship.[2] Van Inwagen has responded in his paper, "When the Will Is Not Free."[3] We have benefited from van Inwagen's insightful response, but we believe that further exploration of the issues will be fruitful. Here we undertake to discuss some of the main points raised by van Inwagen.

I. THE RELATIONSHIP BETWEEN THE ARGUMENTS FOR INCOMPATIBILISM

It is uncontroversial that there are various different apparently plausible formulations of the argument for "incompatibilism" – the view that causal determinism is incompatible with freedom to do otherwise. Some of these arguments explicitly employ a modal principle. This modal principle can be stated in various different ways (and has been given various different names). The basic idea of the modal principle is that if some state of affairs $S1$ obtains and one does not have any choice about (or control over) $S1$'s obtaining, and if $S1$ implies $S2$ and one does not have any choice about (or control over) the fact that if $S1$ obtains, then $S2$ obtains, then it follows that $S2$ obtains and one does not have any choice about (or control over) $S2$'s obtaining. The modal principle works as a kind of modal slingshot: it projects the modal property of "powerlessness" from one state of affairs ($S1$) to another ($S2$).

In his important and influential book, *An Essay on Free Will*, van Inwagen presents three different versions of an argument for

incompatibilism.[4] Only one of these versions – the "Modal Version" – explicitly employs the modal principle, which van Inwagen calls, "Principle Beta." (We shall here follow van Inwagen in calling the principle "Beta".)

Very briefly, the Modal Version of the argument for incompatibilism proceeds as follows. Suppose causal determinism obtains. It follows that a statement describing the genuine features of the universe in the past conjoined with a statement of the natural laws entails that you behave as you do now. You now have no choice about (or control over) the fact that the universe had those features in the past. And given that the laws of nature entail that if the universe had those features in the past you will behave as you are actually behaving now, it follows that you have no choice about (or control over) the fact that if the universe was that way in the past, then you are behaving as you are actually behaving now. Now an application of the Modal Principle yields the result that you have no choice about (or control over) your current behavior.

Although only one version of the argument for incompatibilism presented by van Inwagen in *An Essay on Free Will* makes explicit appeal to Beta, he suggests that all three versions of the argument for incompatibilism will "stand or fall together."[5] In "When the Will Is Free," however, we present a version of the argument for incompatibilism – the "Conditional Version" – which we suggest does not explicitly or implicidy rely on principle Beta. It would follow that, if one were to reject Beta and thus the Modal Version – one could not thereby find fault with the Conditional Version.

The Conditional Version of the argument for incompatibilism proceeds, very roughly, as follows.[6] Suppose that causal determinism is true. It follows that a statement describing the universe in the past, together with a statement of the laws of nature, entails that you behave as you are now. Now one of the following conditionals must be true. (1) If you were to do otherwise now, then the universe would have been different in the past than it actually was; (2) If you were to do otherwise now, then the natural laws would be different from what they actually are; or (3) If you were to do otherwise now, then either the universe would have been different in the past than it actually was or the natural laws would be different from what they actually are. But given the fixity of the past, it is plausible to say that (I) if (1)

were true, then you cannot do otherwise now. And given the fixity of the natural laws, it is plausible to say that (II) if (2) were true, then you cannot do otherwise now. And given the above, it is plausible to say that (III) if (3) were true, then you cannot do otherwise now. So (IV) you cannot do otherwise now.

Our claim in "When the Will Is Free" is that the Conditional Version is an attractive version of the argument for Incompatibilism which does not explicitly or implicitly rely on Beta. In "When the Will Is Not Free," van Inwagen denies our claim that the Conditional Version does not rely upon Beta, and he explains and defends his thesis that all plausible versions of the argument for incompatibilism must stand or fall together. He does this (in part) by considering one of the versions of the argument for incompatibilism which he presented in *An Essay on Free Will* and which does not explicitly appeal to Beta. Let us call this the "Access-to-Possible-Worlds" Version. Van Inwagen then goes on to argue that the Access-to-Possible-Worlds Version implicitly makes such an appeal, and suggests that any version of the argument (including the Conditional Version) must at least implicitly make this sort of appeal.

Van Inwagen points out that the Access-to-Possible-Worlds Version of the argument

appeals to no rules of inference but those of textbook logic, and its two premises

> No one has access to a possible world in which the past is different from the actual past
> No one has access to a possible world in which the laws are different from the actual laws

certainly do not seem, on the surface, to commit their adherents to the validity of Beta. It is therefore a good 'test case' with which to confront my general thesis [that all versions of the argument rely upon Beta in some way or another].[7]

Van Inwagen goes on to ask why someone should accept the two premises. He begins by showing that if one accepts Beta (together with certain trivial assumptions), one can *derive* the principles.[8] Further, he simply states that he does not see any other reason for accepting these premises.[9] Van Inwagen condudes:

As I have said, it seems plausible to *me* to suppose that the point of this example can be generalized. I do not know how to prove this, but I would suppose that what is *in effect* an allegiance to Rule Beta must lurk somewhere, in however inarticulate a form, in the background of any technically satisfactory argument for incompatibilism.[10]

We do not wish to take issue with van Inwagen's derivation of the premises of his apparently nonmodal version of the argument from Beta (together with trivial assumptions). But we do think it is worthwhile to explore the dialectical issues raised by van Inwagen's strategy. The situation could be described as follows. Van Inwagen discusses two versions of an argument for incompatibilism. Version One is explicitly modal in the sense that it explicitly employs Beta. Version Two – the Access-to-Possible-Worlds-Version – is apparently nonmodal in the sense that it does not explicitly employ Beta. Rather, it employs van Inwagen's two premises together with trivial assumptions. Now the question is whether Version Two *implicitly* relies on Beta.

Van Inwagen answers "yes" because one *can* derive the premises of Version Two from Beta, and he does not see any *other* way of deriving the premises (or any other reason why one should accept the premises). But we respond as follows. From the fact that one *can* derive the premises of Version Two from Beta, it does *not* follow that one *must* or *ought* to do so. And there may well be good reasons not to. To explain. The premises of the second version of the argument seem to us to capture plausible intuitive views about the past, the laws, and our freedom. Consider, for example, the premise

> No one has access to a possible world in which the past is different from the actual past.

Surely this premise corresponds to the intuitive picture of the future as branching off a "given" past. The intuitive picture is that we are free to add to and extend a given past: there are various possible paths that all branch off a *single* past, and although we may be able to take various different paths into the future, we cannot now make it the case that some past *other than the actual past* have been the past. The premise gives expression to this powerful intuitive view.

Now we grant that we have not given any sort of technically (or logically) adequate derivation of the premise from the intuitive considerations on which we claim it rests. But we do not see why this is necessary. Surely it is enough to show that the premise is strongly supported by a set of very deep and plausible intuitive considerations; why does one need to provide a formal derivation of the premise?

Consider also the premise

> No one has access to a possible world in which the laws
> are different from the actual laws.

Again, this premise seems to us to be strongly supported by intuitive views about the relationship between human powers and the laws of nature. If one did not accept this premise, how could one adequately explain why we don't think individuals can travel faster than the speed of light (or build machines that would cause objects to do so, and so forth)? In explaining why human agents cannot do certain things, it is natural to appeal to the fact that we cannot violate (or cause to be violated) a natural law. If this is the only – or the most plausible – sort of explanation for the inabilities in question, then there would be support for the premise that does not involve Beta.

As above, we admit that we have not provided a "logically adequate" derivation of the premise. Rather, we have shown that its rejection would leave some of our common practices (of attributions of abilities and inabilities) without any support. Given that these are relatively central practices, and it is unclear what *other* sort of support they could have, why isn't this enough? Why does one *need* a *technical* derivation of the premises?

To explore further the force of this question, recall two important facts. First: Beta has been challenged by various thoughtful philosophers.[11] And second: van Inwagen admits that Beta itself cannot be derived – or at least that he cannot see how to derive it. Given that Beta is contentious and itself cannot be derived, and given that the premises of Version Two can be given strong intuitive support, it seems to us prudent for an incompatibilist to have Version Two as an independent option.

To see this more clearly, suppose that someone challenges Beta in the manner suggested by Slote and Dennett, or in some other way. Asked to defend Beta, van Inwagen must concede that he cannot derive it from more basic, uncontroversial ingredients. Presumably he must say it just seems right and it seems to explain what we are inclined intuitively to say in various contexts. Fine; but how is van Inwagen better off than the proponent of Version Two?

Considered a bit more abstractly, there are two arguments. Let us say that one argument employs a premise from which one can derive

the premises of the second argument but it is not the case that one can derive the premises of the first argument from those of the second. What exactly does this show about the arguments? Surely, it does *not* show that the second argument is in any way superfluous. Indeed, there are examples of the above situation in which one would clearly prefer to employ the second argument. As a rather trivial example, suppose the first argument employs a set of premises, one of which is "A&B". And imagine that the second argument employs a set of premises, one of which is "A". Well, one can derive in a logically satisfactory way the relevant premise of the second argument from the relevant premise of the first. But what does this show? In this case it may well be preferable to employ the second argument since it begins with a weaker (and thus perhaps less contentious) premise!

We thus conclude that van Inwagen has not established that the Access-to-Possible-Worlds Version of the argument for incompatibilism implicitly relies on Beta. And since he believes that the considerations pertinent to this version will also generalize to the Conditional Version, he has not established that the Conditional Version implicitly relies on Beta.

II. THE CONDITIONAL VERSION OF THE ARGUMENT

In the above section we considered van Inwagen's general worries about versions of the argument for incompatibilism in which one cannot "derive" the premises. Van Inwagen also criticizes the Conditional Version for another reason. His criticism here is not that the premises cannot be derived, but rather that the conclusion cannot be derived from the premises.[12]

The Conditional Version, laid out informally above, rests importantly on intuitive views about the fixity of the past and the fixity of the natural laws. These views are here given expression by certain conditional statements, which van Inwagen identifies as "premises" of the argument:

> For any action Y, agent S, and time T, if it is true that if S were to do Y at T, some fact about the past relative to T would not have been a fact, then S cannot do Y at T.

> For any action Y, and agent S, if it is true that if S were to do Y, then some natural law which actually obtains would not obtain, then S cannot do Y.

Van Inwagen correctly states:

But incompatibilism cannot be deduced from these two premises, since neither of the following two propositions can be deduced from determinism: if I had at any time acted differently from the way I in fact acted at that time, something prior to that time would have been different from the way it actually was; if I had at any time acted differently from the way I in fact acted at that time, the laws of nature would be different from what they actually are. ... If the world is deterministic, it does indeed follow that if I had acted otherwise than I in fact have, then either the past would have been different or the laws would be different. But it does not follow from this that if I had acted otherwise than I in fact have, the past would have been different, and neither does it follow that if I had acted otherwise than I in fact have, the laws would be different.[13]

These considerations provide the basis for van Inwagen's claim that our argument is "logically defective."[14] Van Inwagen elaborates:

Interestingly enough, Fischer and Ravizza are aware of this barrier to deducing incompatibilism from their two premises, but they attempt to do so anyway. Since, as we have seen, this cannot be done, there must be some flaw in their argument. It is this. They employ the following argument-form (in the reasoning at the top of p. 428):

$$(p \Box \!\!\rightarrow q) \rightarrow s$$

$$(p \Box \!\!\rightarrow r) \rightarrow s$$

hence,

$$(p \Box \!\!\rightarrow .q \vee r) \rightarrow s.^{15}$$

Van Inwagen is here formalizing the move in the Conditional Version, presented informally above, from the truth premises (I) and (II) to the conclusion that (III) is true. (This is of course only a *part* of the argument to the conclusion that (IV) you cannot do otherwise now.) To see this, let "p" be "you were to do otherwise now," "q" be "the universe would have been different in the past from the way it actually was," "r" be "the natural laws would be different from what they actually are," and "s" be "you cannot do otherwise now." Van Inwagen points out that this argument-form is invalid for reasons similar to those that explain the invalidity of the argument-form discussed in the penultimate quotation.[16] He concludes that our argument is invalid and that he has not yet seen "a counterexample to

the thesis that any logically adequate argument for incompatibilism must make a covert appeal to the validity of Rule Beta."[17]

To respond. First, note that van Inwagen employs two different arrows in his regimentation of the argument. This suggests that he is interpreting the main connectives in the argument differently from the connectives in the antecedents; perhaps he is interpreting the main connectives in terms of material implication. But we never intended the conditionals in our argument to be material conditionals, and there is no reason why the argument should be interpreted in this way. Rather, the argument is to be understood as employing subjunctive conditionals throughout:

$$(p \square\rightarrow q)\square\rightarrow s$$

$$(p \square\rightarrow r)\square\rightarrow s$$

hence,

$$(p \square\rightarrow .q \vee r)\square\rightarrow s.$$

Further, van Inwagen is correct to say that we never presented this version of the argument for incompatibilism as "logically valid." That is, we did not present it as an argument whose conclusion must be accepted simply in virtue of its *form*. Rather, our point was (and continues to be) that it is very reasonable to accept the conclusion, given the form of the argument *and the content of the premises*. And nothing van Inwagen says in any way vitiates our point.

To further explain our position, it will be useful to lay out van Inwagen's concrete example which shows the "invalidity" of the argument-form in question:

Suppose you have a little indeterministic device that sports a button, a red light, and a green light. If you press the button, one light or the other will flash, but it is undetermined which will flash. It would seem to follow that if you had pressed the button a moment ago, either the red or the green light would have flashed, but it is not true (and hence, if every proposition is either true or false, is false) that if you had pressed the button the red light would have flashed and it is false that if you had pressed the button the green light would have flashed. .... Let *p* be 'You pressed me button', *q* be 'The green light flashed,' *r* be 'The red light flashed' and *s* be 'Pressing the button would have a determinate outcome.'[18]

Van Inwagen's example usefully shows how the premises of the argument-form (of the *sub-argument* of our argument) can be true

compatibly with the falsity of the conclusion. More precisely, the point is that there are arguments whose premises have the same *form* as (I) and (II) but which do not entail the corresponding conclusion with the form of (III).

But again, our point is *not* that one must accept the conclusion of the sub-argument of the version of the argument for incompabilisim we proferred simply in virtue of the form of its premises. Rather, we assert that it is very reasonable to accept the conclusion in virtue of the form of the sub-argument *and the content of its premises.* A variant on van Inwagen's example can help to motivate our position. Suppose then that everything is as in van Inwagen's example, except now take *s* to be "Pressing the button would result in the flashing of a light whose color I like."[19] Now it seems to us eminently reasonable to accept the conclusion of the argument based on its form *and the content of its premises.* And we maintain precisely this position about our version of the argument for incompatibilism.

In van Inwagen's example, his analogue to – or, perhaps, instantiation of – (III) is not a necessary truth. But arguably our (III) is a necessary truth. Thus, someone might argue that it is unclear that we have established that (III) *follows* in any interesting sense from (I) and (II).[20] We claimed that anyone who accepts (I) and (II) should also accept (III), in virtue of the form of the argument and the content of the premises. But insofar as (III) is a necessary truth, it might be questionable whether one's tendency to accept it genuinely comes from one's acceptance of the premises.

We would respond to this worry in two steps. First, our version of van Inwagen's example of the lights is a case in which the analogue to (III) seems to *follow* from the analogues to (I) and (II), but not formally. And we contend that it is plausible that our (III) similarly *follows* from our (I) and (II), but not formally. That is, it seems that what is going on in our transition from (I) and (II) to (III) is relevantly similar to the transition from the analogues of (I) and (II) to the analogue of (III) is our version of van Inwagen's example.

But the second step is to note that we need not insist on this point. Suppose that (III) is a necessary truth that does *not* follow in any interesting way from (I) and (II). Then it will still clearly be the case that we have a perfectly good argument from (I), (II), and (III) to (IV) – the conclusion that you cannot do otherwise (given the truth

of causal determinism).[21] And if this is so, then we have a perfectly good argument to the incompatibilist's conclusion which does *not* employ the modal principle Beta.

One might wonder why anyone would wish to take seriously an argument that is not formally valid or technically satisfactory, when other arguments for the same conclusion which *are* formally valid are available. The answer is similar to the position developed in the previous section. The version of the argument for incompatibilism which employs Beta may be formally valid, but Beta is questioned by various thoughtful philosophers. And Beta (evidently) cannot itself be "derived". Now the version of the argument for incompatibilism we presented is not formally valid, but its conclusion ought to be accepted, given its form and the contents of its premises; and its premises have considerable intuitive support. Thus, it is not obvious that one is any better off with the formally valid argument. And of course our view is not that the formally valid argument is somehow to be discredited. Rather, it is that it is useful to see that the incompatibilist can present the basic thrust of his view in a different way – a way that avoids my potential controversy about Beta.

III. MOTIVES, PROXIMITY, AND ACCESSIBILITY

Thus far we have been investigating whether all adequate versions of the argument for incompatibilism depend, at least implicitly, upon principle Beta. We have argued that they do not. Now we want to consider two further claims van Inwagen has defended. First he has maintained that if a "Beta-like" principle is valid it follows – for reasons quite independent of determinism – that agents rarely, if ever, are free to do otherwise. Second, he has argued that it does not follow from this result (and his assumption that freedom to do otherwise is a necessary condition of moral responsibility) that agents can only rarely be held morally accountable for the consequences of their actions. We disagree with both of these claims.[22] In the present section we address the first claim; in the following section we take up the second.

Let us briefly consider van Inwagen's reasons for thinking that if Beta is valid, then we rarely, if ever, are free to do otherwise. First he notes that anyone who accepts Beta also ought to accept

a closely related inference principle, "Beta-prime." Beta-prime is basically a version of Beta that is indexed to an agent and a time. More precisely, Beta-prime tells us that "from Nx,p and Nx,(p → q) deduce Nx,q" (where the two-place operator 'N' is used as follows: 'Nx,p' abbreviates 'p and x now has no choice about whether p').[23]

Next, van Inwagen presents several cases in which the validity of Beta-prime allegedly implies that the agents have no choice other than to act as they do. These cases include situations in which (1) agents refrain from doing something they consider morally indefensible, or (2) they do something that they have an unopposed inclination to do, or (3) they act without reflection or deliberation. The form of argument van Inwagen gives to show that agents have no choice in these sorts of situations is basically the same for all three cases, so it will suffice here to entertain only his argument in support of the first.[24]

Consider, then, a case in which someone proposes that an agent perform an action that he considers morally reprehensible and which he is not (at the moment) even tempted to perform. Van Inwagen imagines a case in which he is asked to bear false witness about a colleague, Smith, in order to block Smith's appointment to Chair of the Tenure Committee, an appointment van Inwagen does not object to in the least. About such a case, he writes:

I am *unable* to do what my colleague has proposed: that is, I am not going to do it, and the fact that I am not going to do it [is] something that I simply have no choice about. The argument for this conclusion – it is an instance of the rule Beta-prime – is this ('A' stands for the proposed act):

> N, I I regard A as indefensible
> N, I (I regard A as indefensible → I am not going to do A)

hence

> N, I I am not going to do A.

In this argument, 'I regard A as indefensible' is short for 'I regard A as an indefensible act, given the totality of relevant information available to me, and I have no way of getting further information, and I lack any positive desire to do A, and I see no objection to *not* doing A, given the totality of relevant information available to me.'[25]

In "When the Will is Free," we contended that this argument fails because the second premise is flawed. The idea behind our criticism can be summarized briefly as follows. According to van Inwagen's

terminology, if an agent regards an action A as indefensible, then he has no positive desire to do A. For the sake of argument we are willing to grant that one cannot perform an action, unless at some point one has a positive desire to do it. So it might seem to follow that if an agent regards A as indefensible (and hence has no positive desire to do A), then he does not have the power to do A. But to leap to such a conclusion would be hasty. Why? Because even if an agent does not at the moment have a positive desire to do A, "he might well have the ability (during the relevant temporal interval) to generate such a desire, and to act on that desire. And it is extremely implausible to suppose that agents quite generally lack the *power* to generate the relevant sorts of desires."[26]

To support this claim, we considered cases in which agents summon a desire to do something indefensible simply to exercise their freedom or to flaunt moral prohibitions. van Inwagen acknowledges the possibility of such examples; in fact, he offers his own version of this type of case. In it, his reading of Sartre persuades him that he is in "bad faith" when he denies that he is radically free to perform acts that are morally indefensible. As a result of this realization he acquires the desire to perform an act that transgresses his deepest moral boundaries, an *acte gratuit*. When his colleague proposes that he lie about Smith to block Smith's appointment, van Inwagen initially is repulsed by the suggestion (which he considers indefensible); yet a moment later, he realizes that this would be a perfect way to perform an *acte gratuit*. And, if he were to allow himself to continue in these reflections, he would summon the desire to lie about Smith, and do so.

Cases such as these, we contend, show that van Inwagen's original claim is mistaken. It does *not* follow from the truth of Beta that we are not free to perform indefensible actions (even though we initially may lack any positive desire to desire to perform them).[27] This does not follow because, given our power to summon certain sorts of desires, the proposition

> N I, (I regard A as indefensible $\Box\!\!\rightarrow$ I am not going to do A)[28]

is not obviously true.

Van Inwagen concedes this last point. Nevertheless, he discounts its force. At most, claims van Inwagen, such examples show that in *some* cases one is free to act indefensibly, but these cases are so "remote" and occur so infrequently that they have little bearing on his main thesis. He summarizes his objection as follows:

It is only in cases in which such potential motives for performing A exist and I can reach them from the starting point "I regard A as reprehensible and I have no desire to perform A" that I have the power or ability to proceed from that starting point to a performance of A. As I have said, I am convinced, on the basis of an examination of my own biography and my modal and counterfactual judgments about the existence of "nearby" potential motives that the cases in which such potential motives so much as exist are very rare.[29]

Certainly van Inwagen is correct to argue that in order for an agent to be free to do something indefensible there must be some "open path into the future" that leads to his performing the indefensible act. We attempted to describe such a path by showing how agents not only have a potential motive to act indefensibly, but also the power to summon this motive (during the relevant temporal interval). Although van Inwagen appears to accept the plausibility of this sort of example in some instances, he wishes to avoid its force by suggesting that "in the vast majority of cases, however, there will be no potential motive ... that is lurking somewhere nearby in logical space."[30] The thought here seems to be that unless there is a potential motive "nearby" the agent does not have the power to acquire the desire and perform the indefensible action – there is no path to this future open to him. But why should an assessment of the agent's power be based on which motives are "lurking nearby in logical space?" In assessing the alternatives open to an agent we do not consider merely those accessible possible worlds that are *most similar* to the actual world. Rather, for an agent to have the power to act indefensibly it suffices that there merely be *some* accessible possible world in which he does so act; this world need not, however, be one "lurking nearby in logical space."

To illustrate this point consider any bizarre action that you have no desire to perform and which you would perform only in a rather distant possible world. For example, imagine that even though you are a singularly unmusical person with an aversion to Scottish music, someone you very much dislike signs you up for bagpipe lessons (as a kind of practical joke). Certainly there would not be a potential

motive to practice the pipes lurking *nearby* in logical space. Never-theless, it seems intuitively plausible to suppose that you do have the power to practice the bagpipes; it would seem odd to deny that there is a path to this future open to you simply because this potential state of affairs does not lie *close* (in some sense) to the actual world.

Similarly, even if one were to concede that only rarely do desires to exercise radical freedom or to flaunt moral laws exist in "nearby" possible worlds, what difference should this make to the claim that an agent has the power to summon such desires and act indefensibly? As long as there is some path open to a possible scenario in which the agent does (during the relevant temporal interval) acquire such a desire, it should not matter how "remote" the motive may be. The force of our examples, then, was never intended to rest on a claim that such motives for acting indefensibly were lurking just around the corner. Rather, we were arguing that agents frequently have the power to summon these rather unusual concerns, and insofar as they have this power, they have the freedom to act indefensibly.

IV. MORAL RESPONSIBILITY FOR THE GOOD

Van Inwagen accepts the traditional view that moral responsibility requires freedom to do otherwise. He also accepts the truth of Beta-prime, and he insists that anyone who accepts this principle ought also to conclude that we rarely, if ever, are free to do otherwise. Given these commitments, one might expect van Inwagen also to concede that we rarely, if ever, are morally responsible for our actions. How-ever, he resists this natural consequence of his views. Instead he argues that even though we presently are unable to do otherwise in the majority of cases, this inability results from habits and character traits that were formed to a certain extent by our past free choices; consequently, we still can be held morally accountable for our present inability to do otherwise and the actions that flow from this feature of our character.

In our earlier paper, we suggested that this strategy runs afoul of ordinary intuitions concerning moral responsibility. We are normally held responsible for the actions that flow from our character traits, even though "much of our character results from the habituation we

receive in early life, and these portions of our character don't seem to be necessarily connected with situations of conflict between duty, inclinations, or incommensurable values [the three sort of cases in which van Inwagen grants that agents are free to do otherwise]."[31] In contrast van Inwagen's approach would maintain that we should *not* be judged responsible for such acts.

To illustrate the tension in van Inwagen's view, we offered the example of a patriotic woman named Betty. Betty's unquestioning loyalty to the United States is the result of her early upbringing, and it has never been tested in any of the three types of conflict situations mentioned above. While traveling abroad, Betty is mistakenly approached by a foreign spy who asks her to betray the U.S. for a large sum of money. She instantly refuses without being tempted in the least. Now, if van Inwagen's arguments concerning Beta were sound, then Betty did not in this situation have any choice about whether to refuse the offer, and hence she should not be judged morally accountable for her action. But this seems to be a *reductio* of his view. Clearly Betty knows what she is doing in turning down the offer, and her action flows freely from her character. Why shouldn't she be considered morally responsible and worthy of a certain degree of praise for her display of loyalty?

In answer to this question, van Inwagen notes that usually assessments of moral responsibility arise in connection with bad actions, not good – "it would be odd indeed to say, 'Find out who the people are who are morally accountable [or even morally responsible] for the excellent safety record in District Three.... ' "[32] Yet, as van Inwagen himself points out, "this oddness may only be a matter of 'conversational implicature.' "[33] We think that it is.

There are many contexts in which it would sound odd to make a certain claim, even though one would be willing to accept that it is true. For example, if Joe is driving the car to the store, we usually would not say, "Joe is *trying* to drive the car to the store." Nevertheless such a statement is true. Similarly, it might sound odd to say, "Betty is morally responsible for declining to betray her country," even though this is true. The simple point is that our sense of moral responsibility is not exhausted by those contexts in which one typically would use locutions that include the phrase "morally responsible." As Strawson has reminded us, to hold someone respon-

sible involves more than a propensity to employ certain phrases; it involves a willingness to adopt certain attitudes toward the person and to react to him in certain kinds of ways – e.g., to treat him with respect, to hold him in contempt, to thank him, to praise or blame him, and so on.[34] Thus, an adequate theory of moral responsibility should recognize that people can be held responsible for both good and bad actions. Insofar as van Inwagen's view requires us to deny that agents can be morally accountable for good acts such as Betty's, this provides a reason to question its ultimate plausibility.[35]

NOTES

[1] Peter van Inwagen, "When Is the Will Free?" in J. Tomberlin, ed., *Philosophical Perspectives IV: Action Theory and Philosophy of Mind*. (Atascadero, Ca.: Ridgeview Publishing Company, 1990), pp. 399–422.

[2] John Martin Fischer and Mark Ravizza, "When the Will is Free." in J. Tomberlin, ed., *Philosophical Perspectves VI: Ethics*. (Atascadero, Ca.: Ridgeview Publishing Co., 1992), pp. 423–451. See also John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control*. (Oxford: Blackwell, 1994), pp. 47–62.

[3] Peter van Inwagen, "When the Will Is Not Free," *Philosophical Studies* 75 (1994), pp. 95–113.

[4] Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983).

[5] Van Inwagen, *An Essay on Free Will*, p. 57.

[6] For a more careful formulation, see "When the Will Is Free," pp. 427–428; and *The Metaphysics of Free Will: An Essay on Control*, pp. 62–63.

[7] Van Inwagen, "When the Will Is Not Free," p. 96.

[8] Van Inwagen, "When the Will Is Not Free," pp. 96–97.

[9] Van Inwagen, "When the Will Is Not Free," p. 97.

[10] Van Inwagen, "When the Will Is Not Free," p. 98.

[11] Michael Slote, "Selective Necessity and the Free-Will Problem," *Journal of Philosophy* 79 (1982), pp. 5–24; and "Review of Peter van Inwagen's *An Essay on Free Will*," *Journal of Philosophy* 82 (1985), pp. 327–330. See also Daniel Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge, Ma.: MIT Press, 1984), pp. 148–149.

[12] Van Inwagen, "When the Will Is Not Free," p. 99.

[13] Van Inwagen, "When the Will Is Not Free," pp. 98–99. van Inwagen invokes David Lewis's counterfactual logic to support his analysis.

[14] Van Inwagen, "When the Will Is Not Free," p. 98.

[15] Van Inwagen, "When the Will Is Not Free," p. 99.

[16] In regard to the argument-form of the penultimate quotation, van Inwagen points out that it is shown to be invalid, on David Lewis's counterfactual logic: "When the Will is Not Free," p. 99.

[17] Van Inwagen, "When the Will Is Not Free," p. 99.

[18] Van Inwagen, "When the Will Is Not Free," p. 111 (van Inwagen's footnote 8).

[19] We are grateful to Stewart Cohen for help with this point.

[20] We are indebted to Stewart Cohen for this point.

[21] The argument also employs the point developed in the informal presentation above, that the conditionals in the antecedents of (I), (II), and (III) exhaust the field of possibilities.

[22] Let us be clear about the nature of our disagreement. In "When the Will Is Not Free," van Inwagen suggests that we are arguing that his claim – "that we are seldom if ever able to act otherwise than we actually do" – *would* entail "the even more unpalatable conclusion that we can seldom if ever be held morally accountable for what we have done" (van Inwagen, 1994, p. 95). This is *not* our position. Since we do *not* accept van Inwagen's assumption that freedom to do otherwise is a necessary condition of moral responsibility, we do *not* think that responsibility would be jeopardized by his contention that we rarely, if ever, have this freedom. Rather, our position is as follows. van Inwagen cannot accept (1) that freedom to do otherwise is a necessary condition of responsibility and (2) that we rarely, if ever, are free to do otherwise, and still argue (3) that in most cases we can be held morally accountable for our actions (in a way that matches common sense).

[23] Van Inwagen, "When Is the Will Free?" p. 408.

[24] We have analyzed all three of the arguments and displayed their weaknesses in our earlier paper, "When the Will Is Free."

[25] Van Inwagen, "When the Will Is Not Free," pp. 100–101.

[26] "When the Will Is Free," p. 433.

[27] Similar arguments show that we could also be free in the other two classes of cases that van Inwagen discusses, i.e., cases in which (i) we do something that we have an unopposed inclination to do or (ii) we act without reflection or deliberation. Hence, van Inwagen is wrong to claim that from the truth of Beta, it follows that we rarely, if ever, are free to do otherwise. We develop these arguments in "When the Will Is Free."

[28] We employ the connective for the subjunctive conditional here, simply to make clear that we believe that this is the appropriate interpretation.

[29] Van Inwagen, "When the Will Is Not Free," p. 106.

[30] Van Inwagen, "When the Will Is Not Free," p. 105.

[31] Fischer and Ravizaa, "When the Will Is Free," p. 443.

[32] Van Inwagen, "When the Will Is Not Free," p. 107.

[33] Van Inwagen continues his explanation of this point as follows: "perhaps the speaker's use of the adverb 'morally' carries the implicature that the state of affairs under discussion is disapproved of by the speaker, despite the fact that it is possible that someone accept the proposition expressed by the speaker's utterance and not disapprove of that state of affairs." (van Inwagen "When the Will Is Not Free", p. 108).

[34] Peter Strawson, "Freedom and Resentment" reprinted in *Free Will*, ed. Gary Watson (Oxford: Oxford University Press, 1982), 59–80; and in John Martin Fischer and Mark Ravizza, ed., *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press, 1993), pp. 45–66.

JOHN MARTIN FISCHER   MARK RAVIZZA
Department of Philosophy   *Department of Philosophy*
University of California   *Santa Clara University*
Riverside CA 92521-0201   *Santa Clara, CA 95053*
USA   *USA*