PETER VAN INWAGEN

WHEN THE WILL IS NOT FREE

(Received 15 September 1993)

In "When the Will Is Free," John Martin Fischer and Mark Ravizza[1] examine and criticize several of the arguments and conclusions of my paper, "When Is the Will Free?"[2] In the present paper, I will reply to their criticisms. In order to save space, I will not recapitulate the arguments of either paper, but will rather suppose that the reader is familiar with both.

In "When Is the Will Free?", I argued for the following three theses (among others): that incompatibilists should accept the validity of the inference-rule that I had earlier called "Beta";[3] that the validity of Beta entails – for reasons quite unconnected with determinism – that we are seldom if ever able to act otherwise than we actually do; that this conclusion, while perhaps unpalatable, does not at any rate entail the even more unpalatable conclusion that we can seldom if ever be held morally accountable for what we have done. All three of these theses have been challenged by Fischer and Ravizza.[4]

I

Fischer and Ravizza contend that the incompatibilist need not accept Rule Beta owing to the fact that there are arguments for incompatibilism that do not depend on Beta. They are, of course, aware that I shall cheerfully concede that there exist sound arguments for incompatibilism that do not *explicitly* appeal to Beta. After all, I have presented two such arguments myself.[5] My position is that all (logically adequate) arguments for incompatibilism must make some sort of implicit or hidden or covert appeal to Beta.[6] Fischer and Ravizza are willing to grant that

Philosophical Studies **75:** 95–113, 1994.

this may be true as regards many of the better-known arguments for incompatibilism, but they insist that it is possible to construct arguments for incompatibilism that make no appal to Beta, however covert. To demonstrate this, they present an argument for incompatibilism that, they contend, has this feature. I shall discuss this argument presently. Before I do that, however, I want to say more than I have about what I mean by saying that all logically adequate arguments for incompatibilism must make at least an implicit appeal to Beta – and about why I think that this is true. I would not know how to go about constructing a general proof of this, but I will look at one case in detail, in the hope that it will seem plausible to the reader – as it seems plausible to me – to suppose that the lessons of this case can be generalized. I will examine one of my own arguments for incompatibilism, and ferret out its hidden commitment to the validity of Rule Beta. The argument I shall examine is the second of the three arguments for incompatibilism in Chapter III of *An Essay on Free Will* (the "access to possible worlds" argument). This 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.

Let us ask: why should someone find these two premises plausible? Why, to be specific, should someone find the second premise plausible? Well, the intuition that underwrites our tendency to assent to this proposition can be broken down and the resulting parts arranged in the form of an argument. One way to do this would be as follows. Suppose that W is a world in which some actual law, L, is a false proposition. If X has – in actuality – access to W, then X has a choice about whether W is actual. But it is a necessary truth that if W is actual then L is false. So if X has access to W, then X has a choice about whether L is true,

which is absurd. It is easy to see that, by the explicit introduction of the uncontroversial rule Alpha (an appeal to which would seem to be implicit in the reasoning) and a little trivial rearrangement, we can turn this persuasive but informal argument into the following formally valid argument (or at any rate every inference in the argument other than the inference of 6. from 5. can be justified formally):

1. $\Box$(W is actual $\rightarrow$ L is false)

hence,

2. $\Box$(L is true $\rightarrow$ $\neg$ W is actual)

hence,

3. N(L is true $\rightarrow$ $\neg$ W is actual) [Rule Alpha]
4. N(L is true)

hence,

5. N($\neg$ W is actual) [3,4 Rule Beta]

hence,

6. No one has access to W.

If one did not accept the validity of Beta, one would not have to accept the validity of this argument. If one did not accept the validity of this argument, one would not – at least so far as I can see – have to accept the validity of the informal argument from which it is derived. And if one did not accept that argument, then I can't see what reason one would have for accepting the second premise of the "access to the worlds" argument for incompatibilism. (Similar reasoning, of course, applies to the first premise.) One could simply say, "Oh, I have access to some worlds in which L is false. That is, some worlds in which L is false are such that I have a choice about whether they are actual. But –

since I reject Beta – I do not admit that this commits me to the absurd conclusion that I have a choice about whether L is true."[7]

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. (Whether or not Beta really is valid, it *seems* self-evidently valid to many people, and lurking inarticulately in the background of Alice's arguments is something that propositions that seem self-evident to Alice – or would if Alice thought about them at all – are very good at.) At any rate, this is what I meant by saying that the incompatibilist should accept Beta: there will be some premise or premises in any technically satisfactory argument for incompatibilism that the incompatibilist would have no reason to accept if he did not accept the validity of Beta.

Fischer and Ravizza, as I have said, believe that they have presented an argument that is a counterexample to this general thesis. Unfortunately, their argument is logically defective. The argument employs two premises, the "principle of the fixity of the past," and the "principle of the fixity of laws":

> 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.

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. At any rate, these conclusions do not follow in the most plausible version of counterfactual logic, David Lewis's. If Lewis's counterfactual logic is correct, '$p \mathrel{\Box\!\!\rightarrow} q. \vee .p \mathrel{\Box\!\!\rightarrow} r$' cannot be deduced from '$p \mathrel{\Box\!\!\rightarrow} .q \vee r$'; for suppose that in all of the closest p-worlds either q is true or r is true, although q is false in some of the closest p-worlds and r is false in the others.

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):

$$(p \mathrel{\Box\!\!\rightarrow} q) \rightarrow s$$

$$(p \mathrel{\Box\!\!\rightarrow} r) \rightarrow s$$

hence,

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

And this argument-form is invalid, for essentially the same reason that the argument-form whose invalidity was shown in the preceding paragraph is invalid.[8] So the argument for incompatibilism that Fischer and Ravizza have constructed is invalid. I conclude, therefore, that we have 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.

II

In "When Is the Will Free?", I discussed several cases in which the validity of a certain "Beta-like" rule, Beta-prime, implied – or so I argued – that the agents described in those cases had no choice about

some matter. (Beta-prime is, essentially, an agent- and present-time-indexed version of Beta. I contended that anyone who accepted Beta should accept Beta-prime.) I argued that the sorts of circumstances represented by these cases were common enough that, if I was right about the agent's having no choice in each of the representative cases, "having a choice" is something that occurs only rarely – if it occurs at all. Fischer and Ravizza contend that I am wrong about each of these cases. The arguments that I used to show that in each of these cases the agent had no choice did not differ from one another in any reality important way, and, as a consequence, the reasons Fischer and Ravizza give for thinking that in each case I was wrong do not differ from one another in any really important way. To save space, I am going to examine their discussion of only one of the cases. I do not think that the reader will find it difficult to adapt what I say concerning their remarks about this one case to their remarks about the others. The case I shall discuss is of this general sort: A certain act is proposed to me; I regard this act as morally reprehensible, and, although one certainly might be *tempted* to perform an act that one regarded as morally reprehensible, in this case I am not even tempted to perform it. Let us say the case is this: A colleague who believes that I do not want Smith to become Chair of the Tenure Committee suggests that I attempt to block his all but inevitable appointment to that position by reporting, falsely, that I have heard him maintain that women are incapable of serious scholarly work (my colleague is a notorious enemy of Smith's and realizes that the lie would have to be told by someone else to be believed); I regard bearing false witness against one's neighbor as morally reprehensible; as a matter of fact, my colleague is misinformed not only about my principles but about my preferences, for I am not at all opposed to Smith's becoming Chair.[9]

I argued that in this case 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 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 relevant 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'.

Fischer and Ravizza object to this argument on the following ground. Grant for the sake of argument that I cannot do A unless at some point I have some sort of desire to do A. It does not follow from this premise that if I regard A as indefensible (and therefore "lack any positive desire to do A"), I cannot do A. This does not follow because from the premise that I have no desire to do A it does not follow that it is not within my power to have a desire to do A.[10]

That I might have it within my power to have a desire to perform A when I in fact have no such desire is shown by the following case. (The case is mine, but it is of the same general sort as the cases they appeal to, and I do not think that it has any special features that misrepresent what they had in mind.) Suppose that I am subject to a general worry about whether I have all my life been attempting to deny responsibility for my acts by implicitly holding that certain acts are – for moral reasons – simply out of the question for me. As a result of reading Sartre, I decide that if I do implicitly hold this, then I am guilty of *mauvaise foi*, a cowardly attempt to deny the awful freedom to which I am condemned by the very fact of being able to see various contemplated acts as alternatives. These reflections create in me the following desire: to perform an act in contradiction with my deepest moral principles (or the features of my consciousness that I sometimes describe that way; but that description is tainted with *mauvaise foi*, since the only way in which a moral principle

can become *mine* is by my acting on it), an *acte gratuit*. Let us add these further suppositions to the case we have constructed, the case of the proposed lie about Smith.

At *t*, my colleague suggests that I tell the lie about Smith. Because I regard the proposed course of action as morally reprehensible, I experience an upsurge of moral revulsion. (And I can't help being aware of the fact that I have no desire to block Smith's appointment by any means, fair or foul). Suddenly, however, "Sartrean" thoughts stir in my mind. I think of my long-standing desire to perform an *acte gratuit*, and it is borne in upon me that one way to satisfy this desire would be to do just what my colleague has proposed. Let us suppose that I thrust the desire to perform an *acte gratuit*, and the reflections concerning my present situation that accompanied it, out of my mind and indignantly refuse my colleague's suggestion. But suppose that if I had not cleared my mind of these things, the desire to perform an *acte gratuit*, together with the other features of my mental landscape at that moment, would shortly have issued in a desire to do A. (Perhaps I was at some level aware of the truth of this counterfactual, and this is one of the reasons I had for hastily thrusting the "Sartrean" thoughts out of my mind.)

If my situation is as we have imagined, the proposition

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

may well be false. At any rate it is not clearly true. For if I had not pushed those thoughts out of my mind – and let us suppose that I had a choice about whether I did this –, then I should have had a desire to do A, and, although this desire would, from its inception, have been "warring against the law of my mind" (against my moral convictions), it might be that I should have a choice about who would win the war. (The "Sartrean" desire that the moral convictions lose the war would also be "in play.") It is important to realize that if the potential desire to do A did come to actuality, the antecedent of the embedded conditional would still be true, for that proposition contains an implicit reference to the present, to *t* or to a moment shortly thereafter; it would be a bit *later* that I came to desire to do A, and thus came to be in a condition that violated the "no positive desire" clause in the definition of 'regard

as indefensible'. It might therefore be that there is at t a "path into the future" – one open to me – that ends in a possible situation in which the antecedent of the embedded conditional is true and its consequent false. And, it would seem, if there is a path to this situation, and if I am able to follow it, then I have a choice about the truth-value of the conditional.

Let us assume that I do have a choice about the truth-value of the conditional (that is, that I have a choice about this in the "augmented" or "Sartrean" version of the story). At most this proves that in *some* possible cases one is able to perform an act that one regards as morally indefensible. I will concede this. The important question is: In what proportion of the possible cases in which one regards an act as morally indefensible is one able to perform it? If the answer is "Only a very small proportion," conceding that such cases exist will have no consequences for my main thesis – that if Beta is valid, then there is precious little free will. And I believe that the answer is indeed "Only in a very small proportion," for it seems that any circumstances that might have consequences relevantly similar to those of the augmented case must be very rare. Suppose we consider not only Sartrean existential bemusement of the kind I have imagined, but any sort of episode whatever in which I have an unrealized potential for acquiring a desire to perform a proposed act that I regard as morally reprehensible. There have been occasions in my life in which I have been in a situation of the following general type:

> Something has suggested to me that I undertake a certain course of action (perhaps another person has suggested it, or perhaps the suggestion has for some unguessable reason arisen from the depths of my unconscious); I regard the proposed course of action as morally reprehensible; at the moment the course of action is suggested to me, and for at least a short while afterwards, I not only regard it as reprehensible, but I haven't the least desire to undertake it; I'm not even *tempted* by the suggestion that I undertake it.

I suppose that it is probably true that, in a certain proportion of these episodes, the passage of a moment of time has produced the following sequel:

> ... but a moment later I do have some desire to undertake this course of action; I *am* to some degree tempted by the suggestion that I undertake it.

But I am quite sure that the proportion of these episodes that have had this sequel is very, very small: in almost all such episodes, no desire to perform the proposed act came to be. (In my own case, certainly, these rare episodes have not been of the recondite "Sartrean" variety,[11] but have been simply episodes in which I suddenly thought of some advantage I'd gain by doing what was proposed, an advantage that had not at first occurred to me.) The conviction I have just recorded is a conviction about my actual past. I also have a corresponding modal or counterfactual conviction, a conviction about the content of the regions of logical space in the immediate vicinity of those occupied by my actual past: In very few of these episodes has there been some non-actual state of affairs that was *close* to actuality, and which was such that, had it become actual, I should have – or should probably have or might well have – acquired a desire to perform the proposed act.

I believe that the schema

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

is true for all, or almost all, of the cases in which 'A' represents a course of action that has actually been proposed to me, and which, at the time the proposal was made, I regarded as morally indefensible. The intuitive picture on which this belief rests could be articulated as follows. Suppose that carrying out A has just now been proposed to me, and that I now regard A as morally indefensible. If I now have a choice about whether I am going to do A a moment from now, then there must be some coherently describable path through logical space from my present condition to my doing A a moment from now, and I must now be *able* to follow this path, must be able to negotiate every twist and turn in it. And, in the vast majority of cases in which, at a certain moment, I regard a proposed course of action as morally reprehensible (and in which I am not at that moment even tempted to undertake this course of action), there is no such path. In "When Is the Will Free?", I tried to make the absurdity of supposing that such a path through logical space existed by imagining the following exchange: "Imagine that [someone] X *does* do A [in the circumstances imagined]. We ask

him, 'Why did you do A? I thought you said a moment ago that doing A would be reprehensible.' He replies:

> Yes, I did think that. I still think it. I thought that at every moment up to the time at which I performed A; I thought that while I was performing A; I thought it immediately afterward. I never wavered in my conviction that A was an irremediably reprehensible act. I never thought there was the least excuse for doing A. And don't misunderstand me: I am not reporting a conflict between duty and inclination. I didn't *want* to do A. I never had the least desire to do A. And don't understand me as saying that my limbs and vocal cords suddenly began to obey some will or other than my own. It was *my* will that they obeyed. It is true without qualification that *I* did A, and it is true without qualification that I *did* A." (pp. 408–9)

It is true – this is the essence of Fischer and Ravizza's point – that such a confession would not *necessarily* have to be a correct description of an episode in which an agent began by regarding a certain act as morally reprehensible (and was not even tempted to perform that act) and ended by, nevertheless, performing that act. It could be that in an episode that started this way there was a potential motive for performing A that was pretty close to being actually present to the agent's mind; it might even be that the agent had a choice about whether this potential motive should become actually present to his mind. The potential motive might be of some philosophically recondite sort (a "Sartrean" or "Augustinian" or "Dostoevskian" motive), or it might be something much more mundane, like the recognition of a momentarily overlooked advantage. In the vast majority of cases, however, there will be no such potential motive, no reason for performing A that is lurking somewhere nearby in logical space. And – to return to my own case – even when there is such a potential motive for my performing A, the existence of this at present merely potential motive will not be relevant to the truth-value of

$$N\ I, (I \text{ regard A as indefensible} \rightarrow I \text{ am not going to do A})$$

unless I now have a choice about whether it is shortly to become one of my actual motives – unless I am now *able* to follow one of the "paths into the future" that passes through my acquiring that motive for performing A.

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 cases in which such potential motives so much as exist are very rare. (And it may well be that only a small proportion of the cases in which the potential motives exist are cases in which I have a choice about whether they are to become my actual motives.) And, of course, I suppose that everyone or almost everyone is like me in this respect.

I therefore continue to insist that if the inference-rule Beta is valid, then cases in which one is able to do otherwise are rare indeed.

III

And I continue to insist that it does not follow from this result that cases in which one can be held morally accountable for the consequences of what one does are rare. In "When Is the Will Free?", I argued that one can be held to moral account for the consequences of an act that one could not at the time have refrained from performing if one can be held to moral account for having had this inability. And I argued that one's present inability to act otherwise in a large class of cases might be due to features of one's character that were the consequences of one's past free choices. Fischer and Ravizza, however, argue that this will not do:

> In the end, however, ... this strategy must fail. 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. (p. 443)

They support this contention by asking the reader to consider the case of a young woman named Betty. Betty is a conventionally patriotic American (this feature of her character is a product of her early socialization, and has never been "tested"). An agent of a hostile foreign power mistakes Betty for someone else and offers her a certain amount

of money if she will betray the United States, an offer she indignantly refuses. (She is not even tempted by the money she has been offered, and her refusal is immediate and unreflective.)

I will grant for the sake of argument that Betty had, in this situation, no choice about whether to refuse the offer. (I should want to build a bit more into the case before I was willing to regard this as a logical consequence of the case, but the other things that would have to be added could be added without affecting the points at issue.) I will also grant for the sake of argument that those features of Betty's present character that are responsible for her now having no choice about whether to refuse the offer are not due to free choices that Betty made at earlier points in her life.[12] (Again, I should want to build a bit more into the case before I was willing to regard this as a logical consequence of the case.) Since I have (as I said in "When Is the Will Free?") a "classical" conception of the relationship between moral accountability and the ability to do otherwise, I should certainly not want to accept the thesis that Betty was morally accountable for having declined to betray her country. Fischer and Ravizza take this to be a *reductio* of my position. They say, "But such a conclusion runs directly counter to our actual practices of holding people morally responsible" (p. 444).

It is far from clear to me whether this is actually true. For one thing, the "actual practices" that we employ in deciding to hold someone morally accountable for a certain state of affairs are almost entirely directed at states of affairs that we regard as ones that ought not to obtain. When we are deciding whether to hold someone morally accountable for x, we are normally trying to determine whether that person is to *blame* for x. I concede that we sometimes say things like "Find out who the people are who are responsible for the excellent safety record in District Three and give them all bonuses." But 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 . . . ." Now this oddness may be only a matter of "conversational implicature"; 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. I will not therefore insist on rejecting out of hand any general, theoretical account of moral accountability that has the consequence that an agent can be held morally accountable for good or indifferent states of affairs. It is clear, however, that "our actual practices of holding people morally responsible" in some way incorporate an asymmetry between bad and good or between approval and disapproval. And it is certainly true that *typical* ascriptions of morally accountability for a state of affairs are cases in which the ascriber thinks the state of affairs a bad thing. If, therefore, Fischer and Ravizza really wish to raise the question whether my theory has implications that "run directly counter to our actual practices of holding people morally responsible," they ought at least to begin with cases in which the state of affairs for which moral accountability is being sought is one that is generally thought bad or is disapproved of by the speaker or something of the sort. If such a case could be found, it would better support their position than the case they have actually used. If no such case could be found, they ought to be ready to explain why the only cases that even seem to support their position against mine are the atypical cases in which it is asked whether an agent is to be held morally accountable for a state of affairs of which the speaker approves.

Let us therefore look at a case that is like the case of Betty, but in which the states of affairs that are being enquired about are ones that all of the readers of this paper would agree ought not to obtain:

> Hansi is an enthusiastic fifteen-year-old member of the Hitler Youth. (The year is 1944.) Hansi hates Jews. (At any rate, Hansi hates a class of people he calls 'Juden'; he has no clear memory of any individual Jews, for all the Jews in his neighborhood disappeared when he was a small boy, and his family and their friends never even allude to their former neighbors; he has beliefs about "Juden" only as a class, and the characteristics he ascribes to the members of this class are the ones he has been taught to ascribe to them in school and at Hitler Youth meetings; it is quite literally true that if his teachers had told him that "Juden" had horns and tails, he would have believed them – for the same reason that he believed them when they told him that Mars had two moons.)

Is Hansi morally accountable for the fact that he hates Jews (or however, exactly, the intentional object of his hatred should be described)? Can we blame him for the fact that he believes that the defeat of Germany in the First World War was due to the machinations of a cabal of Jewish

plutocrats? Is the fact that he once enthusiastically carried a banner at anti-Jewish rallies something that we can morally condemn him for? I would think that "our actual practices of holding people morally responsible" dictate that we *not* blame him – or, to use the philosophers' phrase, "not hold him morally responsible" – for any of these states of affairs (his hatred of Jews in 1944, his beliefs about Jews in 1944, his participation in anti-Jewish rallies in 1944). After all, unlike his elders, he had no choice about what German youth was told about Jews in the decade preceding 1944. Unlike his elders, he had no choice about whether he believed what he was told about Jews. Most of his elders had actually met Jews, and thus had first-hand evidence that Jews were wholly unlike the official Nazi picture of Jews. It is therefore not unreasonable to suppose that they had a choice about whether they would believe the evidence of their senses or believe Nazi propaganda. Whether or not they did have a choice about this, Hansi certainly did not, for he did not have anything corresponding to "the evidence of their senses." And, although this fact may not always be immediately evident to the parents of teenagers, people Hansi's age have a (no doubt biologically based) tendency to believe what they are told by adults in positions of authority over them: if they did not, socialization would be impossible. (And this is a much stronger tendency than the tendency of adults, even German adults, to believe what they are told by the government.) All of the things that I have said in exoneration of Hansi are things that we all believe, at least in some inarticulate form. That is why people who were fifteen-year-old members of the Hitler Youth in 1944 are not regarded by anyone as having been disqualified by that membership for high office in present-day Germany.

I would judge, therefore, that in cases in which one enquires about who is morally accountable for states of affairs one thinks bad, the theses argued for in "When Is the Will Free?" do not run counter to our usual practices for making judgments about moral accountability. What about cases involving states of affairs one does not think bad? What about cases like Betty's?

Insofar as I can make any sense of the question whether Betty (in the situation Fischer and Ravizza have imagined) should be held morally accountable for the state of affairs *Betty declines to betray her country*

– and the question does seem to me to be a very odd one –, I can see no reason to say that a theory of moral accountability that has the consequence that she should not be held accountable for this state of affairs runs counter to our actual practices for making such judgments.

NOTES

[1] *Philosophical Perspectives* 6, 1992, pp. 423–451.

[2] *Philosophical Perspectives* 3, 1989, pp. 399–422.

[3] In *An Essay on Free Will* (Oxford: at the Clarendon Press, 1983), p. 94.

[4] I must say, I should be delighted if Fischer and Ravizza were right about the first two theses. (If they were right about the first two, it would be fine with me if they were also right about the third. The "worst-case scenario," of course, would be their being wrong about the first two theses and right about the third.) I don't like to be refuted any more than anyone else does, but I regard the conclusions of "When Is the Will Free?" as one of the more serious challenges to the plausibility of the incompatibilist position defended in *An Essay on Free Will.* (When I read a draft of the paper at a conference on free will at McGill University in 1986, Dan Dennett said, "Thank you, Peter, for the lovely *reductio* of incompatibilism," and one can appreciate the force of his point without accepting it.) If I were convinced that Fischer and Ravizza were right, I could resume my dogmatic slumbers. But I am not convinced that they are right.

[5] See the first two of the "Three Arguments for Incompatibilism" that are the subject-matter of Chapter III of *An Essay on Free Will.* (The third argument of that chapter explicitly appeals to Rule Beta.) These two arguments were originally presented in "The Incompatibility of Free Will and Determinism," *Philosophical Studies* 27 (1975), and "A Formal Approach to the Problem of Free Will and Determinism," *Theoria* XL (1974) Part 1.

[6] That this – or something very close to it – was so was conjectured by Michael Slote in "Selective Necessity and the Free Will Problem" *The Journal of Philosophy* 79 (1982) pp. 5–24. It should be noted that Slote's article and Chapter III of *An Essay on Free Will* are entirely independent of each other. The roots of the third argument for incompatibilism in Chapter III of *An Essay on Free Will* are in Carl Ginet's "Might We Have No Choice?", which appeared in Keith Lehrer (ed.) *Freedom and Determinism* (New York: Random House, 1966), pp. 87–104.

[7] And this is very like what the compatibilist *would* say in response to the "access" argument for incompatibilism. The compatibilist would say something like this: "I have access either to worlds in which the laws are different or the past is different – or to worlds of both kinds. But since I reject Beta [the compatibilist must, of course, reject

Beta], I am not thereby committed to the absurd conclusion that either I have a choice about what the laws are or that I have a choice about how things were in the past (much less both)."

[8] Take the counterexample to that argument-form that was given in the text: this is a case in which both $p \Box\!\!\rightarrow q$ and $p \Box\!\!\rightarrow r$ are false and $p \Box\!\!\rightarrow .q \vee r$ is true; suppose that s is false, and you have a counterexample to the more complex argument-form. If this proof is too abstract for your tastes, consider the following case. 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. (Whether or not this does follow from our description of the device, let us assume that we have a device of which it is a correct description.) Let p be 'You pressed the button', q be 'The green light flashed', r be 'The red light flashed' and s be 'Pressing the button would have a determinate outcome'. (I have chosen this "relevant" consequent for its intuitive force: with this consequent, the truth of the two premises is more than a mere consequence of the truth-values of their antecedents and consequents and the truth-table for '$\rightarrow$'. I hope no one is going to tell me that by assuming that 'Pressing the button would have had a determinate outcome' is false, I am begging the question. Remember that any false consequent – 'The moon is made of green cheese', for example – would yield a counterexample.)

[9] I have changed a few of the details of this case from the way it was presented in "When Is the Will Free?". The only change worth noting is this: in the case as it was first presented, I preferred that Smith not become the Chair of the Tenure Committee. I gave the case that feature (I think) because I assumed that the kind of desire that it would be appropriate for one to express by a statement like "I prefer that that not happen" was very weak compared with the kind of desire it would be appropriate for one to express by a statement like, "If I acted that way I'd be doing something morally indefensible" – for, surely, to regard a proposed course of action as morally indefensible implies having a desire not to perform it. But, strictly speaking, I should have considered a case in which my moral convictions were not opposed by any desire, even a very weak one.

[10] In Chapter II of *An Essay on Free Will*, I introduced a way of indicating the "scope" of ascriptions of ability by means of brackets. Using these brackets, Fischer and Ravizza's point could be put very compactly by saying that they charge that my argument conflates the following two theses:

I can (do A) when I regard A as indefensible.

I can (do A when I regard A as indefensible).

Their point could be put this way (it is interesting to compare this criticism of my argument with my criticism of the fatalist's argument in Chapter II of *An Essay on Free Will*): the latter is – or we grant this for the sake of argument – indeed false, even necessarily false, but its falsity does not support my conclusion; the falsity of the former would support my conclusion, but I have not shown that it is false.

[11] Or "Augustinian" or "Dostoevskian." See "When the Will Is Free," pp. 436–439. The terms are based in the episode of the stolen pears in *Confessions* and the murder of the pawnbroker in *Crime and Punishment*. An Augustinian desire may be defined as a desire for something evil precisely because it is evil ("Evil be thou my Good," as Milton's Satan says). A Dostoevskian desire may be defined as a desire to place oneself "beyond good and evil." It should be noted that the actual episodes in *Confessions* and *Crime and Punishment* raise no difficulties for my position, for it is clear that the young Augustine did desire to steal the pears and clear that Raskalnikov did desire to murder the pawnbroker: Neither of them did something that he had no positive desire to do. (And the desire in each case was a very strong desire.) The springs of these desires may have risen in the depths of ruined human nature or in deplorable philosophical theories. They may have been perverse desires. Nevertheless, they were as real as the most normal, everyday desire. We can, however, easily imagine other cases: cases in which, owing to the agent's fallen human nature or owing to the agent's addiction to the works of Nietzsche, although the agent does *not* desire to join in the pear-stealing expedition that his peers have proposed to him, or, this very evening, to seek out and murder the aged pawnbroker, he nevertheless has it within his power to desire to do these things.

[12] Fischer and Ravizza have an argument ready in case I am not willing to make this concession. Despite the fact that I am willing to make the concession, I want to discuss the argument because it embodies an important misconception about my general position – a misconception about what I mean by saying that only "rarely" are we free to do otherwise than we in fact do. I discuss the argument at this point because the following quotation (which contains the argument) will be intelligible only to someone who has the story of Betty in mind

> Of course the restrictivist might object that Betty really is responsible for her disposition to patriotism. "Undoubtedly" – the argument goes – "there must have been many more small conflict situations in her life than you have allowed for (or she is even aware of), and these situations taken together account for her present disposition." However, to make such a concession would prove fatal to the restrictivist's position, for it would undermine his central thesis that rarely, if ever, are we in one of these situations in which we are free to do otherwise. (p. 444)

(The "restrictivist" is someone who, like me, allows only a very restricted scope for free will.) I would point out that one no doubt performs hundreds or even thousands of acts every day. Even if there were hundreds of cases every year in which one was free to act otherwise than one did, therefore, it could be that free action was pretty uncommon.

It is also worth noting that it might be that free actions were relatively common during the years of the formation of one's character – childhood through early adulthood – and relatively rare in one's later years.

Department of Philosophy
Syracuse University
Syracuse, NY 13244-1170
USA