

Ridgeview Publishing Company

When the Will Is Free

Author(s): John Martin Fischer and Mark Ravizza

Source: *Philosophical Perspectives*, Vol. 6, Ethics (1992), pp. 423-451

Published by: [Ridgeview Publishing Company](#)

Stable URL: <http://www.jstor.org/stable/2214255>

Accessed: 24/02/2011 13:37

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

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

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

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



Ridgeview Publishing Company is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Perspectives*.

<http://www.jstor.org>

WHEN THE WILL IS FREE

John Martin Fischer and Mark Ravizza
University of California, Riverside

Incompatibilists usually direct their attention to the following worry: if the thesis of causal determinism is true, then none of us is free to do other than what he actually does. But although causal determinism poses the most frequently discussed threat to freedom for incompatibilists, it may not be their only source of worry, at least not for those incompatibilists who also accept the common intuition that most of us, most of the time, are free to do otherwise.

In his article "When Is The Will Free?"¹ Peter van Inwagen offers a creative and systematic development of this less often discussed side of incompatibilism. He maintains (1) that anyone who is an incompatibilist should accept a rule of inference which he calls 'Beta,' and (2) that "anyone who accepts Beta should concede that [even if causal determinism is false] one has precious little free will, that rarely, if ever, is anyone able to do otherwise than he in fact does" (p. 405).² We will call the position suggested by van Inwagen's arguments 'restrictive incompatibilism' or 'restrictivism' for short. This name seems appropriate because restrictive incompatibilists hold both that incompatibilism is true, and that anyone who accepts the truth of incompatibilism must also (in virtue of accepting Beta) accept radical restrictions on one's ability to do otherwise.³ According to this position, if causal determinism is true, we never are free to do otherwise, and if causal determinism is false, we "rarely, if ever" are free to do otherwise.

The conclusion that all incompatibilists allegedly must accept—that we, at best, are only rarely free to do otherwise—will come as a shock to many. However, this conclusion becomes even more disquieting when it is combined with another assumption of the "classical tradition" which van Inwagen and many other incompatibilists embrace. This is the assumption that freedom to do otherwise is a necessary condition of moral responsibility. Accept this premise and the following worry quickly arises: If the restrictivist is right—if incompatibilists are committed to a severe restriction on one's ability to do otherwise—then must they not also accept a similar limitation on the range of states of affairs for which one can be held morally responsible? And if this

is the case, then would not incompatibilism itself seem to be incompatible with many of our most deeply held beliefs about the type of respect, praise and blame merited by persons?

In what follows we will address these issues by outlining the arguments for restrictive incompatibilism and then by discussing some responses to this position. In particular, we will argue that one can accept incompatibilism without a *fortiori* being committed to the restrictivist position. That is, we will maintain that one can accept the thesis that freedom to do otherwise is incompatible with causal determinism without implicitly being committed to the further conclusion that “rarely, if ever, is anyone able to do otherwise than he in fact does” (p. 405). In taking this position, we do not intend to argue *for* the truth of causal determinism or incompatibilism, but rather to argue *against* the restrictivists’ claim that the logic behind the incompatibilist position requires that any incompatibilist also accept severe restrictions on freedom to do otherwise. Finally we will argue that, irrespective of our earlier criticisms, restrictive incompatibilists cannot (as van Inwagen suggests they can) provide a satisfying theory of moral accountability while still remaining within the classical tradition (which accepts that such accountability requires freedom to do otherwise).

I. From Incompatibilism to Restrictive Incompatibilism

Let us begin by considering why incompatibilists purportedly must find themselves with little, if any, freedom to do otherwise. The restrictivist argues that the incompatibilist position rests upon a rule of inference termed ‘Rule Beta’. Beta says that “from Np and $N(p \supset q)$ deduce Nq ” (where “‘ Np ’ stands for ‘ p and no one has or ever had any choice about whether p ’”) (pp. 404-5). To appreciate the reason for this stress on Rule Beta we need only digress for a moment to consider one form of the Consequence Argument which persuades the restrictive incompatibilist that free will is not compatible with causal determinism:⁴

Rule Alpha: From $\Box p$ deduce Np .

(‘ \Box ’ represents “standard necessity”: truth in all possible circumstances.)

Rule Beta: From Np and $N(p \supset q)$ deduce Nq .

Now let ‘ P ’ represent any true proposition whatever. Let ‘ L ’ represent the conjunction into a single proposition of all laws of nature. Let ‘ Po ’ represent a proposition that gives a complete and correct description of the whole world at some instant in the remote past—before there were any human beings. If determinism is true, then $\Box(Po \ \& \ L \ \supset \ P)$. We argue from the consequence of this as follows.

- | | |
|--|-------------------------------|
| 1. $\Box(P_o \ \& \ L. \ \supset \ P)$ | |
| 2. $\Box(P_o \ \supset \ (L \ \supset \ P))$ | 1; modal and sentential logic |
| 3. $N(P_o \ \supset \ (L \ \supset \ P))$ | 2; Rule Alpha |
| 4. NP_o | Premise |
| 5. $N(L \ \supset \ P)$ | 3,4; Rule Beta |
| 6. NL | Premise |
| 7. NP | 5,6; Rule Beta |

If the above argument is sound, then determinism entails that no one has a choice about what she does; hence, determinism is incompatible with freedom to do otherwise. Since the restrictive incompatibilist insists that no one reasonably could take issue with either of the premises or Rule Alpha, he concludes that the soundness of the incompatibilist argument depends upon the validity of Beta. Indeed van Inwagen goes so far as to say that “if one accepts Beta, one should be an incompatibilist, and if one is an incompatibilist, one should accept Beta” (p. 405).

The next step in the restrictivist argument is to claim that any person who accepts Beta should also accept a similar rule of inference termed ‘Beta-prime’. Beta-prime tells us that “from Nx,p and $Nx,(p \supset q)$ 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’) (p. 408). When an agent has no choice about whether a proposition (or statement) obtains we will say that that proposition is “power necessary” for him.

Finally the restrictivist presents three arguments to show that if Beta-prime is valid then (even if causal determinism is false) we are not able to do otherwise in three types of cases which represent the majority of all actions. We will discuss the details of these arguments in section III; here we need only note the three cases. The first is one of duty unopposed by inclination; that is, “no one is able to perform an act he considers morally reprehensible” (p. 405). The second case is one of unopposed inclination; thus, “no one is able to do anything that he wants very much *not* to do and has no countervailing desire to do it” (p. 406). The third case is one in which we act without reflection or deliberation; thus “if we regard an act as the one obvious thing or the only sensible thing to do, we cannot do anything but that thing” (p. 406). Given these points, the restrictivist concludes that the only times an agent *is* free to do otherwise are times in which the agent is confronted with conflicting alternatives such that, even after reflection, it is not obvious to him what to do. Such conflict situations, van Inwagen tells us, occur *rarely* and can be divided into three general categories: (1) “Buridan’s Ass” cases,⁵ (2) cases in which duty or general policy conflicts with inclination or momentary desire, and (3) cases in which one must choose between incommensurable values.⁶

Since the restrictivist holds that all incompatibilists must accept Beta and

hence Beta-prime, and that anyone who accepts Beta-prime must concede that we cannot do otherwise in the cases which make up the majority of all our actions, he therefore concludes that “the incompatibilist must hold that being able to do otherwise is a comparatively rare condition, even a very rare condition” (p. 404).

II. Who Needs Beta?

Before getting involved in the details of the restrictivist’s individual arguments, we should note that an immediate way to circumvent his conclusion is simply to deny the initial contention that “... if one is an incompatibilist, one should accept Beta.”⁷ To support this denial a non-restrictive incompatibilist could simply refer to any of a number of formulations of the Consequence Argument for incompatibilism which do not explicitly make use of modal principles akin to Beta.⁸ If valid, these arguments apparently would give one reason to accept incompatibilism without also requiring one to accept Beta or Beta-prime. Then, even if all of the restrictivist’s remaining arguments should prove to be valid, one could accept incompatibilism without having any corresponding commitment to accept the restrictivist’s conclusion that we rarely, if ever, have free will.

In response to this type of objection the restrictive incompatibilist might insist that all formulations of the Consequence Argument, even those which aren’t explicitly formulated using Beta, must implicitly depend upon *some* rule of inference similar to Beta. Such a response is suggested by van Inwagen’s own claim that all three of his formulations of the Consequence Argument in *An Essay On Free Will* should “stand or fall together.”⁹ This claim is particularly germane to our discussion, because only van Inwagen’s third argument explicitly depends upon Beta. Nevertheless, he writes: “I am quite sure that any specific and detailed objection to one of the arguments can be fairly easily translated into specific and detailed objections to the others; and I think that any objection to one of the arguments will be a good objection to *that* argument if and only if the corresponding objections to the others are good objections to *them*.”¹⁰

Van Inwagen is not alone in holding this view. Even some compatibilists, who in other respects want to take issue with van Inwagen’s reasoning, agree with his intuition that any respectable form of the argument for incompatibilism must depend upon some type of inference akin to Beta. Pursuing this intuition, such compatibilists have sought to attack the incompatibilist’s position by blocking the modal inference on which it purportedly rests. One such “beta-blocker,” Michael Slote, writes: “I want to argue, in particular, that the arguments of GLVW [Carl Ginet, James Lamb, Peter van Inwagen, and David Wiggins] all rest on the questionable form of inference, the very inference from the double modality of ‘Np’ and ‘N(p \supset q)’ to ‘Nq’ which marks

the superiority of the new kind of argument to earlier defenses of incompatibilism.”¹¹ Further support for this position is found in Terence Horgan’s comment that “Slote has described well the deep family resemblances among the various formulations [of the Consequence Argument for incompatibilism], and he too has suggested that the different versions probably stand or fall together.”¹² This shared opinion on the part of compatibilists and incompatibilists alike, along with the debate over the validity of Beta to which it has given rise, support the restrictivist’s contention that anyone who accepts any formulation of the Consequence Argument implicitly is committed to accepting Beta. Thus, the restrictivist might seem to be on firm ground when he insists that “if one is an incompatibilist, one should accept Beta” and with it (assuming the soundness of his subsequent arguments) restrictive incompatibilism.

Nevertheless, we want to argue that this claim is false. Admittedly, many formulations of the Consequence Argument do depend upon intuitions similar to those which underlie Beta. However, the argument for incompatibilism can be formulated in such a way that it does not explicitly make use of Beta, and hence the onus remains on the restrictivist to show how such arguments do, in fact, commit their proponents to accepting Beta.¹³

To illustrate this point consider the following sketch of an argument which is adapted from a parallel argument concerning the incompatibility of God’s foreknowledge and free will.¹⁴ The argument rests upon two principles which are controversial though not implausible. The first principle expresses the fixity of the past; it says not only that one cannot causally affect the past, but also that one cannot so act that the past would have been different from what it actually was. The fixity of the past principle can be formulated as follows:

- (FP) 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 at T do Y at T.

The second principle expresses the fixity of the laws; in a manner similar to FP it says not only that one cannot causally change the laws, but also that one cannot so act that the laws of nature would have been different from what they actually are.¹⁵ The fixity of the laws principle can be formulated as follows:

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

Now consider some act X which agent A actually refrains from doing at T₂. Taking determinism to be the thesis that a complete description of the world at T in conjunction with a complete formulation of the laws entails

every subsequent truth, then if determinism is true, and S_1 is the total state of the world at T_1 , one of the following conditionals must be true:

- (1) If A were to do X at T_2 , S_1 would not have been the total state of the world at T_1 .
- (2) If A were to do X at T_2 , then some natural law which actually obtains would not obtain.
- (3) If A were to do X at T_2 , then either S_1 would not have been the total state of the world at T_1 , or some natural law which actually obtains would not obtain.

But if (1) is true, then (via FP) A cannot do X at T_2 ; similarly, if (2) is true, then (via FL) A cannot do X at T_2 . Finally, if (1)'s truth implies that A cannot do X at T_2 and (2)'s truth implies that A cannot do X at T_2 , then it follows that if (3) is true, then A cannot do X at T_2 . The conclusion of this argument is that if determinism is true, then A cannot do anything other than what he actually does at T_2 . Generalizing this result, the incompatibilist claims that if determinism is true none of us is free to do other than what he does.

The importance of the argument for our purposes, however, is not to raise yet one more banner for incompatibilism.¹⁶ Rather, the argument serves to illustrate that the debate over incompatibilism should not be reduced to a discussion about the validity of Beta. Incompatibilists share basic beliefs about the relationships between free will, determinism, the fixity of the past and the fixity of the laws. But these beliefs can find expression in different forms of argument, not all of which necessarily involve the same commitments. Such arguments show that an incompatibilist can consistently adhere to her position without automatically being committed to Beta or restrictive incompatibilism.¹⁷

As we pointed out above, van Inwagen claims that incompatibilism depends upon Beta, but we have presented an argument for incompatibilism which does not appear to depend in any way upon Beta. We thus conclude that van Inwagen's claim is false. Further, if we are correct, then Slote's strategy (in what has been described by Dennett as a "pioneering article") is not nearly so promising as it might have been supposed to be.¹⁸ Slote alleges that there are counterexamples to modal principles structurally analogous to Beta, and he suggests that Beta is similarly flawed. He concludes that incompatibilism should be rejected. We take issue with Slote's claim that Beta is flawed.¹⁹ But what is relevant to our discussion here is that *even if Slote were correct and Beta were invalid*, one could generate versions of the troubling argument for incompatibilism. Thus, a Beta-blocking strategy cannot easily assuage the panic that might issue from the incompatibilist's argument.

There is another approach which claims that Beta is not necessary in order to generate the incompatibilist's argument. Bernard Berofsky has recently argued that one can develop the argument without the use of Beta.²⁰

Berofsky presents what he calls a 'system of contingent necessity'. This sort of system validates the following kind of principle, with certain restrictions:

P
N($P \supset Q$)
hence, N(Q)

Whereas it is often alleged that this sort of move involves a modal fallacy, Berofsky attempts to justify this inference (with suitable restrictions on the substitution-instances of the propositional variables), and he claims that it provides a way of formulating the incompatibilist's argument in a valid fashion. We share with Berofsky the claim that the incompatibilist's argument does not require Beta. But if we are correct, then the incompatibilist's argument does not even require the validity of Berofsky's principle and his system of contingent necessity.²¹ It is useful to see that the incompatibilist's argument does not require *any* modal principle similar to Beta.

Clearly, the above considerations—that the incompatibilist's argument can be formulated in various ways without the use of Beta—do not in any way bear on the validity of Beta.²² We, in fact, are of the opinion that Beta might well be one of those intractable principles which seems valid but which can neither be easily proved nor disproved. Recognizing this, the restrictive incompatibilist might contend that, independently of its decisive role in many arguments for incompatibilism, Beta should be accepted by all incompatibilists simply because Beta is valid. And these grounds alone would be sufficient to confirm the restrictivist's position.

For the sake of argument, let us entertain this claim. Let us consider that Beta may well be valid, or at the very least that incompatibilists of the van Inwagen sort are committed to its validity. Does it now follow that such incompatibilists must also be restrictive incompatibilists? To evaluate this question, we turn to van Inwagen's three arguments that purportedly establish that if Beta is valid, then rarely, if ever, is one free to do otherwise.

III. Free to Ignore the Obvious

In order to show that most of the time one is not able to do otherwise, the restrictivist presents a series of three arguments. In the first, he argues that no one is able to act in a manner that he considers morally indefensible. The argument runs as follows:

- (1) N I, (I regard A as indefensible).
- (2) N I, (I regard A as indefensible \supset I am not going to do A).

Hence (via Beta),

- (3) N I, (I am not going to do A) (p. 409).

The intuitive idea behind the argument is that at this moment I don't have any choice about the fact that I now consider some action A indefensible, and I also don't have any choice about its being the case that if I regard an action as being morally indefensible then I am not going to do it; these two premises being true, it follows that at this moment I'm not going to do A and I don't have any choice about this. In short, it is power necessary for me that I am not going to do A. Generalizing the results of this argument the restrictivist concludes that "no one is able to perform an act he considers morally reprehensible" (p.405).

Van Inwagen (our model restrictive incompatibilist) then extends the type of reasoning used in this argument about morally indefensible actions to two other cases which, he claims, constitute the majority of all actions: (1) cases of unopposed inclination in which we want very much to do one thing and have no opposing desires; and (2) cases of unreflective action in which we know what the obvious thing to do is after little if any deliberation. In the case of unopposed inclination, we are asked to consider an example in which a person, Nightingale, is anxiously awaiting a phone call which he very much desires to receive. Nightingale has a very strong desire to answer the phone, and no countervailing desires not to do so. The question is: Can Nightingale refrain from answering the phone? The restrictivist reasons that he cannot, and in support of this conclusion he offers the same argument-form used above. Skipping the formalization, the rough idea behind the argument is as follows: (1) At this moment Nightingale does not have any choice about the fact that he very much desires to answer the phone, and (2) he also has no choice about its being the case that if he very much desires to answer the phone (and he has no countervailing desire to refrain from doing so), then he is going to answer the phone; these two premises being true, it follows that at this moment Nightingale is going to answer the phone and he doesn't have any choice about this. Van Inwagen concludes that "no one is able to do anything that he wants very much *not* to do and has no countervailing desire to do it" (p. 406).

In the last argument, which is supposed to cover the broadest range of actions, the restrictivist turns to actions which "with little or no deliberation...just seem—or would seem if we reflected on them at all—to be the obvious thing to do in the circumstances" (p. 412). Again we are asked to consider a situation in which a phone rings and a person immediately answers it without giving the matter a second thought. Following the same style of reasoning as in the Nightingale example, the argument claims that the agent is not free to refrain from answering the phone. Roughly the argument runs as before: (1) At the moment the phone rings, the person has no choice about the fact that he has no reason not to answer the phone immediately or to deliberate about answering it; (2) furthermore, he has no choice about its being the case that if he hasn't any reason not to answer the phone then

he is going to answer it. From these two premises it follows that at the moment the phone rings, the agent is going to answer it and he has no choice about this. Generalizing this conclusion and that of the preceding argument van Inwagen concludes:

There are therefore, few occasions in life on which—at least after a little reflection and perhaps some investigation into the facts—it isn't absolutely clear what to do. And if the above arguments are correct, then an incompatibilist should believe that on such occasions the agent cannot do anything other than the thing that seems to him to be clearly the only sensible thing to do" (p. 415)

Does an incompatibilist have to accept this conclusion? We think not. To challenge these arguments, we want to take issue with the second premise in each. The most detailed defense of premise (2) is offered in the first argument, here van Inwagen maintains that the second premise is true because the following conditional is a necessary truth and no one has a choice about a necessary truth.

(C1) If X regards A as an indefensible act, given the totality of relevant information available to him, and if he has no way of getting further relevant information, and if he lacks any positive desire to do A, and if he sees no objection to *not* doing A (again, given the totality of relevant information available to him), then X is not going to do A (p. 407).²³

Van Inwagen claims that the restrictivist's three arguments are similar, and thus we assume that van Inwagen imagines that there are conditionals parallel to (C1) which are supposed to support the parallel premisses of the latter two arguments. Here, we will begin by discussing the latter two arguments—pertaining to unopposed inclination and unreflective action. We will deny the claim that the relevant conditionals successfully support the second premisses of these arguments. We shall focus our remarks on the argument concerning unopposed inclination; this argument appears to us to be the stronger of the latter two restrictivist arguments, and the considerations adduced against it can readily be applied to the third argument. Then we will turn to van Inwagen's first argument—concerning indefensible actions. Although we are departing from van Inwagen's order of presentation, our criticism can be developed more naturally in this fashion.

In his second argument, the restrictivist argues that in cases of unopposed inclination the agent cannot do other than what he actually does (despite the intuitive impression that he can so act). The argument has the same form as the argument concerning indefensible actions sketched above, but now the second premise (upon which we shall concentrate) is:

(2) $\neg X, (X \text{ has an unopposed inclination to do } A \supset X \text{ is going to do } A).$

And parallel to the conditional which allegedly supports the second premise of the argument about indefensibility, we have:

- (C2) If X very much desires to do some act A given the totality of relevant information available to him, and if he has no way of getting further relevant information, and if he lacks any positive desire to perform any act other than A, and if he sees no objection to doing A and refraining from doing anything else (again, given the totality of relevant information available to him), then the person is not going to do anything other than A.

Now, the only way in which (C2) can support premise two of the argument is if (C2) is *power necessary* for the relevant agent. That is, (C2) must be true and X must have no choice about whether C2 is true. (This is parallel to the point made above that it is in virtue of the fact that no one has any choice about the truth of (C1) that premise (2) of the first argument is supported.)

The problem with the argument can be made clear by employing the following rather familiar sort of strategy. (C2) admits of two interpretations. On one interpretation, (C2) is plausibly thought to be true and power necessary, but it does not support the second premise of the argument. And on the other interpretation the second premise is supported but (C2) is not plausible. Thus, there is *no* interpretation according to which it is the case that both (C2) is plausibly thought to be power necessary and the relevant premise of the argument is true.

Let us first consider the interpretation according to which (C2) is plausibly taken to be true and power necessary. This interpretation is motivated by the basic idea that action requires some sort of “pro-attitude”—say, a desire. That is, it might be argued that actions are distinguished from mere events in virtue of being preceded (in a suitable way) by special sorts of events: “volitions.” Further, it might be claimed that a volition must be based (in a suitable way) on at least *some* desire. If these claims were true, it would follow that it would be impossible for an agent to perform an *action* without having some desire to do so. We suppose that the necessity of desire for action could be posited even by a theorist who does not believe in volitions. In any case, it is a plausible conceptual claim that it is impossible for an agent to perform an action without having some desire to perform the action in question.²⁴

The key point is that the alleged conceptual truth cannot support premise two of the argument. Note that the alleged conceptual truth can be regimented as follows:

- (C2*) It is not possible that the following state of affairs obtain: that X performs an act other than A without having any desire to perform such an act.

And note further that (C2*) does *not* imply

(2) N, X (X has an unopposed desire to do A \supset X is going to do A).

As long as there is no *obstacle* to the agent's having the desire to do other than A during the relevant temporal interval, we believe that (2) can be false compatibly with the truth of (C2*). (2) would be false if, despite the fact that X has an unopposed desire to do A, he *could* refrain from doing A; and, given that (during the relevant temporal period) X *can* acquire this sort of desire, we believe that it is reasonable to suppose that X can do other than A. (We will argue for this below.)

That (C2*) fails to imply (2) can be seen by considering this simple analogy. It is uncontroversially true that it is not possible that the following state of affairs obtain at all points in some temporal interval: Jones is sitting and Jones is standing up. But this conceptual truth does *not* imply that, if Jones is sitting at some point in some temporal interval, then Jones cannot stand at some point in that interval. Thus, even if (C2*) were true—and it does seem plausible to us—it would not successfully secure the truth of the second premise of van Inwagen's argument.

Now let us interpret (C2) such that it does entail (2):

(C2**) If X does not desire to do other than A, X cannot do other than A.

We concede that (C2**) supports (2), but at the price of plausibility. This is because, even if an agent does not actually desire to do other than A, he might well have the ability (during the relevant temporal interval) to generate such a desire, and to act on this desire. And it is extremely implausible to suppose that agents quite generally lack the *power* to generate the relevant sorts of desires.

We elaborate. Just about anybody can summon up the worry that he is not free to do otherwise. That is, one can worry that, despite the pervasive intuitive feeling that frequently we have genuine freedom to do various things, we do not in fact have such freedom. (Indeed, anyone who thinks about the restrictivist's argument certainly has reason to worry that he might not be free to do otherwise in many contexts.) This worry can then generate *some* reason (perhaps, a desire) to do otherwise simply to prove that one can do so.²⁵ Thus, barring special circumstances—to which the restrictivist does not allude in his arguments—even an agent who actually does not have any desire to do other than A can have the power to generate such a desire (during the relevant temporal interval). And insofar as: (i) the agent *can* generate the desire to do other than A, (ii) the agent can try to act on this desire, and (iii) if he were to try to act on this desire, he would succeed, then we believe that the agent *can* (during the relevant temporal interval) do other than A.²⁶ The leading idea here is that there is no reason to suppose that agents *generally* lack the power to generate (in some way or another) reasons to

do otherwise, the power to try to act on those reasons, or the power to succeed in so acting.

Consider van Inwagen's own example in which Nightingale wants very much to answer the phone as soon as it rings. If Nightingale can call to mind the doubt that he is able to do otherwise in such situations, this very doubt can give him a reason to pause before picking up the receiver. (Perhaps he simply does not answer the phone on the first ring, but waits until it rings five times; this suffices, he might feel, to prove he was free to do otherwise.) In this scenario, Nightingale's worry has transformed a normally routine phone call into a situation in which Nightingale must decide between two conflicting desires: (1) a desire to answer the phone as soon as it rings, and (2) a desire to prove to himself that he doesn't have to answer it as soon as it rings. We claim that insofar as: (i) the agent can generate a desire of the second sort, (ii) he can try to act on this desire, and (iii) if the agent were to try to act on this sort of desire, he would succeed in doing other than A, then the agent *can* (during the relevant temporal period) do other than A, even though he actually lacks any desire to do other than A.²⁷

We believe that the above considerations show that, even if an agent actually lacks any desire to perform a given act, he *can* perform that action, insofar as certain conditions are met. These conditions involve the ability to generate certain reasons and to translate these reasons into action. Further, we suggested that it is extremely plausible to suppose that (absent special assumptions about causal determinism or particular psychological or physical impairments) these conditions are frequently met.²⁸ Thus, we believe that (C2**) is not in general true. We have argued, then, that whereas (C2*) is plausible, it does not imply (2); and whereas (C2**) implies (2), it is not plausible.

In order more clearly to highlight our position, it is useful to consider the complaint that we have simply missed van Inwagen's point.²⁹ Van Inwagen's claim is that if in some possible world, W1, Nightingale has a strong, unopposed inclination to answer the phone as soon as it rings, then, Nightingale is going to answer the phone as soon as it rings and he is not able in W1 to do otherwise. But—the objector continues—all your reconstruction of the example shows is that if in some other possible world, W2, Nightingale's motivational set is changed so that he has two conflicting inclinations, then Nightingale in W2 is able to refrain from answering the phone as soon as it rings. Nightingale's ability in W2, however, is a function of his having opposing inclinations, and in itself this doesn't show that Nightingale in W1, without the opposing inclinations, is able to do otherwise. The issue, then, is not what Nightingale can do in W2 with a different motivational set, but rather what Nightingale can do in W1 given that his motivational set is just as van Inwagen stipulates.

We reply that, as long as Nightingale is genuinely *able* (during the relevant

temporal interval) in W1 to generate a desire to answer the phone, then he is *able* in W1 to answer the phone. Insofar as W2 is *genuinely accessible* to Nightingale, then W2 is relevant to what Nightingale *can* do in W1. It is only if W2 is not so accessible that it is irrelevant to Nightingale's abilities in W1. Of course, we rely here on the fact, if some world W2 is in the appropriate sense accessible to W1, then W2 may be relevant to the modal properties of individuals in W1.

IV. Free To Act Indefensibly, Free To Act Crazyly

Thus far we have argued against the restrictivist's argument that in cases of unopposed inclination the agent is not free to do otherwise. We believe that the same considerations apply, *mutatis mutandis*, to the argument concerning unreflective actions. Thus, we believe that we have pointed to a way of salvaging the intuition that, even if Beta were true, individuals are often free to do otherwise in contexts of unopposed desire and unreflective action. Now let us turn to van Inwagen's parallel argument concerning indefensible actions.

Having developed the criticism of the argument pertaining to unopposed desire, it is now extremely simple to explain what is wrong with the argument concerning indefensible actions. In fact, our objection to the argument concerning indefensible actions is precisely the same as the objection to the argument concerning unopposed desires.

Recall that van Inwagen adduces (C1) in support of premise (2) of the argument:

- (C1) If X regards A as an indefensible act, given the totality of relevant information available to him, and if he has no way of getting further relevant information, and if he lacks any positive desire to do A, and if he sees no objection to *not* doing A (again, given the totality of relevant information available to him), then X is not going to do A (p. 407).

Given that van Inwagen adduces (C1) in support of (2), it is clear that he is interpreting the second premise of the argument in the following way:

- (2) N, X (X regards A as an indefensible act and X lacks any desire to do A \supset X does not do A).

To proceed as above. (C1) can be interpreted so as to claim that the following state of affairs is not possible: that X regards A as indefensible, has no desire to perform A, and performs A. But (C1), so interpreted, does not imply (2). Alternatively, (C1) could be interpreted so as to claim that if X regards A as indefensible and X lacks any desire to do A, then X cannot do A. But, so interpreted, (C1) is false, insofar as X can (in the relevant temporal

interval) generate the desire in question.

But in the context of unopposed desire discussed above, it is not supposed that the agent believes that the act in question is *indefensible*. Might this belief constitute an obstacle to generating a reason (or desire of the sort discussed above) to perform the act? That is, is the context of indefensible acts relevantly different from the contexts of unopposed desire and unreflective action?

The examples adduced by van Inwagen in support of (C1) suggest that morally indefensible actions *do* have some special status such that one literally is unable to bring oneself to desire to do (and to do) them.³⁰ To make this point, van Inwagen begins with an example presented by Daniel Dennett in which Dennett makes the claim that he is unable to torture innocent victims for small sums of money.³¹ Van Inwagen observes that the point of the example is not so much that Dennett would not be able to torture these innocents if he so chose, but rather that, given Dennett's character, he simply is *unable to make such a choice* (and, presumably, unable to generate the relevant desire). Van Inwagen wishes to extend this line of reasoning to show that he also could not slander a colleague to prevent that colleague's appointment to Chairman of the Tenure Committee, and similarly that none of us could do anything that he considers indefensible.

Now, we certainly grant that there may be *some* actions—call them “unwilling” actions—which a particular agent literally cannot bring himself to choose to do (and to do); and some (although not necessarily all) of these unwilling actions may be ones that are morally indefensible.³² Indeed, Dennett's example of torturing innocents seems to be just such a case. We wish to emphasize, however, that it does not follow from an action's being morally indefensible that it is *unwilling*. That is, we suggest that the Dennett/van Inwagen point here gains plausibility from their focusing on a proper subset of the relevant cases: those morally indefensible actions which are *also* unwilling. But an indefensible action is not *eo ipso* unwilling. Thus, we wish to block the move from the specific case of one's not being able to torture innocents to the general claim that “no one is able to perform an act he considers morally reprehensible” (p. 405).

We believe that there *can* be cases in which an agent believes that an act is morally indefensible and nevertheless has a desire to perform it (of the sort mentioned above) and indeed successfully acts on this desire. And it is in general plausible to suppose that agents have the *power* to generate this sort of desire. In order to support our claim that the context of indefensible action is not relevantly different from the other two contexts, we present the following examples in which an individual believes that the act in question is indefensible but nevertheless has a desire to perform it and does indeed perform it.

Consider first Augustine's famous account of the theft of pears in his boyhood. Shortly before this passage, Augustine is wondering about the reason

for his stealing pears for which he had no desire, and after acknowledging the view that all action must be for the sake of some apparent good, he dismisses this explanation in his own case:

.now that I ask what pleasure I had in that theft, I find that it had no beauty to attract me. It did not even have the shadowy, deceptive beauty which makes vice attractive" (*Confessions* II, vi, Pine-Coffin translation).

Let my heart now tell you what it sought when I was thus evil for no object, having no cause for wrongdoing save my wrongness. The malice of the act was base and I loved it—that is to say I loved my own undoing, I loved the evil in me—not the thing for which I did the evil, simply the evil (*Confessions* II, iv, Sheed translation).

Augustine's reflections are disturbing precisely because they exemplify one man's ability not only to do something he takes to be morally indefensible, but to be drawn to the action precisely because it is so indefensible. This is not to say that Augustine did not see the robbery as having some desirable consequences. He himself admits that he would not have committed the crime had it not been for his companions and the "thrill of having partners in sin" (*Confessions* II, viii). However, simply because Augustine wanted something from his thieving, this does not show that he saw the thieving as good, or that he believed it conformed to an overall system of values he was willing to defend. A person might see the pilfering of pears as wholly indefensible and still desire to do it, if for no other reason than to assert one's ability to act against moral value. Indeed Augustine's comments suggest that he saw his attraction to evil as being intimately connected to this desire for a perverse sort of freedom and power—a freedom to ignore the Good:

What was it, then, that pleased me in that act of theft? Which of my Lord's powers did I imitate in a perverse and wicked way? Since I had no real power to break his law, was it that I enjoyed at least the pretence of doing so, like a prisoner who creates for himself the illusion of liberty by doing something wrong, when he has no fear of punishment, under a feeble hallucination of power? Here was a slave who ran away from his master and chased a shadow instead! What an abomination! What a parody of life! What abysmal death! Could I enjoy doing wrong for no other reason than that it was wrong?

I loved nothing in it except the thieving, though I cannot truly speak of that as a 'thing' that I could love, and I was only the more miserable because of it (*Confessions* II, vi-viii, Pine-Coffin Translation)

A different type of rebellion, but one which expresses a related yearning to flout moral prohibitions, is found in the story of a character quite distinct from St. Augustine: Dostoevsky's Raskalnikov. Recall that at the outset of the story, Raskalnikov is contemplating killing and robbing the old pawnbroker, Alena Ivanovna, and as he does so, he is keenly aware of the evil at hand; he knows such acts are morally reprehensible and he is repulsed by his own musings:

“Oh God, how repulsive! Can I possibly, can I possibly.. no, that’s nonsense, it’s ridiculous!” he broke off decisively “How could such a horrible idea [i.e., to rob and murder Ivanovna] enter my mind? What vileness my heart seems capable of! The point is, that it is vile, filthy, horrible, horrible!” (*Crime and Punishment*, I 1)

In spite of this moral aversion, Raskalnikov nonetheless finds that he is able to do the indefensible: he takes a borrowed axe to the head of not only Alena Ivanovna but her sister as well. Later, as he thinks back on the murder and robbery, Raskalnikov dismisses the only reasonable motive for the crime: “If it all has been done deliberately and not idiotically,” he ponders, “if I really had a certain and definite object, how is it I did not even glance into the purse and don’t know what I had there, for which I have undergone these agonies and have deliberately undertaken this base, filthy, degrading business?” (*Crime and Punishment*, II.2). Raskalnikov knows that he did not kill the old woman, as a more typical criminal might have, for her money. And later, as he confesses to Sonya, the deeper motivation behind the crime comes out:

“I realized then, Sonya,” he went on enthusiastically, “that power is given only to the man who dares stoop and take it There is only one thing needed, only one—to dare .I wanted to *have the courage*, and I killed I only wanted to dare, Sonya, that was the only reason!”
 . “what I needed to find out then, and find out as soon as possible, was whether I was a louse like everybody else or a man, whether I was capable of stepping over the barriers or not Dared I stoop and take the power or not?”
 “Listen when I went to the old woman’s that time, it was only to *test myself* Understand that!” (*Crime and Punishment*, V 5)

Raskalnikov’s remarks are of interest to us because they give an example of a man who (1) knows that robbery and murder are morally indefensible, (2) is not driven to perform these acts in the pursuit of some good which can be separated from the crime itself, and (3) nonetheless does rob and murder two people. Indeed, what is most important about Raskalnikov for our purposes, is that, given a straightforward reading, he seems drawn to murder the aging pawnbroker, precisely to see if he *can* do it: He wants to discover if he has the power to ignore moral prohibitions; he wants to know if he is free to do the morally indefensible.³³

What is striking about the crimes of both Augustine and Raskalnikov is that, unlike a more mundane robbery in which the wrongdoing is merely a means to material gain, the motive behind their crimes is inextricably bound up with a desire to do wrong and to flout moral constraints. This is not to say that the motivations of Augustine and Raskalnikov can be assimilated in every respect. Whereas Augustine seeks the freedom to do evil in order to rebel against the good, Raskalnikov seeks this freedom to show that he is beyond good and evil. But the crucial point for our discussion is that both

men claim to do what the restrictive incompatibilist says they cannot—freely perform an act that is perceived by the agent to be morally indefensible.

We have argued, then, that the context of indefensibility is not relevantly different from the context of (say) unopposed desire: an agent can generate a certain sort of desire to perform an action even though he believes that the action is morally indefensible. Thus, our critique of van Inwagen's argument about contexts of unopposed desire (and unreflective action) can be extended to apply to his argument about contexts of indefensible actions. Someone might object that our examples only pertain to contexts in which agents believe that the relevant actions are *morally* indefensible, rather than indefensible from some broader (perhaps "all-things-considered") perspective.³⁴ But it is clear that van Inwagen has in mind the notion of moral indefensibility. Further, if the broader notion of indefensibility were employed, this would substantially reduce the incidence of contexts of indefensibility (thus vitiating the restrictivist's claim that we are rarely free to do otherwise). Finally, we do not see why individuals cannot generate desires (perhaps they would be "weak-willed desires") to do things which they consider to be indefensible, all things considered.

Before leaving the question of whether or not we are free to act indefensibly, we want to consider a final worry about such freedom which is suggested by two interesting examples recently formulated by Susan Wolf. Wolf asks us to consider what it would mean for an agent to have the ability to act against everything he believes in and cares about:

It would mean, for example, that if the agent's son were inside a burning building, the agent could just stand there and watch the house go up in flames. Or that the agent, though he thinks his neighbor a fine and agreeable fellow, could just get up one day, ring the doorbell, and punch him in the nose. One might think that such pieces of behavior should not be classified as actions at all—that they are rather more like spasms that the agent cannot control. If they are actions they are very bizarre, and an agent who performed them would have to be insane. *Indeed, one might think he would have to be insane if he had even the ability to perform them.* For the rationality of an agent who could perform such irrational actions as these must hang by a dangerously thin thread.³⁵

Before directly discussing these examples, a word of qualification is in order. Wolf originally presents these examples to illustrate what it would mean for an agent's actions not to be determined by any interests whatsoever. One of the points she is making, if we have understood her properly, is that a person whose actions weren't determined by *any* interests could hardly be said to be acting at all. Rather his behavior, since it did not reflect any interests or intention, would seem more like spasms or the bizarre movements of an insane person. Understood in this fashion Wolf's claim is certainly unobjectionable; indeed, this insight seems merely to reflect the (alleged) conceptual truth discussed above that all behavior, if it is to be considered action at all,

must reflect some pro-attitude.

Our interest in Wolf's examples comes from another more substantive claim which is also suggested by her examples and subsequent comments; this is the suggestion that anyone who even had the *ability* to perform indefensible acts (like allowing her children to burn, or punching her neighbor in the nose for no good reason) would have to be insane. This claim is not the trivial one that anyone whose bodily movements did not reflect her interests would be insane; rather it is the more interesting and substantive claim that anyone who even had the ability to act against all seemingly good interests would be insane. A similar sentiment is found in the following passage by Daniel Dennett: "But in other cases, like Luther's, when I say I cannot do otherwise I mean I cannot because I see so clearly what the situation is and because my rational control faculty is *not* impaired. It is too obvious what to do; reason dictates it; I would have to be mad to do otherwise, and since I happen not to be mad, I cannot do otherwise."³⁶ Both Wolf and Dennett seem inclined to slide from the claim that 'doing X would be crazy' to a stronger claim that 'anyone who had the ability to do X would be crazy'. If this "Wolf/Dennett slide" were correct, then, since most of us are not crazy, it would seem to follow that most of us are not able to act in a crazy, indefensible manner—a conclusion which the restrictive incompatibilist would of course welcome.

We think the conclusion reached via the Wolf/Dennett slide is false. In fact, a strategy similar to the one used earlier to expose the fallacy in van Inwagen's arguments also can be deployed here to make clear the problem with this slide. The conclusion that sane people are not free to do insane things is supposed to follow from the claim that it is not possible for someone to do something that is crazy without actually being crazy. But there are two ways to interpret this claim. On one interpretation, the claim is true, but it fails to support the desired conclusion; on the other interpretation, the conclusion does follow, but the claim is false.

On the first interpretation, the initial claim is construed to mean that the following state of affairs is not possible: that an agent be sane and perform a crazy action. So interpreted the claim may be true, but it certainly does not imply the conclusion that no sane person has the *ability* to act crazily. In order to reach this conclusion, the initial claim needs to be strengthened so as to claim that if an agent is sane then it is not possible for her to do crazy things. But so interpreted the claim seems false. After all, what reason is there to think that the mere *ability* to act crazily should call one's rationality into question?

With respect to other vices, it is customary to accept a distinction between having an ability and exercising it. For example, having the ability to eat and drink to excess does not imply that one is intemperate; nor does having the ability to flee from the battlefield, a coward make. Indeed this distinction seems applicable to a wide range of character traits—having the ability to

act generously does not make one generous, having the ability to act dishonestly does not make one a liar, and so forth. The point here is simply that having the power to act in a certain way does not entail that someone is the type of person who will act that way. And given this general fact, why should we expect the case to be otherwise with indefensible actions like punching one's neighbor for no good reason?³⁷ Why should simply having the ability to act crazily render one crazy? Why should there be this asymmetry between the "ability to act crazily" and other dispositional notions?

An example might be helpful here. A traditional view has it that if we have a free will at all, we must have a perfect, God-like free will. Roughly the idea behind the view is that whereas there can be impediments to action—i.e., one can be unable to act in accordance with one's will—there cannot be any impediments to willing.³⁸ We raise this view not to defend it, but rather to assume, for the purposes of this example, that it is true. (If one prefers science fiction and fantasy to tradition, then simply imagine that you happen upon a magical ring, and after placing it on your finger, you discover that it has bestowed upon you the infinitely free will described by the traditional view above: a will that enables you to choose or not choose any option you desire irrespective of your morals or best interests). Now one thing should be clear: simply because the range of your choices has been increased (thanks to the ring), your ability to listen to reason has not been decreased. Having this freedom does not somehow mute the voice of conscience, or leave you with no way to know which course of action is the most rational; it merely gives you the ability to pick a less optimal path if you so will. Like the motorist who reaches a junction from which she can take either a scenic parkway heading directly toward her destination, or a one-lane dirt road that crawls through acres of sanitary landfill in the wrong direction, you more than likely will pick the most reasonable alternative. But surely we won't judge the motorist to be crazy simply because she is at a junction where she can choose a route which is not in her interests, and neither should we judge you crazy simply because you have the power to choose against your interests.

"Still"—one might complain—"being at a crossroads scarcely shows that one is free to turn as she pleases. After all, no sane motorist ever *will* take the dirt path, and similarly no sane person ever *will* knowingly act against her interests. Indeed having the freedom to act this way would appear to be less a blessing than a curse; for why would anyone ever want the ability to behave in such a contrary fashion? It short it would seem that the power to act both irrationally and immorally, if we have it at all, is hardly as much of an ability, as it is a *disability*—a character flaw which needs to be overcome."³⁹

Two points are raised by this worry. One is easily dealt with; the other broaches a broader issue which we can only touch upon in the context of this discussion. As to the first point—that a sane motorist will never actually

choose the dirt road, and a sane person will never actually act against her best interests—we can agree that in most cases this is true.⁴⁰ Nevertheless, as we saw above, the fact that someone never *will* act against her interests does not entail that she *cannot* do so. For surely there is nothing incoherent about a person having a power which she never exercises. Having given this response, however, we are lead straightaway to the second, more complicated worry: why would a sane person ever want to have a power that she will never exercise, especially a power to act against all of her morals and best interests?⁴¹

But, lamentably, to ask whether we would *want* to have something is, of course, not the same as asking whether we *have* it, for it might turn out that we have the freedom to act indefensibly even though this is hardly a freedom we *would like* to have. Hence, this worry cannot aid the restrictive incompatibilist in securing his position.⁴²

The Wolf/Dennett slide *is* a slide, and it is not well-motivated; the fact that doing X would be crazy does not (in itself and without further argumentation) imply that anyone who had the *ability* to do X would be crazy. Just as agents with the *power* to be gluttonous need not *be* gluttons, agents with the *power* to act crazily need not be considered crazy.

V. Restrictive Incompatibilism and Moral Responsibility: Tracing

Thus far we have argued that incompatibilists need not accept the restrictive incompatibilist's claim that "rarely, if ever, is anyone able to do otherwise than he in fact does." But what if our criticisms have not persuaded? What if incompatibilists still believe that they are conceptually committed to the thesis of restrictive incompatibilism? In closing we want to entertain this possibility and in particular to direct our attention to the following question: how would an incompatibilist account of moral responsibility be affected if one were convinced that most of the time we are not able to do otherwise? Answering this question will show that, even if incompatibilists did not find our previous objections compelling, they should wish they had.

As we mentioned at the outset of this paper, the restrictive incompatibilist identifies himself with what van Inwagen terms the 'classical tradition.' This tradition holds that there is an intimate connection between free will and moral responsibility, such that if there were no free will—if nobody were ever able to do otherwise—then there would be no moral responsibility. This requirement does not, of course, mean that there aren't particular instances in which a person might still be held accountable even though at the time of the action he was unable to do otherwise. (Van Inwagen's example of the drunk driver is such a case). However, it does suggest that any state of affairs for which we are responsible must be able to be traced back to some prior

free action. To capture this “tracing” principle van Inwagen offers the following rule:

An agent cannot be blamed for a state of affairs unless there was a time at which he could so have arranged matters that that state of affairs not obtain (p. 419)

This type of principle does not bode well for any incompatibilist who feels compelled to accept the restrictivist’s conclusions, but still hopes to remain within the classical tradition. Remember the restrictive incompatibilist must hold that there are only three situations in which we are able to do otherwise: Buridan cases, cases in which duty conflicts with inclination, and situations of conflict between incommensurable values. Conjoin this premise with the above tracing principle, and now the restrictive incompatibilist is committed to showing that all states of affairs for which we are responsible can be traced back to one of these three kinds of situations. But why should we think that everything for which we are responsible can be traced back to some free choice between equally attractive alternatives, duty and inclinations, or incommensurable values?

The most promising strategy for the incompatibilist to adopt at this point is to argue that these kinds of conflict situations are precisely the ones through which our characters are formed; hence, we can accept his theory and still be responsible for all states of affairs which come about as a result of actions that are produced by our characters. In the end, however, even 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.

Consider a young woman, call her Betty, who has spent all of her life in a small, rural community. Like most of the citizens of her town, Betty’s family is still proud to be American, and over the years Betty has gradually, almost imperceptibly, internalized a certain degree of patriotism. Being raised mostly during the apathy of the Reagan years, Betty has never been in any situation where her mild patriotism has come into conflict with any of her short-term inclinations or other values. Indeed she has never given the matter much thought—for Betty, being a loyal American has come as naturally as flying the flag on Independence Day. Even though this mild patriotism is a fixed feature of Betty’s character, the restrictivist must hold that she is not yet responsible for it; he is committed to this view because Betty has not yet been in a conflict situation in which she was able to make a free choice that would have prevented her from having her patriotic disposition. Imagine now that Betty travels abroad for the first time, and through a series of strange coincidences, a singularly incompetent foreign agent mistakes her for a young American soldier who has expressed an interest in selling government secrets. He approaches Betty and asks her, in so many words, to betray her country.

Of course, Betty thinks that treason is morally indefensible; she has a strong desire not to do it, and with scarcely a moment's deliberation she turns down the agent's offer without waiting for any further explanation. For the restrictive incompatibilist, Betty clearly was not able to do anything but what she did. Moreover given that her action resulted from features of her character which in turn could not be traced back to some earlier free decision, it seems that he should say that Betty is not responsible for the ensuing state of affairs that Betty declined to betray her country. But such a conclusion runs directly counter to our actual practices of holding people responsible. Indeed if Betty is not responsible in this case, then it would appear that the restrictivist's position requires that he severely limit the domain of moral responsibility, for a great many of our everyday actions result from other character traits and dispositions which, like Betty's patriotism, are not able to be traced back to one of these situations of conflict between duty and inclination or between incommensurable values.

Of course, the restrictive incompatibilist 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. Thus, we leave the restrictive incompatibilist with a dilemma: either accept a severe restriction on the range of states of affairs for which we can be held morally accountable, or else reject the claim that most of the time we are unable to do otherwise. Van Inwagen claims that restrictive incompatibilism can be embedded within a traditional approach to moral responsibility via a tracing theory; we have argued this claim is false.

VI. Conclusion

We have not in this paper intended to argue for incompatibilism, nor have we attempted to explain how the will can be free in an indeterministic world. Rather, we have tried to make several more minimal points. First we sought to undermine the restrictive incompatibilist's position by challenging van Inwagen's initial claim that "if one accepts Beta, one should be an incompatibilist, and if one is an incompatibilist, one should accept Beta" (p. 405). In particular we presented an argument to show that accepting Beta is not a necessary condition of incompatibilism. (In passing we also cited several compatibilist strategies which allege that accepting Beta is not a sufficient condition of incompatibilism.) The argument that one can be an incompatibilist without having to accept Beta has ramifications beyond the scope of our discussion of restrictive incompatibilism. If incompatibilism can

be secured without explicitly using Beta, then a recent trend—exemplified by Slote—which sees a deep family resemblance among various formulations of the argument for incompatibilism and which hopes to undermine them all by questioning the validity of Beta must be seen to have a more limited scope than its proponents might previously have hoped.

Second, we argued that even if an incompatibilist does accept Beta, he need not accept the restrictivist's thesis that one rarely, if ever, is free to do otherwise. Specifically we challenged the restrictivist's claim that persons are unable to (1) perform actions they consider morally indefensible, (2) refrain from performing actions which they strongly desire to perform, and (3) refrain from performing actions which they take to be the only sensible thing to do. In connection with these arguments we questioned a slide (suggested in the writings of Wolf and Dennett) which sought to move from the claim that 'doing X would be crazy' to the stronger conclusion that 'anyone who can do X must be crazy'.

Finally we examined the consequences that restrictive incompatibilism would have for a traditional theory of moral responsibility. We concluded that if incompatibilists were indeed committed to the thesis that we rarely are able to do otherwise, such a commitment would bode ill for any incompatibilistic theory of responsibility which still hoped to remain within the classical tradition.^{43,44}

Notes

- 1 Peter van Inwagen, "When Is The Will Free?" in *Philosophical Perspectives*, 3, *Philosophy Of Mind And Action Theory*, ed. James Tomberlin (Atascadero, California: Ridgeview Publishing Co, 1989), pp 399-422 All subsequent page references will be to this article unless otherwise noted
- 2 We will follow van Inwagen's usage and treat 'free will' as "a philosophical term of art" According to van Inwagen to say of someone that she "has free will" means roughly that she sometimes is free to do other than what she, in fact, does
3. In "When Is The Will Free?" van Inwagen primarily argues for the second half of this thesis; he doesn't purport to offer a detailed defense of the argument that free will is incompatible with causal determinism. He does, however, provide such a defense in his excellent book *An Essay On Free Will* (Oxford: Clarendon Press, 1983). Hence we believe that when taken together "When Is The Will Free?" and *An Essay On Free Will* can reasonably be construed as a defense of the position we are calling 'restrictivism'
- 4 The following formulation of the Consequence Argument is quoted from "When Is The Will Free?" p 405 See also *An Essay On Free Will*, pp 55-105, and Carl Ginet, "In Defense of Incompatibilism," *Philosophical Studies* 44 (November 1983), pp 391-400
- 5 Van Inwagen uses this term broadly to include both standard Buridan cases in which "one wants each of two or more incompatible things and it isn't clear which one he should (try to) get, and the things are interchangeable" (p 415), and cases which he calls "vanilla/chocolate cases" These are situations in which "the

alternatives are not really interchangeable (as two identical and equally accessible piles of hay) but in which the properties of the alternatives that constitute the whole difference between them are precisely the objects of the conflicting desires" (p. 415)

6. Because the central thrust of our criticisms lies elsewhere, we will not directly take issue with this claim. However, it is worth noting that van Inwagen's contention that these sorts of cases (i.e. cases in which we are free) occur only *rarely* is highly debatable. Indeed it seems more plausible to suppose that these cases occur as often, if not more often, than the three types of cases in which van Inwagen claims we are *not* free. We are grateful to Carl Ginet and Nancy Schaubert for calling this point to our attention.
7. Van Inwagen claims that "if one accepts Beta one should be an incompatibilist, and if one is an incompatibilist, one should accept Beta" (p. 405). In what follows we discuss how certain *incompatibilists* would take issue with the latter half of this claim—i.e., "if one is an incompatibilist, one should accept Beta." However, it is worth noting that the former half of the claim—"if one accepts Beta, one should be an incompatibilist"—would also be contested by certain *compatibilists*. Such compatibilists argue that one can accept Beta and still take issue with the basic argument for incompatibilism, they do so either by challenging the fixity of the past (van Inwagen's premise 4) or by challenging the fixity of laws (van Inwagen's premise 6). (For a discussion of former type of compatibilism which might be called "multiple-pasts" compatibilism see Jan Narveson, "Compatibilism Defended," *Philosophical Studies* 32 (July 1977), pp. 83-87, André Gallois, "van Inwagen On Free Will And Determinism," *Philosophical Studies* 32 (July 1977), pp. 99-105, Richard Foley, "Compatibilism And Control Over The Past," *Analysis* 39 (March 1979), pp. 70-74, Keith Lehrer, "Preferences, Conditionals, And Freedom," in *Time And Cause* ed. Peter van Inwagen (Dordrecht. D. Reidel Publishing Co., 1980), John Martin Fischer, "Incompatibilism," *Philosophical Studies* 43 (January 1983), pp. 127-37. For a discussion of the latter type of compatibilism, which might be called "'local-miracle' compatibilism", see David Lewis, "Are We Free To Break The Laws," *Theoria* 47 (1981, Part 3), pp. 113-121, Carl Ginet, "In Defense of Incompatibilism"; John Martin Fischer, "Incompatibilism", and Fischer, "Freedom And Miracles," *Noûs* 22 (June 1988), pp. 235-252, and Kadri Vihvelin, "How Are We (And Are Not) Free To Break The Laws Of Nature" (manuscript).
8. Indeed van Inwagen himself offers two formulations of the argument for incompatibilism that do not depend on any rule of inference like Beta. See van Inwagen's presentation of his First Formal Argument and his Second "Possible Worlds" Argument in *An Essay On Free Will* pp. 55-93. For another example see Carl Ginet, *On Action* (Cambridge: Cambridge University Press, 1990), pp. 90-123.
9. *An Essay On Free Will*, p. 57
10. *Ibid*
11. Michael Slote, "Selective Necessity and the Free-Will Problem," *Journal of Philosophy*, 79 (January 1982), p. 9.
12. See Terence Horgan, "Compatibilism and The Consequence Argument," *Philosophical Studies*, 47 (1985), p. 339.
13. *Pace* van Inwagen, Slote, and Horgan, we want to argue that a "finer-grained" approach to the various arguments for incompatibilism is needed which recognizes that not all formulations make use of the same inference rules or involve the incompatibilist in the same commitments. For example, whereas van Inwagen's

modal argument makes use of principle Beta (*Essay On Free Will*, p. 94), his First Formal Argument uses a different "entailment" principle: "If *s* can render *r* false, and if *q* entails *r*, then *s* can render *q* false" (*Essay On Free Will*, p. 72). Other arguments for incompatibilism rely on still a different type of "transfer" principle "S cannot do X, In the circumstances doing X is doing Y, Therefore, S cannot do Y" (See Fischer's discussion in "Scotism," *Mind*, 94 (April 1985), pp. 231-243) Other philosophers also employ similar principles For discussions of such principles and their roles in incompatibilistic arguments see Philip L. Quinn, "Plantinga On Foreknowledge And Freedom," in *Alvin Plantinga*, eds. James E. Tomberlin, and Peter van Inwagen (Dordrecht: D. Reidel Publishing Co., 1985), Thomas B. Talbott, "Of Divine Foreknowledge And Bringing About The Past," *Philosophy And Phenomenological Research* 46 (March 1986), pp. 455-469, David Widerker, "On An Argument For Incompatibilism," *Analysis* 47 (January 1987), pp. 37-41, and Widerker, "Two Forms Of Fatalism," in *God, Foreknowledge, And Freedom*, ed. John Martin Fischer (Stanford: Stanford University Press, 1989), and Ginet *On Action*. Although all of these Beta-like principles bear some resemblance to one another, it is clear that (on the surface at least) they are not identical. Moreover, as we argue below, the incompatibilist's argument can be formulated in such a way that it makes use of neither Beta, van Inwagen's entailment principle nor any transfer principle. In addition to the argument we shall present, van Inwagen's second argument—the "possible worlds" argument—is an example of an incompatibilist argument which does not depend on any principles of this sort.

- 14 For a detailed presentation and discussion of this argument see Fischer, "Scotism"
- 15 The issues here are complex and delicate—see Lewis, "Are We Free to Break the Laws?", Ginet, "In Defense of Incompatibilism", Fischer, "Incompatibilism" and "Freedom and Miracles", and Vihvelin, "How We Are (And Are Not) Free To Break The Laws"
- 16 In fact, we do not take this argument to be a definitive proof of incompatibilism. For one type of compatibilist response to this argument, see John Martin Fischer, "Power Over The Past," *Pacific Philosophical Quarterly*, 65 (1984), pp. 335-350. The criticisms in this article suggest that a stronger version of the argument for incompatibilism might well be something like the "possible worlds" argument which van Inwagen develops in *An Essay On Free Will*. This argument has the advantage that it relies on neither Beta nor on an overly strong fixity of the past claim which denies one even non-causal power over the past.
17. One might object that although the above argument does not explicitly employ Beta, it would not be sound unless Beta were valid because some principle like Beta is what leads one to accept FL and FP. We, however, do not see how this objection could be developed to show that FP or FL is indeed formally dependent on Beta, nor do we see how the alleged counterexamples to Beta could be successfully translated into criticisms of the above argument. Rather it seems to us that such principles about the fixity of the past and the laws have an independent appeal, and hence one could accept FP and FL without having to accept anything like Beta as a general rule of inference.
- 18 See Daniel C. Dennett, *Elbow Room: The Varieties Of Free Will Worth Wanting* (Cambridge: The MIT Press, 1984), p. 148, and Slote, "Selective Necessity and the Free-Will Problem."
19. John Martin Fischer, "Power Necessity," *Philosophical Topics*, 14 (Fall 1986), pp. 77-91.
20. Bernard Berofsky. *Freedom from Necessity: The Metaphysical Basis of Responsibility* (New York: Routledge and Kegan Paul, 1987).

- 21 Indeed, recognizing that the incompatibilist's argument can be formulated without either Beta or Berofsky's principle calls into question much of the motivation for developing such a system of contingent necessity. The machinery of the system of contingent necessity developed by Berofsky is useful insofar as one wishes to have a modal version of the incompatibilist's argument, but it is important to see that this machinery is not necessary in order to generate the incompatibilist's argument
- 22 For further discussion concerning the validity of Beta see John Martin Fischer, "Introduction Responsibility And Freedom," in *Moral Responsibility*, ed. Fischer (Ithaca Cornell University Press, 1986), pp 9-61, and Fischer "Power Necessity "
- 23 Although van Inwagen calls this conditional, '(C)', it will be useful for our purposes to call it, '(C1)'
24. It is of course not clear that this conceptual claim is true A Kantian theorist of action might argue that actions can be motivated by reason alone and that desire is not a necessary precursor of genuine action. Thus, we do not wish to suggest that the (Humean) conceptual claim is obviously true, rather, we only suggest that it has a certain plausibility Further, it is clear that if the Humean conceptual point is indeed false, then van Inwagen's argument is even in worse shape: in this case even the weaker interpretation would issue in a falsehood and thus no support for (2)
25. In *De Fato* Alexander suggests that when one's freedom is called into question it can be reasonable to do something (that might on other occasions be seen as irrational) simply in order to demonstrate one's ability to do otherwise "Next it is not by compulsion that the wise man does any one of the things which he chooses, but as himself having control also over not doing any one of them For it might also sometimes seem reasonable to the wise man *not* to do on some occasion what would reasonably have been brought about by him—in order to show the freedom of his actions, if some prophet predicted to him that he would of necessity do this very thing" (*De Fato* 200 2-7)
- 26 The notion of successfully acting on a desire is ambiguous between being moved by the desire and actually succeeding in getting the object of one's desire We mean to adopt the latter interpretation.
- 27 Of course, a critic might object that this scenario presupposes that we always do have the ability to call such a worry to mind. However, nothing the restrictivist has said suggests that an incompatibilist must deny that we have *this* ability, and until such an argument is given it seems reasonable to adhere to the common wisdom that we are free to think as we will And one cannot here point out that *if* causal determinism were true—together with incompatibilism—it would follow that we would not have the power in question This is because the restrictivist's argument is supposed to show that Beta implies that we are rarely free to do otherwise, even if determinism were false.
28. Admittedly van Inwagen does want to construe his example in such a way that the incompatibilist must agree that the person is unable to call to mind any reason for not answering the phone To ensure this condition, he writes: "But we might also imagine that there exists no basis either in my psyche or my environment (at the moment the telephone rings) for any of these things [i e., things that would give me a reason not to answer the phone or that would keep me from answering it] We may even, if you like, suppose that at the moment the telephone rings it is causally determined that no reason for not answering the phone will pop into my mind in the next few seconds " (p 413) We will agree that if a person's motivational set is such that he has no reason to or pro-attitude toward answering

the phone, then he will not answer the phone. This is simply an instance of the sort of consideration which supports the alleged Humean conceptual point. However, what is at issue is whether a person with such a motivational set *can* answer the phone. As far as we can tell, the restrictivist has not presented any argument to show that a person with this motivational set lacks the power to call to mind the worry that he might be unable to refrain from answering the phone. If a person has this power, then (even if he actually has no reason or desire to refrain from answering the phone) he does have the power to call to mind a reason not to answer the phone. Given that certain other conditions are satisfied, it is plausible to suppose that he has the power to refrain from answering the phone.

Of course, if it is supposed that causal determinism obtains, then the incompatibilist must say that the agent does not have the power to generate the relevant reasons and thus lacks the power to refrain from answering the phone. But in the context of an assessment of restrictivism, it is not fair to assume causal determinism; after all, the restrictivist's claim is that, even if causal determinism were false, we would rarely be free to do otherwise.

- 29 We are grateful to Sarah Buss, Nancy Schaubert and Eleonore Stump for each calling this objection to our attention.
- 30 See pp. 406-407
31. Dennett, *Elbow Room*, pp. 133ff
32. Harry Frankfurt uses the term "unthinkable" to describe actions which an agent cannot bring himself to will to perform. According to Frankfurt some acts will be unthinkable for an agent because of his moral inhibitions, but "on the other hand, the considerations on account of which something is unthinkable may be entirely self-regarding and without any moral significance." See Harry G. Frankfurt, "Rationality And The Unthinkable," chap. in *The Importance Of What We Care About* (Cambridge. Cambridge University Press, 1988), p. 182. Another reason why an agent may be unable to will something is given by Lehrer's examples of agents who cannot bring themselves to choose to do something because they suffer from a pathological aversion. See Keith Lehrer, "Cans Without Ifs," *Analysis* 29 (October 1968), pp. 29-32, and Lehrer, "Can' In Theory And Practice: A Possible Worlds Analysis," in *Action Theory*, ed. Myles Brand and Douglas Walton (Dordrecht: D. Reidel Publishing Company, 1976), pp. 241-270
- 33 We do not mean to suggest that this is a complete analysis of Raskalnikov's complex character; rather for the sake of brevity we want to limit our comments about his motivations to those which emerge from the passages cited. A more complete analysis would undoubtedly have to consider among other things: (1) the fact that Raskalnikov claims to have felt *beforehand* that he would know *after* the crime that he was only a louse and not an extraordinary man (III 6) and his later demise which seems to confirm this suspicion, (2) his later insistence which challenges the previous claim and suggests that he now, like the extraordinary man, feels no guilt for his crime (Epilogue.2), (3) his claim to have suffered his downfall through "some decree of blind fate" (Epilogue 2), (4) the promised repentance at the end of the book. All these facts point to the need further to refine and revise our abbreviated sketch of Raskalnikov; however, we leave this task to more capable literary critics. For an interesting discussion of these issues, see the collection of critical essays on Raskalnikov in the *Norton Critical Edition of Crime and Punishment*, ed. George Gibian (New York: W. W. Norton & Company, 1975)
- 34 For example, one might object that Raskalnikov commits his crime in order to

show that he is the “extraordinary man” for whom all things are permitted, and therefore he must view his act as being defensible from the broader perspective available to such a person. Even on this reading, however, Raskalnikov still must be seen as doing something he takes to be indefensible given the constraints of conventional morality, for to want to be the “extraordinary man” is to want nothing less than to be free to ignore such moral imperatives.

35. Susan Wolf, “Asymmetrical Freedom,” in *Moral Responsibility* p. 206 emphasis added. Watson outlines this worry in his excellent article “Free Action and Free Will,” *Mind* 96 (April 1987), pp. 145-172
36. *Elbow Room*, p. 133
37. It is interesting to note that Wolf, like van Inwagen, begins with an example—that of allowing one’s children to be incinerated—which is an action most people would find both indefensible and unthinkable, and then moves to an example—that of punching one’s neighbor—which most people would just find indefensible. Since we want to focus on the question of whether the mere ability to do indefensible things does indeed make one crazy, we will concentrate on her second example.
38. For example, Descartes in his Fourth Meditation claims. “It is free will alone or liberty of choice which I find to be so great in me that I can conceive no other idea to be more great, it is indeed the case that it is for the most part this will that causes me to know that in some manner I bear the image and similitude of God.” For a modern defense of the view that idea of an unfree will is inconceivable see Brian O’Shaughnessy, *The Will. A Dual Aspect Theory* (Cambridge: Cambridge University Press, 1980); and Rogers Albritton, “Freedom of Will and Freedom of Action,” Presidential Address, *Proceeding of the American Philosophical Association* (November 1985). As Watson points out the truth behind such claims seems to be that “our concept of the will is such that there is no such thing as failing to will; willing is necessarily successful”, from this point, however, “it does not follow that one cannot be prevented from willing, not by having obstacles placed in the path, but by having one’s will pushed as it were toward one path or another”—see Gary Watson, “Free Action and Free Will,” *Mind*, 96 (April 1987), p. 163
39. Watson raises this worry in “Free Action and Free Will,” p. 164.
40. Although we have suggested that examples like those of Augustine challenge even this intuition for if we take Augustine at his word, he seems to be a case of someone who did act against his best interests, in a manner he believed to be indefensible, and still was not crazy. And if this true, if one can exercise this freedom and not be insane, then surely one can simply possess this freedom without being crazy.
41. Wolf puts this latter point well when she asks: “Why would one want the ability to pass up the apple when to do so would merely be unpleasant or arbitrary? Why would one want the ability to stay planted on the sand when to do so would be cowardly and callous? To want autonomy in other words, is to want not only the ability to act rationally but also the ability to act irrationally—but this latter is a very strange ability to want, if it is an ability at all.” (Wolf, *Freedom Within Reason* (Oxford. Oxford University Press, 1990), pp. 55-56.)
42. Although authors like Dennett and Wolf have argued that the freedom to act indefensibly is not only not wanted by rational agents, but further that it may even be a liability to them, other writers have viewed this freedom in quite an opposite fashion; they have pointed to it as a primary source of human dignity. Jeffrie Murphy asks: “Does not each person want to believe of himself, as a part

of his pride in his human dignity, that he is *capable* of performing, freely and responsibly performing, evil acts that would quite properly earn for him the retributive hatred of others? And shouldn't he at least sometimes extend this compliment to others?" (See Jeffrie G. Murphy and Jean Hampton, *Forgiveness And Mercy* (Cambridge: Cambridge University Press, 1988), p. 102.) Presumably, part of the intuition here is that our respect for others stems from seeing them as responsible agents who, even though they are able to do the bad, refrain from doing so and choose instead to act in accord with morality. Indeed Watson suggests (in his article "Free Action And Free Will") that a Kantian conception of moral agency which emphasizes an ability to set ends requires this type of freedom. On this view, we want the freedom to act indefensibly, not because we want to be irrational, but rather because this freedom underlies our unique status as moral agents, it gives us the ability to reorder our values and to change radically the ends which govern our actions.

43. With respect to this last point, it is worth noting that there are good reasons for abandoning this tradition. For a representative sample of articles discussing this position, see Fischer, *Moral Responsibility*, pp. 143-249. Also see John Martin Fischer, "Responsiveness and Moral Responsibility," in *Responsibility, Character, and the Emotions*, ed Ferdinand Schoeman (Cambridge Cambridge University Press, 1987), pp 81-106; Mark Ravizza, "Is Responsiveness Sufficient For Responsibility" (manuscript), and Fischer and Ravizza, "Responsibility And Inevitability" *Ethics* 101 (January 1991).
44. We are grateful to Sarah Buss, David Copp, Carl Ginet, Jonathan Lear, Nancy Schaubert, and Eleonore Stump for their helpful comments on a previous draft of this paper. Previous versions of this paper were read at the UCLA Law and Philosophy Discussion Group, and at the University of California, San Diego