

MARK RAVIZZA

SEMI-COMPATIBILISM AND THE TRANSFER OF
NON-RESPONSIBILITY

(Received 15 September 1993)

I. THE PROBLEM

Background. Philosophers who think that moral responsibility is incompatible with causal determinism usually do so because they accept something like the following argument:

- (1) Freedom to do otherwise is a necessary condition of moral responsibility.
- (2) Freedom to do otherwise is incompatible with causal determinism.

Moral responsibility is incompatible with causal determinism.

Traditionally compatibilists have responded to this argument by accepting the first premise and arguing that the second is false. That is, they accept that moral responsibility requires freedom to do otherwise, but they deny that such freedom is incompatible with determinism. This response has been challenged in recent years, however, by a new form of argument for the incompatibility of freedom and determinism which allegedly avoids the type of modal confusions which long have been the target of compatibilists' objections. The "Consequence Argument" (as this new style of argument has been called) can be formulated in a variety of ways, but typically it is presented by appealing to what may be thought of as a "transfer of powerlessness principle."¹ Such a principle says roughly – I consider a more careful formulation of the principle and argument below – that if an agent has no choice about p, and also has no

choice about the fact that p leads to q , then it follows that the person has no choice about q . Using this intuitively plausible principle, the thought behind the Consequence Argument may be sketched as follows. Causal determinism is the claim that a complete description of the world at a given time and a complete statement of the laws of nature, together entail every fact about the world after that time. If determinism is true, then all our choices and actions are a *consequence* of the events in the distant past and the laws of nature. So, given determinism, there is some state of the world in the distant past P which is connected by the laws of nature to any action A that one performs in the present. But since no one has a choice about the distant past P , and no one has a choice about the laws of nature that lead from P to A , it follows (given the transfer of powerlessness principle) that no one has a choice about any action A that is performed in the present – i.e., no one has any choice to do or to choose otherwise.

Confronted with this powerful, new argument for the incompatibility of freedom and determinism some thinkers have sought new strategies to insulate our ascriptions of moral responsibility from worries posed by the Consequence Argument. One such strategy circumvents these worries by rejecting the traditional acceptance of premise (1) – i.e., the assumption that freedom to do otherwise is a necessary condition of moral responsibility. This approach garners support by appealing to a group of examples first formulated by John Locke, and more recently by Harry Frankfurt, which suggest that agents can act freely and be morally responsible even though they are not free to do otherwise.² Using these “Frankfurt-type” examples, philosophers sympathetic to this approach have argued for a new brand of compatibilism: “semi-compatibilism.”³ According to this view, moral responsibility does not require freedom to do otherwise; and hence, one can be a compatibilist about responsibility and determinism without having to take a stand on the compatibility of freedom and determinism. Given this position, the semi-compatibilist is able to reconcile moral responsibility with causal determinism, even if (assuming the soundness of the Consequence Argument) determinism is incompatible with freedom to do otherwise.

Semi-compatibilism, thus, avoids worries posed by the Consequence Argument, but it may face a related challenge.

A Problem for Semi-Compatibilism. Peter van Inwagen has argued that the same type of transfer principle which underlies the Consequence Argument can also be used to formulate an argument that shows *directly* (without needing to make any assumptions about freedom to do otherwise) that determinism is incompatible with moral responsibility.⁴ In Van Inwagen's "Direct Argument," the transfer principle is stated in terms of responsibility (rather than choice); it says, roughly that if no one is responsible for *p*, and no one is responsible for the fact that *p* leads to *q*, then it follows that no one is responsible for *q*. Employing this principle, the idea behind the Direct Argument can be sketched in a manner analogous to the Consequence Argument: If determinism is true, then there is some state of the world in the distant past *P* which is connected by the laws of nature to any action *A* that one performs in the present. But since no one is responsible for the state of the world *P* in the distant past, and no one is responsible for the laws of nature that lead from *P* to *A*, it follows that no one is responsible for any action *A* that is performed in the present.

Obviously if such a Direct Argument were sound, then both traditional compatibilists and semi-compatibilists would face a similar set of difficulties in meeting incompatibilists' objections, and thus one important motivation for adopting a semi-compatibilist approach would be lost.⁵ Let me review the dialectic. The Consequence Argument threatens the traditional view of moral responsibility – which holds that freedom to do otherwise is a necessary condition of responsibility – by arguing that such freedom is incompatible with determinism. The semi-compatibilist hopes to protect our ascriptions of responsibility from this threat by arguing that since responsibility does not require freedom to do otherwise, it can be reconciled with determinism, even if the Consequence Argument is sound. Against this view, Van Inwagen presents an argument formally similar to one version of the Consequence Argument which claims to establish directly, without any appeal to freedom to do otherwise, that determinism and responsibility are incompatible. This Direct Argument makes similar assumptions about the fixity of the past and the laws, and employs a transfer principle analogous to that found in the Consequence Argument. Thus, the semi-compatibilist is faced with the following task: to explain why, although the Consequence

Argument might show that freedom to do otherwise is incompatible with determinism, the Direct Argument does not show that determinism is incompatible with moral responsibility.⁶

In this paper, I take up this task and seek to defend semi-compatibilism against Van Inwagen's Direct Argument. Toward this end, I first review Van Inwagen's formal presentation of the Direct Argument and its relation to his formulation of the Consequence Argument; then I present a series of counterexamples intended to show that the Direct Argument is invalid because it rests on a faulty transfer principle. Finally, I compare the role that such transfer principles play in the Consequence Argument and the Direct Argument, and I offer a suggestion for why a transfer of non-responsibility might obey different rules than a transfer of powerlessness.

II. THE DIRECT ARGUMENT

Van Inwagen's Arguments. Having sketched the intuitive idea behind both the Consequence Argument and the Direct Argument, let us compare Van Inwagen's more careful presentation of these arguments in order to illustrate their formal similarity and reliance upon analogous transfer principles. First, although the Direct Argument is syntactically similar to a modal formulation of the Consequence Argument, the two arguments differ semantically. Whereas the Consequence Argument employs a modal operator "Np" which is read as "p and no one has, or ever had, any choice about p," the Direct Argument uses an operator "Np" which is read as "p and no one is, or ever has been, even partly morally responsible for the fact that p." Such a change is required because the Direct Argument aims to establish the incompatibility of determinism and responsibility independently of any considerations of freedom to do otherwise.

Second, both arguments depend upon analogous rules of inference (which Van Inwagen calls "Alpha" and "Beta" in the Consequence Argument, and "(A)" and "(B)" in the Direct Argument):

Rule Alpha/(A): From $\Box p$ deduce Np (' \Box ' represents "standard necessity": truth in all possible circumstances.)

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

Beta is Van Inwagen's formulation of the transfer of powerlessness principle; it says roughly that if no one has ever had a choice about p , and no one has ever had a choice about the fact that $(p \supset q)$, then no one has ever had a choice about q . Similarly Rule (B) is a transfer principle, but it applies to cases in which the agents lack even partial responsibility for the facts in question. Thus, I will refer to (B) as a "transfer of non-responsibility" principle. (B) captures the intuition that if no one is, or ever was, responsible for p , and no one is, or ever was responsible for $(p \supset q)$, then it follows that no one is, or ever was responsible for q .

Finally both the Consequence Argument and the Direct Argument employ the following notation: Let "P" abbreviate any sentence expressing a true proposition. Let "L" abbreviate a sentence expressing the conjunction into a single proposition of the laws of nature. Let " P_0 " abbreviate a sentence that gives a complete description of the world at some moment in the remote past before there were any humans.

Using the notation the argument may be formulated as follows.

Since determinism is assumed to be true, it follows that $\Box(P_0 \& L \supset P)$. From this consequence the argument proceeds as follows:

- | | | |
|----|-----------------------------------|---------------------------------|
| 1. | $\Box(P_0 \& L \supset P)$ | |
| 2. | $\Box(P_0 \supset (L \supset P))$ | 1; modal and sentential logic |
| 3. | $N(P_0 \supset (L \supset P))$ | 2; Rule Alpha/(A) |
| 4. | NP_0 | Premise |
| 5. | $N(L \supset P)$ | 3,4; Rule (Beta)/(B) |
| 6. | NL | Premise |
| 7. | NP | 5,6; Rule Beta/(B) ⁷ |

If this argument, or more precisely this argument-form is sound, it follows (depending upon how the operator N is interpreted) that if determinism is true, either no one has any choice to do otherwise, or no one

is even partly responsible for anything that he does. The point to be stressed about Van Inwagen's presentation of these arguments is that the Direct Argument differs from the Consequence Argument only in its interpretation of the modal operator N: in the Direct Argument, N is interpreted in terms of moral responsibility (instead of choice) and this results in the argument depending upon slightly different premises about the past and laws, and a different transfer principle. The question to be asked is, does this difference affect the soundness of the argument?

Evaluating the Argument. At first glance, the Direct Argument appears to be sound. In addition to the ordinary rules of modal and sentential logic, the argument depends only upon two intuitive rules of inference and two highly plausible premises. Van Inwagen holds that Rule (A) is beyond dispute, for surely no one is, or ever has been, even partly responsible for any fact that is logically necessary. Similarly, he argues that both premises are true, because no one is, or ever has been, responsible for the state of the world at some time before any humans existed, and likewise no one is, or ever has been, responsible for the laws of nature. Accepting the initial plausibility of these claims – and it seems one should – the compatibilist is left with Rule (B) being the point where the Direct Argument is most likely to be vulnerable to criticism.

Unfortunately, the status of (B) is not easily ascertained. Or so Van Inwagen claims. According to him Rule (B), like its analogue Beta, is one of those intractable principles that seems valid but which is neither easily proved or disproved. There is little hope that (B) could be derived from generally accepted inference rules, for it is difficult to see how a rule concerning the concept of moral responsibility could be deduced from inference rules concerning only non-moral concepts.⁸ In fact, Van Inwagen concedes that perhaps the best, and only real, defense of (B) lies first in noting its intuitive plausibility and then in defying its critics to produce any cases that violate the rule.

Recognizing that (B) is the aspect of the Direct Argument most open to dispute, Van Inwagen summarizes the dialectical position of his argument by posing the following challenge to anyone who would reject its conclusion: "If the compatibilist [with respect to responsibility and

determinism] wishes to refute this argument – and, of course nothing obliges him to do this – here is what he will have to do: he will have to produce some set of propositions intuitively more plausible than the validity of (B) and show that these propositions entail the compatibility of moral responsibility and determinism, or else he will have to devise a counterexample to (B), a counterexample that can be evaluated independently of the question of whether moral responsibility and determinism are compatible.”⁹

I believe that Van Inwagen’s challenge can be met. To do so, I will accept the latter task he sets for the compatibilist: I will propose a number of counterexamples to show that (B) is invalid. I begin by discussing counterexamples involving responsibility for consequences;¹⁰ then I show how similar examples also can be constructed for cases of responsibility for action.

III. COUNTEREXAMPLES TO (B): PRE-EMPTIVE OVERDETERMINATION

Why Accept (B)? The initial plausibility of Rule (B) comes, I think, from the fact that in most ordinary cases it intuitively seems as if non-responsibility can be “transferred” in the way the principle suggests. That is, if a person is not responsible for one thing, and also not responsible for that thing leading to another, then the person seemingly is not responsible for the other. And Van Inwagen appeals to such ordinary cases to bolster support for (B). For example, he cites the following case. If (1) no one is responsible for the fact that John is bitten by a cobra on his thirtieth birthday; and (2) no one is responsible for the fact that if John is bitten by a cobra on his thirtieth birthday, then he dies on his thirtieth birthday; it seems to follow that (3) no one is responsible for the fact that John dies on his thirtieth birthday.¹¹ Van Inwagen’s example is not unique. Similar cases are easily found. To take just one: if (1) Mrs. Brown is not responsible for the fact that the rainfall in Northern California is much below normal in the winter of 1988; and (2) she is not responsible for the fact that if the rainfall in Northern California is much below normal in the winter of 1988, then the state rations water in the summer of 1988; it seems to follow that (3) Mrs. Brown is not

responsible for the fact that the state of California rations water in the summer of 1988.

Rule (B), then, does seem to capture some intuitions about the transfer of non-responsibility. Yet what explains these intuitions? Despite the purported independence of the Direct Argument from any concerns with alternative possibilities, it does seem that part of what drives our intuitions in these cases may be little more than the sense that an agent is not responsible for something that is inevitable. The idea here is that if there are conditions present (for which an agent is not responsible) and these conditions are sufficient to ensure a given outcome, then it seems as if the agent should not be held responsible for that outcome.¹²

But the Frankfurt-type cases have allegedly supplied reasons to doubt such general intuitions concerning the lack of responsibility for the inevitable. In particular, these examples show that conditions can be present which are sufficient to ensure that an agent perform a certain task, but as long as these conditions play no role in what *actually* happens, the agent can still be responsible for the action. Adopting the strategy behind these examples suggest one promising way to undermine the initial plausibility of Rule (B).

The Plan of Attack. To develop this criticism of Rule (B), I will first show that familiar versions of the Frankfurt-type examples call into question a principle closely related to (B) which I will call “(B*).” Then I will extend these examples to show that with certain modifications they can also serve as counterexamples to (B). I follow this strategy for three reasons. First, although Van Inwagen formulates his argument by appealing directly to the validity of Rule (B), other incompatibilists have formulated similar arguments by starting with the related principle (B*); hence, it is worth showing that the Frankfurt-type examples invalidate both (B) and (B*).¹³ Second, by formulating (B) in a way that requires that no one is, or ever has been, even partly responsible for the fact in question, Van Inwagen has made it quite difficult to find natural counterexamples to (B). Indeed the counterexamples which I will eventually propose are rather artificial precisely because of this difficulty. In light of the artificiality of the examples, our intuitions concerning how

responsibility should be ascribed in such cases might seem less than clear. To help clarify these intuitions, I want to start with fairly standard Frankfurt-type examples about which our intuitions are more certain, and then gradually build up to the more unusual cases. Finally, given the artificial nature of the last set of counterexamples, some might be tempted to think that the examples themselves are so contrived that they do not pose a serious threat to (B). I believe this type of criticism is in error. The factual oddity of the examples does not stem from any principled moral difference between the examples I eventually will offer and the usual Frankfurt-type examples. By beginning the discussion with the more standard versions of the Frankfurt-type examples, I hope to illustrate the connection between these examples and my counterexamples to (B), and in this way defuse some of the strangeness of this latter group.

First Counterexample. Consider a case called ‘Avalanche.’ Betty is a double-agent who has been instructed to start an avalanche that will destroy an enemy base at the foot of a large, snow-capped mountain. To accomplish her mission, Betty places dynamite in the cracks and crevices of a glacier near the top of the mountain. At T1 she pushes the plunger detonating the explosives and starting an avalanche. The avalanche rumbles down the hill, gaining ever greater force, until some time later, say T3, it crushes the enemy outpost. Assume that the success of the mission depends upon the base being destroyed at exactly T3. Given that Betty acts freely in setting the explosives and in starting the avalanche, it seems that she is responsible for her action and for the consequence to which it leads: *that the enemy base is crushed by an avalanche at T3*. Unbeknownst to Betty, however, another soldier from her army, Ralph, is hiding slightly below her on the mountain. Betty’s commanding officers had reason to doubt her loyalty, and to test it they assigned her the task of destroying the enemy camp. But to ensure that the mission succeeded, they secretly sent Ralph along with instructions to start the avalanche himself (if Betty did not) by using explosives placed a few feet below Betty’s. In this way, if Betty had not detonated the explosives at T1, Ralph still would have had time to detonate his own explosives at T2, thereby ensuring the consequence *that the enemy base*

is crushed by an avalanche at T3. This example is structurally similar to a Frankfurt-type example. Since Ralph, the counterfactual intervener, plays no role in the actual sequence of events, it seems his presence should not affect Betty's responsibility for destroying the enemy camp. This is true even though Ralph's presence is sufficient to ensure that the camp is destroyed at T3. Of course, Ralph's presence might have a bearing on Betty's responsibility for certain types of consequences. For example, given Ralph's possible intervention, one could plausibly argue that Betty is *not* responsible for a certain type of "modalized" fact like *it being the case that inevitably the enemy camp is destroyed by either Betty or Ralph*. But insofar as Ralph's presence has absolutely no bearing on what actually takes place, it certainly should not affect Betty's responsibility for the "non-modalized" fact *that the enemy camp is crushed by an avalanche at T3*. Intuitively Betty is responsible for this consequence because (1) she is responsible for her act of detonating the explosives, and (2) her action brings about the avalanche which destroys the enemy base.¹⁴

Cases of pre-emptive overdetermination, like "Avalanche," provide good reason to doubt the validity of (B). For even though

- (1) Betty is not responsible for the fact that Ralph is present;

and

- (2) she is not responsible for the fact that if Ralph is present, then the enemy base is destroyed at T3;

still

- (3) she *is* at least partly responsible for the fact that the base is destroyed by an avalanche at T3.

Of course such examples are not, strictly speaking, counterexamples to Van Inwagen's Rule (B). (B) requires that *no one* is, or ever was, even partly responsible for the facts in question, and in "Avalanche," it is quite natural to argue that *someone* (perhaps Betty's commanding officer) is partly responsible for the fact that Ralph is present, and for

the fact that if Ralph is present, then the enemy base is destroyed by an avalanche at T3. But though this is a factual departure from the requirements of (B), it is not clear that this should make any moral difference. Let me explain.

Rule (B) vs. Rule (B)*. As presently formulated, the operator, N, used in Rule (B) does not refer to any particular person (or group of persons). It requires simply that *no one* is, or ever has been, even partly responsible for the fact in question. Typically, however, when we speak of “not being responsible,” it *is* with reference to particular persons. That is, we are speaking of certain persons not being responsible for particular actions, omissions, consequences, and so forth.¹⁵ To cover this more common range of cases, an operator like N which concerns non-responsibility could be formulated in a way that specifically indexes the agents involved. To make “N” relative to some agent S, I will use the notion “N_S” where “N_S” is read as “p and agent S is not, and never has been, even partly responsible for the fact that p.” Substituting this operator into (B) yields

Rule (B*): From N_Sp and N_S(p ⊃ q) deduce N_Sq.

It seems to me that (B*) appeals to precisely the same intuitions concerning the transfer of non-responsibility that (B) does, and that anyone who is persuaded by (B) ought also to accept (B*).¹⁶ In fact, the Direct Argument can be formulated using (B*) as well as (B) to argue that for any agent and any true proposition, it follows from the truth of determinism that the agent is not responsible for that proposition.¹⁷ Thus, there does not appear to be any *morally* significant reason for distinguishing between the two principles. Accepting this point, the following argument emerges: (i) since any Frankfurt-type example, like “Avalanche,” is clearly a counterexample to (B*), and (ii) since Rule (B) rests on the same intuitions as (B*), then (iii) the Frankfurt-type examples which challenge (B*) also should give one reason to question the validity of Rule (B).

Of course, it would be false to claim that (B) and (B*) are logically equivalent, or that the falsity of (B*) logically implies the falsity of (B).¹⁸

But this is not the aim of the preceding discussion. Rather the point is simply to note that a general strategy which has been highly influential in other areas – i.e., a Frankfurt-type strategy – can be applied to the present debate in order to call into question the intuitions that support a rule like (B*). Hence, anyone who accepts the force of the Frankfurt-type examples against the Principle of Alternative Possibilities should be persuaded by the above arguments against Rule (B*) as well.¹⁹ And insofar as the very same intuitions that lead one to accept (B*) also lead one to accept (B), the Frankfurt-type strategy provides good grounds for doubting the validity not only of Rule (B*) but also of Rule (B).

An incompatibilist still might resist this point by arguing that (B) is significantly different from (B*). He might insist that the transfer of non-responsibility follows different rules in the general case where “no one is, or ever has been, even partly responsible,” and in support of this claim he could cite the fact that it is much more difficult to find plausible counterexamples to (B) than to (B*). But this objection seems mistaken. Although counterexamples to (B) are more difficult to find in ordinary circumstances that are counterexamples to (B*), this difficulty does not stem from any principled moral reason. That no moral difference is at stake is evinced by the fact that even though Frankfurt-type examples are most naturally constructed using counterfactual interveners who are responsible agents, this is not a necessary feature of such examples.²⁰ To illustrate this point, I offer below two Frankfurt-type examples in which no one is, or ever has been, even partly responsible for the conditions which are sufficient to ensure the outcome in question.

Counterexamples to (B). Call the following case “Erosion.” It is exactly like “Avalanche” except that in this instance the counterfactual intervener, Ralph, is to be replaced by natural forces that have no conscious design and bear no responsibility. As before imagine that Betty plants her explosives in the crevices of the glacier and detonates the charge at T1 causing an avalanche that crushes the enemy fortress at T3. Unbeknownst to Betty and her commanding officers, however, the glacier is gradually melting, shifting, and eroding. Had Betty not placed the dynamite in the crevices, some ice and rocks would have broken free starting a natural avalanche that would have crushed the enemy camp at

T3. As in “Avalanche,” Betty acts freely, and she is responsible for the consequences of her action. This is true though conditions were present, for which no one was even partly responsible, that were sufficient to bring about the consequence *that the enemy camp is destroyed by an avalanche at T3*. Thus, “Erosion” is counterexample to (B), for

- (1) The glacier is eroding and no one is, or ever has been, even partly responsible for the fact that it is eroding;
- (2) If the glacier is eroding then there is an avalanche that crushes the enemy base at T3, and no one is, or ever has been, even partly responsible for this fact;

but given Betty’s responsibility, it is *not* true that

- (3) There is an avalanche that crushes the enemy base at T3, and no one is, or ever has been, even partly responsible for this fact.

“Erosion” shows that Rule (B) is invalid as regards consequences. A similar strategy can also be deployed to show that (B) is invalid in the case of actions. Consider a Frankfurt-type case similar to “Avalanche”; call it the “Dutiful Soldier.” Once again, Betty detonates the explosion at T1 resulting in the enemy camp being crushed by the avalanche at T3. Unbeknownst to Betty, she has been hypnotized by the army, and if she had even begun to hesitate from doing her duty, she would have been overcome by an irresistible post-hypnotic suggestion to push the plunger and start the avalanche. This standard Frankfurt-type example clearly is a counterexample to (B*). Insofar as the hypnotic suggestion plays no actual role, Betty is responsible for her action. This is true even though conditions are present (for which Betty is not responsible) that imply that Betty does push the plunger at T1. Now, if Betty is responsible in this case, it seems she also is responsible in the following version of the example which has been modified slightly so that it is a counterexample not only to (B*) but to (B) as well. Imagine that unbeknownst to Betty she is born with a strange genetic trait (for which no one is responsible) that compels her to obey orders if she even begins to disobey them. In this example, Betty once again freely does her duty

and detonates the explosion. The compulsive dutiful trait plays no role in her action. Nevertheless, if Betty had even begun to shirk her duty, her innate character trait to obey orders would have come to the fore and compelled her to start the avalanche. Assuming that no one is responsible for the trait – and I stipulate that in this example Betty is born with her strong sense of duty by chance²¹ – we once again have a counterexample to (B), this time for actions.

Cases like “Erosion” and the “Dutiful Soldier” illustrate that agents can be responsible for their actions and the consequences of those actions, even though there are conditions present (for which they are not responsible) that are sufficient to ensure the given actions and consequences. These examples constitute clear counterexamples to Rule (B). I conclude, therefore, that as presently formulated the Direct Argument is invalid because it depends upon an invalid transfer principle.

IV. OBJECTIONS: SIMULTANEOUS VS. PRE-EMPTIVE OVERDETERMINATION

A Worry About the Counterexamples. The counterexamples to (B) discussed above are all cases of pre-emptive overdetermination. In each case there is some condition for which Betty is not responsible – e.g., the eroding glacier in “Erosion,” the compulsive dutifulness in “Dutiful Soldier” – and this condition is sufficient to ensure the event in question. Call such a condition an “Ensuring Condition.” In each of the Frankfurt-type examples above, the Ensuring Condition plays a role only in the alternative scenario. In the actual sequence of events it is pre-empted by actions for which Betty *is* responsible, and hence intuitively she is accountable for her actions and their consequences. Indeed, the Frankfurt-type strategy seems to require such pre-emption, since it justifies holding agents responsible only if the Ensuring Condition is not actually efficacious in bringing about the event in question.

Acknowledging this point, a critic might argue as follows: “All that has been presented thus far are instances in which an agent is responsible for something which is pre-emptively overdetermined. But these cases are irrelevant to the matter at hand. Ultimately our discussion is concerned with the ostensible lack of responsibility that stems from

the truth of determinism, the laws of nature, and some initial state of the world. Unlike the role an Ensuring Condition plays in a Frankfurt-type example, where it is causally efficacious only in the *alternative* sequence of events, these conditions are causally efficacious in the *actual* sequence. Thus, the counterexamples presented above do not undermine (B) in any sense that is relevant to the compatibility of responsibility and determinism. What is needed is a counterexample in which an Ensuring Condition does play a role in the actual sequence, and the agent still is responsible. Barring such an example, an incompatibilist is still free to argue that if (i) there are conditions for which no one is, or ever has been, even partly responsible, and (ii) these conditions are sufficient to ensure a given event, and (iii) these conditions *play a role in the actual sequence that brings about this event*, then it follows that no one is responsible for the event in question. Since the laws of nature and the initial state of the world, together with the truth of determinism, do satisfy these conditions and entail every subsequent fact, it follows that determinism is incompatible with moral responsibility.”

The critic’s worry here stems from the fact that although Frankfurt-type examples show that a person can be responsible in cases where an Ensuring Condition only would come into play in some alternative sequence, these examples do not say anything about an agent’s responsibility in situations where an Ensuring Condition plays a role in the actual sequence. Hence, given that the worry about determinism regards conditions that do play a role in the actual sequence, Frankfurt-type counterexamples appear to be irrelevant.

Simultaneous vs. Pre-emptive Overdetermination. Is the critic right? Does it matter whether the Ensuring Condition plays a causal role in the actual sequence of events? With respect to the examples given above, some philosophers have worried that cases of simultaneous overdetermination might be sufficiently different from cases of pre-emptive overdetermination.²² For example, one might feel that if the erosion actually does start an avalanche at exactly the same moment that Betty’s dynamite explodes, then she is no longer responsible for the consequence *that an avalanche starts at T1 which crushes the enemy base at T3*. Call this “Erosion*.” The intuition here is that, unlike a

Frankfurt-type example, the effects of erosion in “Erosion*” cannot be set aside on the ground that they play no role in what actually happens. On the contrary, the erosion is part of the actual sequence that causes the avalanche, and insofar as this makes it inevitable *that an avalanche starts at T1 which crushes the enemy base at T3*, Betty should not be held responsible for this consequence. Of course, this judgment will only affect the *content* of Betty’s responsibility and not the *degree* to which Betty can be lauded for her loyalty and praised for the success of the mission. Betty is still responsible in this case, as in “Erosion,” for the act of detonating the explosives which are sufficient in themselves to start the avalanche, and hence her moral credit should not be diminished. However, the critic here will urge that this misses the point. What is important he will say, is that there is a difference in the content of responsibility between cases of pre-emptive and simultaneous overdetermination. Whereas in cases of pre-emptive overdetermination, an Ensuring Condition plays no role in what actually happens, and hence the agent is still responsible for the inevitable consequence, in cases of simultaneous overdetermination, an Ensuring Condition does play a role in what actually happens, and as a result the person is not responsible for the inevitable outcome. If this analysis is accepted, then determinism would seem to be incompatible with responsibility, for obviously if everything were causally determined, then the laws of nature and the initial state of the world would play a role in bringing about everything that happens, and hence no one would be responsible.

I think this criticism is mistaken. Admittedly our intuitions concerning cases of simultaneous overdetermination like “Erosion” are fairly tentative, and with a bit of persuasion they can be pulled in either direction. However, I believe this is due more to the usual nature of the examples, and to the theoretical framework (borrowed from the Frankfurt-type examples) used to analyze these cases, than to any morally significant reason. If, for a moment, this emphasis on actual and alternative sequences is set aside, and the examples approached in a fresh light, much of this apparent difficulty can be dispelled.

Collective Responsibility. Consider first a case which usually falls under the rubric of “collective responsibility.”²³ This will serve as a counter-

example to the validity of (B*) for cases of simultaneous overdetermination. In this example, "Joint Assassins," Charley wants to assassinate the President of the United States. As the President steps to the podium at T1, Charley aims at her head and fires a shot. At precisely the same moment, another assassin, Julie, from another part of the crowd also fires a shot at the President's head. Both bullets find their mark simultaneously, and the President is killed at T2. Neither assassin knew the other assassin was present, and either bullet alone would have been sufficient to kill the President. In this case, it seems natural to suppose that both persons are at least partly responsible for the consequence *that the President is killed at T2*. Each agent may in fact be fully responsible for the consequence, but at the very least each assassin is partly responsible for it.²⁴ After all, since neither agent was aware of the presence of the other agent, and each one's bullet was sufficient to kill the President, it would be highly implausible to argue that the fortuitous presence of the other assassin should somehow create a situation in which neither assassin was even partly morally responsible for the fact that the President is killed.

Granting that each assassin is at least partly responsible for the assassination of the President, "Joint Assassins" present a clear counterexample to (B*). Surely Julie is not even partly responsible for the fact that Charley shoots the President in the head at T1, nor is she even partly responsible for the fact that if Charley shoots the President in the head at T1, then the President dies at T2. Nevertheless, Julie is at least partly responsible for the fact that the President dies at T2.

Next consider an analogous example – "Joint Avalanche." This example is similar to "Avalanche" except for the fact that in addition to Betty, there is another soldier who starts the avalanche at exactly the same time Betty does. Assume that neither Betty nor the second soldier realizes that the other is present. If one accepts that Julie is at least partly, if not fully, responsible in a case like "Joint Assassins," then certainly a similar conclusion should follow here. Moreover, it seems that Betty's responsibility for the avalanche should not be affected even if the example is modified slightly so that the second event that triggers the avalanche is initiated by natural causes, rather than by a responsible agent. For example, assume that instead of another soldier waiting to

start an avalanche, there is a wandering goat. Unbeknownst to Betty, just as she detonates her explosives at T1, the goat jars loose a large boulder which also starts the avalanche that crushes the enemy base at T3. Both Betty's explosives and the large boulder are sufficient to start the avalanche, and each causally contributes to initiating the avalanche. In this case

- (1) the goat knocks the boulder loose at T1 and no one is, or ever has been, even partly responsible for this fact;

and

- (2) if the goat knocks the boulder loose at T1, then the avalanche crushes the enemy camp at T3, and no one is, or ever has been, even partly responsible for this fact;

and still

- (3) the enemy camp is crushed at T3, and Betty *is* at least partly responsible for this fact.

In short, if one accepts the common view about collective responsibility in the case of "Joint Assassins," one should also accept that a person can be at least partly responsible in a case like "Joint Avalanche." And if this is right, then Rule (B) is invalid for cases of simultaneous overdetermination just as it was invalid for cases of pre-emptive overdetermination.²⁵

Ensuring Conditions in Actual and Alternative Sequences. Before leaving this discussion, I would like to clear up a possible source of confusion. In the examples concerning pre-emptive overdetermination a great deal of weight was placed upon the distinction between what happens in the actual sequence of events and what would have happened in some alternative scenario. I argued that the person's responsibility in these examples depends upon the fact that the Ensuring Condition is pre-empted – insofar as this condition is pre-empted, it plays no role in what actually happens, and hence the agent still may be responsible for his actions and their consequences. However, the subsequent discussion

of simultaneous overdetermination may have appeared to undermine the importance of this distinction, for in these cases the Ensuring Condition is not pre-empted – it *does* play a role in the actual sequence, and nevertheless the agent is responsible. How are these claims to be reconciled?

The answer lies in attending more carefully to why an Ensuring Condition might be thought to undermine a person's responsibility. The point which emerges from this discussion of pre-emptive overdetermination is that if an Ensuring Condition plays *no* role at all in what actually happens, then clearly it should not undermine the person's responsibility, even though it makes the outcome of his actions inevitable. This point can be granted, however, without accepting as a corollary the converse point that whenever an Ensuring Condition *does* play a role in what actually happens, then the person's responsibility should be undermined. As the cases of simultaneous overdetermination illustrate, what matters is not just whether an Ensuring Condition plays a role in the actual sequence, but also *what type* of role this Condition plays. For example, in the case of "Joint Assassins," even though the other assassin is causally efficacious in the actual sequence, his action neither undermines Julie's free action of shooting the President, nor does it interfere with the way in which Julie's action affects the world – i.e., the bullet is fired, finds its way to the President's head, and kills her just as it would have had the other assassin not been present. Hence, Julie is responsible.²⁶

V. BETA AND RULE (B)

Who Needs Beta and (B)? If the above arguments are sound, then a transfer of non-responsibility principle like (B) is invalid. Accepting this point, an interesting contrast emerges concerning the role of Beta in arguments for the incompatibility of freedom and determinism, and the role of (B) in arguments for the incompatibility of responsibility and determinism. Many philosophers, compatibilists and incompatibilists alike, have thought that the success of the incompatibilist's argument concerning freedom to do otherwise hinges upon the validity of some

“transfer of powerlessness” principle akin to Beta.²⁷ Acceptance of this view has encouraged compatibilists to undermine the incompatibilist’s position by blocking the modal inference on which it supposedly rests.²⁸ However, ultimately this type of “Beta-blocking” strategy is much less promising than might initially have been thought. First, it remains highly controversial whether any convincing counterexamples to Beta can be constructed.²⁹ (Certainly, Frankfurt-type examples will not work.) Second, and more importantly, even if counterexamples to Beta were found, the intuitions behind the Consequence Argument could simply be reformulated in ways that do not directly require any modal principles like Beta.³⁰ To illustrate this point, consider the following sketch of an argument which is based on a parallel argument for the incompatibility of free will and divine foreknowledge.³¹

The argument rests upon two principles which are controversial though not implausible. The first principle express 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 is 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. 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 the 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 then none of us is free to do other than what he does.

My purpose in citing this argument is not to offer yet another endorsement for incompatibilism,³² but rather to illustrate that the intuitions behind incompatibilism can be formulated in ways that do not depend directly upon a transfer of powerlessness principle. As a result, the debate over the compatibility of determinism and freedom to do otherwise cannot be reduced to a debate over the validity of Beta.

But whereas the debate about Beta is something of a red herring as regards the compatibility of determinism and freedom, the analogous debate over the validity of Rule (B) may be precisely what is at issue concerning the compatibility of determinism and responsibility. In the first place, if the above arguments are sound, then there are convincing counterexamples to Rule (B). Second, unlike the Consequence Argument which can be formulated without appealing to any transfer principles, the Direct Argument appears to depend crucially upon the transfer of non-responsibility principle. Of course, the incompatibilist might hope that just as the Consequence Argument can be reformulated

without using Beta, so analogous reformulations could be found for the Direct Argument that make no use of (B). But this seems unlikely. Other formulations of the Consequence Argument (like the argument just given) depend directly upon the intuition that one does not have the power to alter the distant past and/or laws of nature, and these intuitions are not easily translated into claims concerning responsibility.³³ As a result, the incompatibilist must either refute the counterexamples presented above, or else concede that the Direct Argument, as presently formulated, is invalid.³⁴

Why Is (B) Invalid, but Beta Valid? Noting this contrast between the role of Beta in the free will debate and the role of (B) in the responsibility debate, one might wonder why a transfer of powerlessness principle like Beta should most likely be valid, and yet an analogous transfer principle concerning responsibility be invalid. Part of an answer to this question can be found by looking more closely at why the Frankfurt-type counterexamples work against (B) but not against Beta. Speaking metaphorically, in both the cases of pre-emptive and simultaneous overdetermination there are several “paths” leading to the same outcome. The counterexamples to (B) are set up so that the Ensuing Condition for which no one is responsible is “on one path” (call this the “Ensuring Path”), while the agent’s free action and its corresponding effects make up the other path. Insofar as the Ensuring Condition does not inappropriately affect the causal path which “runs through” the agent’s action to the outcome in question, the agent can still be responsible for the outcome. Intuitively the agent is responsible in these cases because “on his path” he acts freely and the world is “sensitive” to his action in just the way it would have been had the Ensuring Path not been present.³⁵ According to this picture, Rule (B) is invalid. For even though

- (1) there is some Ensuring Path leading to a particular outcome and no one is, or ever has been, even partly responsible for this fact;

and

- (2) if there is this Ensuring Path, then the outcome is reached, and no one is, or ever has been, even partly responsible for this fact;

it does *not* follow that

- (3) the outcome is reached and no one is, or ever has been partly responsible for this fact.

An agent still can be responsible for the outcome in question if his causal path is of the proper sort and the Ensuring Condition plays no inappropriate role in what actually happens along this causal path.

But when we turn to Beta, we see that an analogous strategy cannot be deployed to invalidate this principle. Beta is concerned with a transfer of powerlessness, and here the presence of even a single Ensuring Condition suffices to render an agent powerless to prevent the outcome in question. Given that

- (1) there is some Ensuring Path leading to a particular outcome and no one is, or ever has been, free to do anything to prevent this fact;

and

- (2) if there is this Ensuring Path, then the outcome is reached, and no one is, or ever has been, free to do anything to prevent this;

it *does* seem to follow that

- (3) the outcome is reached and no one is, or ever has been, free to do anything to prevent this fact.³⁶

Thus, in contrast to Rule (B), in evaluating the validity of Beta it is irrelevant whether the Ensuring Condition affects the actual causal trajectory which runs through the agent's action to the outcome. The mere presence of an Ensuring Path, by definition, renders the agent powerless to prevent the outcome. Hence, neither the Frankfurt-type examples

nor the cases of simultaneous overdetermination will show that Beta is invalid.

In sum, whereas Rule (B) is vulnerable to certain types of counter-examples, Rule Beta is not, and the preceding picture helps to explain why. As regards a transfer of powerlessness, if an agent is powerless to prevent a path that ensures a given outcome, then he is powerless to prevent that outcome (regardless of whether the Ensuring Condition actually plays a role in the path that leads to the outcome); in contrast, as regards a transfer of non-responsibility, even if some Ensuring Path is present for which the agent is not responsible, as long as this path does not play an inappropriate role in the causal trajectory in which an agent freely brings about a given outcome, that agent can still be responsible for the outcome.

Why (B) Seemed Plausible. This metaphorical talk of “various paths leading to an outcome” can also help us to see why the examples chosen by incompatibilists initially make Rule (B) seem so plausible. Consider one of the examples Van Inwagen uses to elicit support for (B):

- (1) N Plato died in Antiquity
- (2) N(Plato died in Antiquity \supset Plato never met Hume)
- (3) *hence*, N Plato never met Hume.³⁷

In this example, the premises do entail the conclusion. But the validity of this inference stems not from any general truth about the transfer of non-responsibility, but rather from a special feature of the example itself. In contrast to a case like “Avalanche” – where there are a number of different paths to the outcome, including one in which Betty’s free action initiates a causal chain that brings about the destruction of the enemy camp – in the above example there is no path in which anyone can so act that Plato meets Hume. Since agents cannot initiate causal paths that flow into the past, the Ensuring Condition (i.e., the fact that Plato died in antiquity) precludes there being any path in which an agent freely acts to bring about Plato’s acquaintance with Hume. It is this feature of the example, not the general validity of (B), which leads us to accept the inference from the premises to the conclusion.

A similar picture also seems to be what persuades one to accept the other example Van Inwagen offers in support of (B). In this case Van Inwagen argues:

- (1) N John was bitten by a cobra on his thirtieth birthday
- (2) N (John was bitten by a cobra on his thirtieth birthday \supset John dies on his thirtieth birthday)
- (3) *hence*, N John died on his thirtieth birthday.³⁸

This example, call it “Snake Bite,” gains plausibility only to the extent to which one pictures there being only one path to John’s death and that on that path no agents freely act to bring about the death. Change this interpretation of the situation, and the status of the example becomes much more problematic. For instance, imagine that just as the cobra bites John, an assassin (who is unaware of the cobra) shoots John in the head. The bullet and the venom kill John simultaneously, and each is sufficient to bring about his death. This case is now similar to “Joint Assassins.” As argued above, if one assassin acting alone is responsible for the victim’s death, then it would be highly implausible to think that the fortuitous addition of another assassin should result in *neither* assassin being even partly responsible for the death of the victim. And if this conclusion follows in “Joint Assassins,” then a similar point should hold in this modified version of “Snake Bite.” The assassin in “Snake Bite,” then is at least partly responsible for the fact that *John died on his thirtieth birthday*, and hence the example is a counterexample to (B).

In short, the examples Van Inwagen uses to support (B) derive their plausibility from exploiting special features of the situation which are not common to all cases involving a possible transfer of non-responsibility. These examples depend upon there not being any other paths to the outcome in which an agent freely acts to bring about the outcome in question. Once this feature is removed, even Van Inwagen’s own examples, like “Snake Bite,” become problematic. Of course, an incompatibilist may still feel that causal determinism is a condition that precludes there being any path in which an agent can freely act to bring about a given outcome. But this claim has been shown to be little more than an opinion in search of justification. The very resources which the

incompatibilist had hoped to use to show that determinism is incompatible with responsibility have all been found wanting. Let us review the dialectic.

Conclusion: Responsibility and Determinism. The traditional explanation for why responsibility and determinism should be incompatible is that responsibility requires freedom to do otherwise, and this type of freedom is not compatible with causal determinism. But as Frankfurt has argued, an explanation of this sort may be repudiated by appealing to examples which illustrate that freedom to do otherwise is not a necessary condition for responsibility. Next, incompatibilists argue that the incompatibility of responsibility and determinism follows directly from a transfer of non-responsibility principle: since no one is responsible for the past and the laws, and no one is responsible for the fact that together these entail every subsequent fact, it follows that no one is responsible for any fact at the present time. Yet this strategy fails too. By appealing to cases of pre-emptive overdetermination like those found in Frankfurt-type cases, the transfer of non-responsibility principle is shown to be invalid. Finally, to counter this last objection, incompatibilists can argue that a transfer of non-responsibility validly applies only to cases in which the conditions in question play a role in the actual sequence. However, this refined argument is refuted by noting that Rule (B) is invalid even in cases of simultaneous overdetermination; as a case like “Joint Avalanche” shows, the mere presence of an Ensuring Condition in the actual sequence does not suffice to undermine an agent’s responsibility. Accepting these points, one might still argue that determinism is incompatible with responsibility because determinism in the actual sequence would involve a type of force akin to that traditionally associated with other factors that intuitively undermine responsibility (such as hypnosis, direct brain manipulation, and the like). But what would justify such an assertion? The incompatibilists’ usual appeals to freedom to do otherwise and transfers of non-responsibility have been discredited. And without these usual explanations, incompatibilists cannot simply insist that causal determinism undermines responsibility without begging the very question they hoped to settle.³⁹

NOTES

¹ The term “transfer of powerlessness” is used by John Martin Fischer in “Introduction: Responsibility and Freedom,” in *Moral Responsibility*, ed. Fischer (Ithaca: Cornell University Press, 1986), 19. Peter van Inwagen uses this term “Consequence Argument” to designate his argument for the incompatibility of free will and determinism, and he offers three different formulations of the argument. See Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983). I will use this term in a slightly broader sense to refer to any argument in the new family of arguments for incompatibilism: these new arguments, in contrast to older arguments which appealed mainly to the idea of universal causation, begin with general principles about the fixity of the past and the laws of nature, and then argue to the conclusion that determinism is incompatible with freedom to do otherwise. For a representative sampling of the different ways in which the Consequence Argument can be formulated see Carl Ginet, “In Defense of Incompatibilism,” *Philosophical Studies* 44 (November 1983): 391–400; Carl Ginet, *On Action* (Cambridge, Cambridge University Press, 1990); James W. Lamb, “On a Proof of Incompatibilism,” *Philosophical Review* 86 (January 1977): 20–35; Peter van Inwagen, “A Formal Approach to the Problem of Free Will and Determinism,” *Theoria* 40 (1974): 9–22; Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983); and David Wiggins, “Towards a Reasonable Libertarianism,” in *Essays on Freedom of Action*, ed. Ted Honderich (London: Routledge & Kegan Paul, 1973), 31–36. In the following discussion of the Consequence Argument, I follow Van Inwagen’s presentation in *An Essay on Free Will*, pp. 56ff.

² See John Locke, *Essay Concerning Human Understanding* (Oxford: Clarendon Press, 1894), Book II, Chapter xxi, Section 8–11; and Harry Frankfurt, “Alternate Possibilities and Moral Responsibility,” in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), 1–10. A representative sample of other arguments in this debate are collected in *Moral Responsibility*, ed. Fischer.

³ See John Martin Fischer, “Responsiveness and Moral Responsibility,” in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, ed. Ferdinand Schoeman (Cambridge: Cambridge University Press, 1987), 102.

⁴ Peter van Inwagen, *An Essay on Free Will*, 183ff.

⁵ Of course, avoiding the difficulties posed by the Consequence Argument is not the only, nor even the most important, motivation for adopting a semi-compatibilist theory. The primary motivation comes from the Frankfurt-type examples which illustrate that moral responsibility does not require alternative possibilities.

⁶ In stating the challenge this way, I do not intend to endorse the Consequence Argument; rather, the issue I wish to explore is whether the semi-compatibilist can remain agnostic with respect to the soundness and validity of the Consequence Argument, and yet still maintain that the Direct Argument is invalid.

⁷ This argument is taken from Van Inwagen, *An Essay on Free Will*. See pp. 55–105

for his presentation of the Consequence Argument, and p. 185 for his presentation of the Direct Argument.

⁸ *Ibid.*, 187. Van Inwagen also agrees that appealing to the methods of formal semantics will be of little help, since anyone who doubts (B) will probably also question the attempts formally to defend Beta and (B) by appealing to possible-world semantics. See Van Inwagen's discussion of this strategy in connection with Beta on pp. 96–97 and in connection with (B) on p. 186.

⁹ *Ibid.*, 188.

¹⁰ In the discussion that follows I shall treat consequences as consequence-universals. For our purposes, the distinction between consequence-particulars and consequence-universals can best be understood in terms of a criterion of individuation. A consequence-particular is individuated in terms of causal genesis. On this view, the causal path leading to an event is an essential feature of the event; thus, a consequence-particular as *y* if and only if *x* and *y* have the same causes. In contrast, a consequence-universal is individuated more coarsely: the same consequence-universal can be brought about via any number of different causal paths.

¹¹ Van Inwagen, 187.

¹² Of course, the proponent of the Direct Argument would probably resist this suggestion, since the argument is not supposed to depend upon any intuitions concerning freedom to do otherwise. However, if the incompatibilist does resist the suggestion and is willing to grant, for the sake of argument, that one can be responsible for something which is inevitable, then I find it hard to see why anyone should still be persuaded that a person cannot be responsible for a given outcome, simply in virtue of not being responsible for conditions which are sufficient to ensure that outcome.

¹³ For one version of this type of argument, see John Martin Fischer, "Incompatibilism," *Philosophical Studies* 43 (January 1983): 127–137; cf. Carl Ginet, "Might We Have No Choice?" in *Freedom and Determinism*, ed. Keith Lehrer (New York: Random House, Inc., 1966), 87–104.

¹⁴ This intuitive notion of being responsible for consequences that one freely brings about needs to be refined in order to avoid certain objections that arise in connection with consequences that are inevitable. Such objections are discussed and the necessary refinement offered in John Martin Fischer and Mark Ravizza, "Responsibility and Inevitability," *Ethics* 101 (January 1991): 258–278; and Fischer and Ravizza, "Responsibility for Consequences," *Festschrift for Joel Feinberg*, ed. Jules Coleman and Allen Buchanan (forthcoming). To avoid overcomplicating this paper, I do not discuss these more technical points here. Rather, the following examples are developed in greater detail than is usually required for Frankfurt-type cases. This is done so that the agents are not only *intuitively* responsible for the consequences, but they would come out responsible given any of the theories developed in papers that discuss the relevant objections. See Bernard Berofsky, *Freedom From Necessity* (London: Routledge and Kegan Paul, 1987); Robert Heinaman, "Incompatibilism Without the Principle of Alternative Possibilities," *Australasian Journal of Philosophy* 64 (1986): 266–276; and

William Rowe, "Causing and Being Responsible for What Is Inevitable," *American Philosophical Quarterly* 26 (1989): 153–160.

¹⁵ Strictly speaking, Van Inwagen's formulation of (B) relates not to actions, omissions, or consequences, but to true propositions. Having noted this, I will continue to speak about agents being responsible for actions, omissions, and consequences, since this is closer to our ordinary usage and such language can be translated straightforwardly into talk about true propositions.

¹⁶ Van Inwagen himself argues for an analogous point concerning Beta and a principle he calls Beta-prime (which is simply Beta with an operator indexed to persons and times). And if he is willing to accept that Beta and Beta-prime stand or fall together, there seems to be no principled reason why he should not accept the similar point concerning (B) and (B*). See Peter van Inwagen, "When Is the Will Free?" in *Philosophy of Mind and Action Theory*, Vol. 3, *Philosophical Perspectives*, ed. James Tomberlin, (Atascadero, California: Ridgeview Publishing Co., 1989), 409.

¹⁷ For one version of this type of argument, see Fischer, "Incompatibilism." Fischer's argument for the incompatibility of responsibility and determinism is a transformation of a modal argument for the incompatibility of free will and determinism which Carl Ginet, "The Conditional Analysis of Freedom," in *Time and Cause*, ed. Peter van Inwagen (Dordrecht: Reidel Publishing, 1980), 171–186; and Ginet "In Defense of Incompatibilism," *Philosophical Studies* 44 (November 1983): 391–400. It is interesting to note that both Fischer and Ginet *begin* their arguments with agent-indexed principles akin to (B*) and Beta-prime, then universally quantify these respective principles to get (B) and Beta. The point to be emphasized here is that if a principle like (B) is construed as merely a universally quantified form of an agent-indexed principle like (B*), then it would be highly unlikely that there would be any significant moral difference between the principles.

¹⁸ Even when these principles are applied to conditions for which *no* agent is even partly responsible, as in the Direct Argument, there seems to be a semantic difference in how these principles express the transfer of non-responsibility. (B*) says roughly, "For any agent whatsoever, given the agent is not even partly responsible for p, and he is not even partly responsible for $(p \supset q)$, then he is not even partly responsible for q." In contrast, (B) says roughly, "Given that no one is even partly responsible for p, and that no one is even partly responsible for $(p \supset q)$, then no one is even partly responsible for q." As formulated (B*) is still open to Frankfurt-type counterexamples like "Avalanche" in a way that (B) is not. Thus, even though both (B) and (B*) may be used to formulate versions of the Direct Argument, these formulations will not be equivalent. In particular, the agent-indexed formulation may be susceptible to certain types of counterexamples to which Van Inwagen's Direct Argument is immune.

A similar point holds if we compare two popular versions of the Consequence Argument. Whereas Van Inwagen formulates the Consequence Argument by directly appealing to Beta (a principle which is not indexed to agents), Ginet in his presentation of the Consequence Argument, begins with an agent-indexed "transfer of powerless-

ness principle" analogous to (B*), and then he universally quantifies this agent-specific principle to argue that if determinism is true then no one is responsible. I am grateful to John Fischer for a helpful discussion on these points.

¹⁹ And note that even an incompatibilist such as Van Inwagen accepts the force of Frankfurt's examples against the Principle of Alternative Possibilities (which says that a person is morally responsible for what he has done only if he could have done otherwise): "If Frankfurt has made out a good case for the falsity of PAP (and I think he has), then it would seem that he has undercut the debate between the 'compatibilists' and the 'incompatibilists' (to use the contemporary jargon) in a way very similar to the way in which Hobbes and others undercut the debate between the libertarians and the necessitarians." Peter van Inwagen, "Ability and Responsibility," in Fischer, ed. *Moral Responsibility*, 154.

²⁰ Frankfurt makes this point in "Alternate Possibilities and Moral Responsibility," note 4.

²¹ To avoid any possible objections this may need to be specified further. For example it should be emphasized that Betty is not aware that she had the trait; hence, she could not have "taken responsibility" for having this trait, and in this way become responsible for the actions and consequences that flow from it.

²² William Rowe suggests that this distinction might be important in his article, "Causing and Being Responsible for What is Inevitable." The worries discussed in this section were suggested to me in discussions at different times with William Rowe and David Schmidtz, although neither endorsed the objection as I now present it.

²³ The following example is borrowed from Gregory Mellema, "On Being Fully Responsible," *American Philosophical Quarterly* 21 (1984): 190.

²⁴ Indeed, Mellema uses this example to illustrate that even in cases where agents are not solely responsible for some outcome, they still can be fully responsible for it; see Mellema, "On Being Fully Responsible," 190. Whether the agents are fully responsible is perhaps debatable. For example, Frankfurt argues for a stronger sense of "fully responsible" in which a person is fully responsible for those things "which come about because of what he does and which would not come about if he did otherwise." See Harry G. Frankfurt, "What Are We Morally Responsible For?" in *How Many Questions? Essays in Honor of Sidney Morgenbesser*, ed. Leigh S. Cauman, Isaac Levi, Charles Parsons, and Robert Schwartz (Indianapolis: Hackett Publishing Co., 1983): 326. Fortunately, there is no need to enter into this dispute here, for the only point that is needed for the sake of the current argument is that each of the assassins is at least partly responsible, and this point is accepted by both parties in the dispute. See also R. S. Downie, "Collective Responsibility," *Philosophy* 44 (1969): 66–69; Joel Feinberg, "Collective Responsibility," *Journal of Philosophy* 65 (1968): 674–688; Peter A. French, *Collective and Corporate Responsibility* (New York: Columbia University Press, 1984); Michael J. Zimmerman, *An Essay on Moral Responsibility* (Totowa, N.J.: Rowan & Littlefield, 1988).

²⁵ Note that all of the counterexamples concern responsibility for consequence-

universals. Obviously formulating parallel examples for actions would require even more imaginative cases, but the strategy would essentially remain the same.

²⁶ This suggestion is developed and defended in Fischer and Ravizza, "Responsibility and Inevitability," and "Responsibility for Consequences."

²⁷ For example, 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" (Michael Slote, "Selective Necessity and the Free-Will Problem," 9). Echoing this opinion, Terence Horgan writes, "Slote has described well the deep family resemblances among the various formulations [of the Consequence Argument] and he too has suggested that the different versions probably stand or fall together" Terence Horgan, "Compatibilism and the Consequence Argument," *Philosophical Studies*, 47 [1985]: 339).

The view that the Consequence Argument implicitly requires some Beta-like transfer principle is even supported by Van Inwagen. He offers three formulations of the Consequence Argument, only one of which explicitly depends on Beta. Yet despite the differences between these formulations, van Inwagen insists that all three versions of the Consequence Argument should "stand or fall together." Further, he claims that "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*" (*An Essay on Free Will* 57).

²⁸ Michael Slote develops this strategy in "Selective Necessity and the Free-Will Problem."

²⁹ For a discussion of this point, see John Martin Fischer, *Control: The Metaphysics of Free Will* (Oxford: Basil Blackwell, forthcoming 1994).

³⁰ This point is discussed in greater detail in John Martin Fischer and Mark Ravizza, "When the Will Is Free," in *Ethics*, Vol. 6, *Philosophical Perspectives*, ed. James Tomberlin, (Atascadero, Calif.: Ridgeview Publishing Co., 1989): 423–451. Pace van Inwagen, Slote and Horgan, Fischer and Ravizza 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 the principle Beta (*Essay on Free will*, 94), his Final 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*, 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]: 231–243.) Other philosophers also employ similar principles. For discussions of such principles 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): 455–469; David Widerker, 'On an Argument for Incompatibilism,' *Analysis* 47 (January 1987): 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 these Beta-like principles bear some resemblance to one another, it is clear that (on the surface at least) they are not identical" ("When the Will Is Free," note 13).

³¹ I cite this argument as it is given in Fischer and Ravizza, "When the Will Is Free," 427–428.

³² Nor do I take this argument to show irrefutably that freedom to do otherwise is incompatible with determinism. For one type of objection to this argument, see John Martin Fischer, "Power Over the Past," *Pacific Philosophical Quarterly* 65 (1984): 335–350.

³³ The suggestion here is that if principles like (FP) and (FL) were translated into claims about responsibility (rather than ability), they would lose their initial intuitive plausibility.

³⁴ An incompatibilist might also respond to my arguments by arguing that although the proposed examples are counterexamples to Rule (B), they are not counterexamples to some closely related transfer principle which is still strong enough to show that responsibility is incompatible with determinism. It is beyond the scope of this essay to carefully follow out this next cycle in the dialectic; however, it seems to me that development of this objection might parallel analogous attempts to avoid the Frankfurt-type counterexample by reformulating the Principle of Alternative Possibilities.

For example, in order to avoid counterexamples involving pre-emptive overdetermination, such as "Erosion," one could adopt a transfer principle, call it "(B')," which says roughly: if no one is responsible for p, and no one is responsible for the fact that p leads to q (*because* of p), then no one is responsible for q. Such a principle would rule out a counterexample like "Erosion" for although it is true that (1) no one is responsible for the fact that glacier is eroding, and (2) no one is responsible for the fact that if the glacier erodes, then the enemy base is destroyed, it is not true that the enemy base is destroyed *because* of the erosion. (This is not true because the avalanche caused by Betty's explosives pre-empts the avalanche caused by erosion).

But although (B') avoids counterexamples involving pre-emptive overdetermination, it is not clear that it avoids the examples involving simultaneous overdetermination. For instance, in "Joint Avalanche" it is true both that (1) no one is responsible for the goat starting the avalanche, and (2) no one is responsible for the fact that if the goat starts an avalanche, then the enemy base is destroyed because of this avalanche. But it is not true that (3) no one is responsible for the destruction of the enemy base (since Betty is also partly responsible for this).

To avoid counterexamples such as “Joint Avalanche” an incompatibilist might appeal to a further reformation of (B), say “(B’), ” which says roughly: if (1) no one is responsible for p, and (2) no one is responsible for the fact that p leads to q (*only because of p*), then (3) no one is responsible for q. (B’) appears to avoid counterexamples like “Joint Avalanche” for in this case, since Betty also starts an avalanche which crushes the enemy base, condition (2) of (B’) is not satisfied: the enemy base is not crushed *only* because of the goat. The problem with (B’), however, is that it may be so restrictive that an incompatibilist could not use it to show that responsibility is incompatible with causal determinism. That is, it is not clear that our present actions are performed *only* because of the distant past and laws of nature. Rather, as defenders of the Frankfurt-type examples have argued, it seems natural to claim that agents act as they do also because they have reasons and want to act this way. For a related discussion about the senses of “because” relevant to responsibility and determinism in Frankfurt-type cases see: David Blumenfeld, “The Principle of Alternate Possibilities,” *Journal of Philosophy* 68 (June 1971): 339–345; and Robert Cummins, “Could Have Done Otherwise,” *Personalist* 60 (October 1979): 411–414. I am grateful to the members of the UCLA Law and Philosophy Discussion Group for a helpful discussion on this point.³⁵ This suggestion and the notion of the world being “sensitive” to action are developed in Fischer and Ravizza, “Responsibility and Inevitability,” and “Responsibility for Consequences.”

³⁶ Here I use Fischer’s formulation of Beta, since this formulation brings out the notion of unavoidability less ambiguously than does Van Inwagen’s version of Beta (which speaks of agents “having a choice about something”). See John Martin Fischer, “Introduction: Responsibility and Freedom,” in *Moral Responsibility*, 19.

³⁷ Van Inwagen, *An Essay on Free Will*, 187.

³⁸ *Ibid.*

³⁹ I am grateful to Matthew Carnes, Jonathan Lear, John Martin Fischer, David Schmitz, and the members of the UCLA Law and Philosophy Discussion Group for useful comments on earlier drafts of this paper.

Department of Philosophy
University of California at Riverside
Riverside, CA 92521-0201
 USA