Some Unsound Arguments for Incompatibilism

Andrew M. Bailey

Biola University

December 2005

§0. INTRODUCTION

In this paper, I contend that several arguments for the incompatibility of determinism and moral responsibility are unsound. In §1, I exposit a widely influential argument for the incompatibility of moral responsibility and determinism. The argument relies on the Principle of Alternate Possibilities, a principle famously subject to so-called Frankfurt-style counterexample. In §2, I consider a 'direct' argument for incompatibilism. But the transfer principle deployed by the argument will be subject to Frankfurt-style counterexamples too, casting doubt on the argument's validity. In §3, I argue that strengthening the transfer principle is not sufficient to avoid these counterexamples. In §4, I consider a radically different strategy: an argument for incompatibilism employing a *provably valid* transfer principle. This strategy, too, will be subject to criticism via Frankfurt-style counterexample. In doing all this, I dismantle an entire class of motivations for incompatibilism about determinism and moral responsibility.

§1. THE STANDARD ARGUMENT

Ted Warfield sketches an argument for incompatibilism with respect to moral responsibility and determinism, indicating that it has had some influence.[1] It may be put as follows:

<u>The Standard Argument</u>

P1. Determinism implies that, whatever one does, it is not the case that one could have done otherwise.
P2. One is morally responsible for what one does only if one could have done otherwise.
C1. Determinism is incompatible with moral responsibility.[2]

---

[1] O'Connor also describes the standard argument, noting that it is a "typical" motivation for incompatibilism with respect to determinism and moral responsibility. O'Connor (1993a): 345. Instances of similar lines of reasoning include Chisholm (1966): 12-14, van Inwagen (1983): 161-162, and Ginet (1995): 85-86.
[2] Warfield (1996): 219.

One might reason toward P1 with something like Peter van Inwagen's consequence argument.[3]

But in a classic article, Harry Frankfurt proposes a series of ingenious cases as counterexamples to

P2, an iteration of the Principle of Alternate Possibilities (PAP). Here is one:

<div align="center">Frankfurt-Style Case 1</div>

> Suppose someone—Black, let us say—wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide something other than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do and that he does do, what he wants him to do. Whatever Jones's initial preferences and inclinations, then, Black will have his way… Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it.[4]

It is *prima facie* plausible to think of such a case as a counterexample to PAP. It presents

us with a subject, Jones, who could not have done otherwise, but who intuitively is still morally

responsible for what he has done, precisely that state of affairs that PAP assured us would not be.

There seem to be grounds, then, to reject as false a premise of the standard argument.

§2. THE DIRECT ARGUMENT

The standard argument makes use of PAP as a premise. But suppose an argument could be

constructed toward the same end without use of PAP at all: a direct argument. Following van

Inwagen, I cast this as an argument by conditional proof: assume determinism, the thesis that

---

[3] van Inwagen (1983): 93-104. For plausible accounts of the consequence argument that (successfully, it seems to me), answer a variety of criticisms see Huemer (2000) and van Inwagen (2002).
[4] Frankfurt (1969): 149-50.

given the past and the laws of nature, there is only one possible future. From this assumption and

two modal principles, derive the conclusion that no one is responsible for anything.[5]

The argument employs a modal operator, 'N.' Np is true just in case 'p and S isn't even

partly morally responsible for the fact that p.'[6] The conclusion of the direct argument is that for

any true proposition p, given determinism, Np. van Inwagen proposes two principles in the logic

of N:

A: $\Box p \vdash Np$
B: Np & $N(p \supset q) \vdash Nq$

Letting P_o express the complete state of the world at a time before there were any humans

(as a proposition), and L, a statement of the laws of nature (as a proposition), the following

argument can be generated:

<p style="text-align:center">The First Direct Argument</p>

1. $\Box((P_o$ & $L) \supset p)$ (Assumed, consequence of determinism)
2. $\Box(P_o \supset (L \supset p))$ (1, elementary propositional and modal logic)
3. $N(P_o \supset (L \supset p))$ (2, A)
4. NP_o (Assumed, no responsibility for the distant past)
5. $N(L \supset p)$ (3, 4, B)
6. NL (Assumed, no responsibility for the laws of nature)
7. Np (5, 6, B)

Premise 1 follows trivially from our definition of determinism. Premises 4 and 6, too, seem

prima facie true. Pick a subject *S*. It is plausible to think that *S* is not even partly responsible for

the state of the world in the distant past (long before *S* was born), or for the laws of nature. Indeed,

it is plausible to think this of all human subjects. So the conclusion can be extended generally by

---

[5] This section is largely a presentation of van Inwagen's argument for the incompatibility of moral responsibility and determinism, which in turn is modeled after his original consequence argument for the incompatibility of free will and determinism. Van Inwagen (1980).
[6] Van Inwagen's reading of the N operator is unrestricted. Though mine is indexed to a subject, this difference is immaterial to our current purposes.

substituting any human subject for S and any true proposition for p.[7] If determinism is true, no one is morally responsible for anything.

The first direct argument has its virtues. First, a counterexampled PAP is not an explicit premise (as it was in the standard argument), so it's not immediately obvious that it is subject to the same worries. Second, it neatly avoids many criticisms of van Inwagen's original consequence argument for the incompatibility of free will and determinism (after which it is modeled).[8] Third, the conclusion of the first direct argument is philosophically interesting (even startling), and given determinism, threatens our conception of ourselves as morally responsible agents, as subjects, and as *persons*. This is a conclusion worth attending to. Note that the conclusion has force even for the semi-compatibilist (who agrees that determinism precludes alternative possibilities, though holding such possibilities to be irrelevant to moral responsibility).

Even if it is a deliverance of current science that exhaustive global determinism is false (because indeterminism obtains on the 'micro' level, let's say), the argument is still of interest. Suppose only that 'macro' level events are determined (eg, the neural events leading up to me typing this sentence); we might call this 'close enough' determinism. Such determinism would seem to envelop in its menacing grasp those very events we might care about the most in this context: our own actions. Whether the determinism in question is exhaustive or only close enough, it threatens something many care about.

---

[7] That is, for any p such that p is made true earlier than q. This restriction is necessary to ward off counterexamples like those David Widerker proposes to a parallel rule of inference in the consequence argument. Widerker (1987). Crisp & Warfield and O'Connor agree on something like this restriction as a solution. See O'Connor (1993b), Warfield (1996), and Crisp & Warfield (2000).

[8] Two powerful criticisms in particular (to the consequence argument) are circumvented by casting the argument in terms of moral responsibility rather than free will. First, while we might plausibly have counterfactual control over the past or the laws (casting doubt on the consequence argument's analogues to premise 4 or 6), this control is almost certainly not sufficient to ground ascriptions of moral responsibility. See Lewis (1981) and Horgan (1985). Second, a logical consequence of α and β (free will counterparts to principles A and B) is agglomeration (closure under conjunction introduction) for N. Agglomeration is *invalid* on certain free will readings of N, but not for any moral-responsibility-laden reading of N. See Johnson & McKay (1996).

The first direct argument has both an interesting conclusion and seemingly innocuous premises; it's thus clear that the heavy lifting is being done by the principles it employs: A and B. But surely no one is responsible for a logically necessary truth, so A seems as plausible principle as any. Let us turn our suspicious gaze to B, then, and inquire whether the first direct argument is valid.

Recall what B claims: if S is not even partly morally responsible for the fact that p, and S is not even partly morally responsible for the fact that $p \supset q$, then S is not even partly morally responsible for the fact that q. The principle has some plausibility, as the following example is designed to show. Let p be 'Hurricane Katrina ravages the gulf coast', and q, 'Countless lives are lost.' On the supposition that Np and N($p \supset q$), is it not reasonable to infer Nq? That is, someone S isn't even partly morally responsible for the fact that p, and S isn't even partly morally responsible for the fact that $p \supset q$. It seems obvious *on these grounds alone* to conclude that S isn't even partly morally responsible for q either. This is principle B in action.

But Frankfurt-style counterexamples are near at hand:

<u>Frankfurt-style Case 2</u>

[Green]… walks along a beach and, noting that there is a child drowning, dives into the water and rescues the child. Though Green has had a device implanted in his brain, the device does not play any role in Green's decision to save the child (and his subsequent action). That is, the device monitors Green's brain activity but does not actually intervene in it. Let us suppose that this is because the scientists can see that Green is about to decide to save the child and to act accordingly. But let's also suppose that the scientists would have intervened to bring about a decision to save the child if Green had shown an inclination to decide to refrain from saving the child. That is, were Green inclined to decide on his own not to save the child, the scientists would ensure electronically that he decide to save the child and also that he act to carry out this decision.[9]

John Martin Fischer notes of the case:

---

[9] Fischer (1986): 41.

> ... Green is not morally responsible for the fact that the scientists are ready to intervene, and he is not responsible for the fact that, if they are so ready, he will save the child. But he does seem to be morally responsible for saving the child.[10]

So a case bearing all the hallmarks of Frankfurt-style counterexamples to PAP spells trouble for B, and hence for the first direct argument.[11] It points us to a scenario in which some subject S isn't even partly morally responsible for the fact that p (scientists are ready to intervene), or for $p \supset q$ (if the scientists are ready to intervene, the child is saved), but S is at least partly responsible for the fact that q (the child is saved). B as stated is false. The incompatibilist, then, must rework B, or develop a sophisticated response the Frankfurt-style counterexample strategy.

§3. ANOTHER DIRECT ARGUMENT

Suppose we were to logically strengthen principle B as employed by the first direct argument. Might Frankfurt-style counterexamples then be set aside? Warfield has thought as much, advancing another direct argument for the incompatibility of moral responsibility and determinism. His makes use of this principle:

Beta $\square$: Np & $\square$($p \supset q$) $\vdash$ Nq

No argument for Beta $\square$ is offered by Warfield, though it seems as intuitively plausible as the original B. That is, it is not hard to cook up a variety of cases in which the Beta $\square$ inference holds.

Warfield's direct argument is the very model of simplicity and elegance. 'N,' 'P$_o$,' and 'L' remain defined as above.[12] The assumptions of the argument will remain nearly the same as before; we need only add that S isn't even partly morally responsible for the *conjunction* of the

---

[10] Fischer (1986): 60-61.

[11] Fischer (1986) presents the earliest application of Frankfurt-style counterexamples to transfer principles like B (so far as I know). For further defense of Frankfurt-style cases as defeaters to B, see Fischer & Ravizza (1998): 151-169 and Fischer (2004). Ravizza (1994) employs a similar case to undermine B involving *overdetermination* (rather than counterfactual intervention, as in standard Frankfurt-style cases).

[12] Following van Inwagen, Warfield gives an unrestricted reading of 'N,' in contrast with the subject-indexed version I have offered, though as before, the difference is not material to our current purposes.

laws of nature and some state of the world in the distant past. This is formally stronger than the

premises of the first direct argument (though equally plausible).

<div align="center">The Second Direct Argument</div>

1. $\Box((P_o \ \& \ L) \supset p)$	(Assumed, consequence of determinism)
2. $N(P_o \ \& \ L)$	(Assumed, no responsibility for the distant past and the laws of nature)
3. Np	(1, 2, Beta $\Box$)

Where any true proposition can be substituted for p, and any subject for S, the same result

is generated: Np.

Warfield assures his reader that no Frankfurt-style case can be contrived to unseat Beta $\Box$.

Such a case must have at least these features: it must involve a genuine and intuitive instance of

moral responsibility for S for some proposition q; it must involve the logical entailment of q by

some proposition p, and it must be that S isn't even partly morally responsible for the fact that p.

It's suggested that these very features preclude the use of a Frankfurt-style case as a

counterexample to Beta $\Box$, for judging that S is responsible for q is nothing more than judging that

determinism and moral responsibility are compatible. As he notes:

> It is hardly of interest to point out that the assumption of the compatibility of determinism
> and moral responsibility implies that Rule Beta $\Box$ is invalid. As van Inwagen pointed out
> some time ago, one who hopes to produce an interest refutation of incompatibilist transfer
> of nonresponsibility principles must produce cases which do not presuppose the
> compatibility of determinism and moral responsibility.[13]

This is surely correct, but I will here argue that such an objectionable presupposition of

compatibilism is not in fact necessary. The compatibilist about determinism and moral

responsibility affirms something like the:[14]

Compatibility Thesis (CT): It is possible that no one is even partly morally responsible for

---

[13] Warfield (1996): 223.

[14] I here ignore some minority positions: multiple-past compatibilism or small miracle compatibilism (*a la* David Lewis), and an (uninstanced, so far as I know, but possible) iteration of compatibilism which denies the validity of $(P_o \ \& \ NL) \vdash N(P_o \ \& \ L)$.

the conjunction of the past and the laws and no one is even partly morally responsible for the fact that the conjunction of the past and the laws entails every true proposition p, *but* someone might still be (at least partly) morally responsible for the fact that p.

Van Inwagen and Warfield are right that it is dialectically unseemly to assume CT in a refutation of the second direct argument or its transfer principle. But CT is not needed to generate a Frankfurt-style counterexample to Beta □, only:

> Weak Compatibility Thesis (WCT): It is possible that some S isn't even partly morally responsible for the fact that p and S isn't even partly morally responsible for the fact that p entails q *but* S might still be (at least partly) morally responsible for the fact that q.

CT and WCT are logically distinct (though CT implies WCT). CT is true iff there is a globally deterministic world in which someone is morally responsible. But even on the supposition that there are no such worlds, WCT might still be true.

To be dialectically above-board we would be wise to assume neither the truth of CT nor the denial of WCT. But can we concoct a plausible Frankfurt-style case that doesn't assume CT—a case *without* global determinism that still meets the conditions of Beta □? I think so. Consider the following:

Frankfurt-style Case 3

Jethro has never been all that successful as a graduate student in philosophy. His peers and professors do not appreciate him, and he will soon loose his stipend, his place in the department life, and his already-small office space—that is, unless he captures an 'A' in at least one of his courses. Jethro's recently deceased brother, Joel, has left behind a stack of philosophy papers. Coincidentally, one of these papers is a perfect match for Jethro's Kripke seminar, brilliantly argued, and bursting with original insight and rigor. Jethro carefully (and indeterministically) deliberates, freely decides to submit, and in fact submits Joel's paper as his own.

Unbeknownst to Jethro, his faculty advisor is worried about his future. She also happens to be a leading neurologist skilled in the construction of (very) tiny robots. Out of concern for Jethro, she has recently dropped a nanobot into his coffee. The nanobot worms its way into Jethro's brain and monitors Jethro's deliberation. Should Jethro begin to form a decision other than the decision to use Joel's paper, the nanobot would cause Jethro to form the decision to submit Joel's paper. As it happens, Jethro chooses to use Joel's paper. So in the actual sequence of deliberation, the nanobot made no difference; Jethro has chosen what to do, and all on his own.

We have in Frankfurt-style Case 3 a counterfactual intervener who makes no difference in the actual sequence, a hapless subject, and an exclusion of alternate possibilities for that subject—all the classic ingredients of Frankfurt-style cases. Jethro is morally responsible for submitting Joel's paper as his own; that he could not have done otherwise (due to the nanobot) is immaterial. Now a detail: let the conjunction of the nanobot's being placed in Jethro's brain by his advisor and the laws of nature (propositionally stated) be p. Let q be 'Jethro submits Joel's paper as his own.' Finally, suppose that p entails q; in all the worlds where the actual laws obtain and the nanobot is placed in Jethro's brain, Jethro submits Joel's paper as his own.

Frankfurt-style case 3 is a counterexample to Beta $\square$; some subject (Jethro) isn't even partly morally responsible for the fact that p, nor is the subject even partly responsible for the fact that p entails q. And yet Jethro is still at least partly morally responsible for q.

Note several things. First, we have not packed any global determinism into the case. Jethro dwells in an *indeterministic* world, in fact, and deliberates indeterministically in the actual sequence. Warfield is mistaken, then, in thinking that any counterexample to Beta $\square$ must involve the conjunction of global determinism and some subject's being morally responsible for some fact (that is, compatibilism). Second, absent any special assumptions, it's hard to see how merely adding one *pocket* of determinism (as we have done) would alter our intuitive judgment of about Jethro's moral responsibility. Thus, Warfield's claim that "Frankfurt cases cannot be used to refute my argument."[15] is false. In Frankfurt-style Case 3, we have a counterexample to Beta $\square$, the principle on which Warfield's argument relies.

§4. A VALID DIRECT ARGUMENT

The first and second direct arguments have made use of a common strategy: first provide

---

[15] Warfield (1996): 222.

intuitive grounds for some modal principle concerning the transfer of non-responsibility across an implication or entailment, and then employ this principle in an argument for the incompatibility of determinism and moral responsibility.

But there is another way I now consider: use some uncontroversial rules of modal and counterfactual logic to derive a modal principle concerning the transfer of non-responsibility across an implication or entailment, and then use this principle as a premise in an argument for the incompatibility of determinism and moral responsibility.[16]

This strategy has advantages. First, it provides support for a transfer principle independent of intuition pumps. Those giving little credence to the deliverances of intuition will find this appealing. Second, it sets high success standards for objections to the principle. If one wishes to disagree with the derived principle (and the derivation is valid), she must disagree with some uncontroversial rues of modal and counterfactual logic. But such a move would in most cases be contrived and unmotivated.

To derive a modal principle concerning the transfer of nonresponsibility using only uncontroversial rules of modal and counterfactual logic, we must first analyze moral responsibility in terms of modal and counterfactual terms. I now give such an analysis. Though I do not claim it is the ultimately correct analysis of moral responsibility, it will be close enough to suggest how this final incompatibilist strategy plays out.

A few preliminaries. I will assume that our concept of moral responsibility allows us to speak of someone's being morally responsible for some fact (or the truth of some proposition). Whatever it is that we are *primarily* responsible for, I take it that the concept can also be applied to

---

[16] For the central idea of this section I am entirely indebted to Tom Crisp. Huemer uses a similar maneuver (employing a provably valid transfer principle to dodge counterexamples) in his defense of the consequence argument for the incompatibilism of free will and determinism. Huemer (2000): 539-540. van Inwagen defends a logically equivalent position (though less explicitly and for different reasons). van Inwagen (2002): 167.

facts or propositions (this is, after all, the language we have already used in the first and second direct arguments).

I also assume that acts are things to which duties, obligations, and whatnot can attach, and that these duties, obligations, and whatnot have something to do with moral responsibility. There are other kinds of responsibility, and they may not involve duties, obligations, and whatnot—but here I am not concerned with them.

Let 'O' be the predicate '*S* should or should not perform act *a*.' So O*a* is true iff *a* is an action to which duties are attached for *S*, whether to perform *a* or to refrain from performing *a*. I leave open the question of what conditions are necessary and sufficient for a subject to have such a duty to act (or refrain from acting). Further, let 'P' be the predicate '*S* performs act *a*.'

The analysis:

MR: *S* is at least partly morally responsible for the fact that *p* if and only if *p* & there is some action *a* (such that *S* could perform *a*) & O*a* & if it were the case that P*a*, it might have been the case that not-*p*.

Two questions emerge. Does MR plausibly capture into its analysis the concept of moral responsibility, and is B or a B–like rule provably valid on a reading of the N operator that employs the analysis?

On the first question. It is plausible to think that a necessary condition of *S*'s being morally responsible for *p* is there being an action to which are attached some moral duties. It is *moral* responsibility that is under the microscope, after all, and not legal, financial, causal responsibility, or whatnot—and moral responsibility is linked in some way to what we should or should not do. For a subject's actions to which are attached no duties, there can be no *moral* responsibility.

Perhaps more contentious, though still quite ecumenical, is the 'such that *S* could perform *a*' restriction. *S* is not morally responsible for some fact unless *S* *can do* something (leave 'can' unanalyzed). This is so because *S* is not morally responsible for some fact unless *S* has some moral

duty attached to some action, and it follows from the attachment of a moral duty to an action for S that *S can* perform that action. It follows, that is, on the assumption of a weak ought-implies-can principle.[17]

But does MR capture the necessary and *sufficient* conditions of someone's being morally responsible for some fact? This is difficult to confirm, but examining a few paradigmatic cases will prove illuminating.

Lilith and Jane, Part One

Lilith duly deliberates and decides to murder her twin sister Jane (Lilith has always resented Jane). Lilith buys a gun, points it at Jane, and pulls the trigger. Lilith didn't *have* to do this, but she chose to anyways. Jane dies as a result of the gunshot wound.

Given only these facts, it seems that Lilith is at least partly morally responsible for the proposition 'Jane is dead.' And she is morally responsible in a particular way, deserving *blame* (at least partly) for the fact that Jane died. Now see that Lilith has met the conditions outlined in MR. It is true that 'Jane is dead,' there is some action such that Lilith could have performed (putting down the gun without pulling the trigger), this is an action to which there are attached moral duties or obligations for Lilith (she has a duty or obligation to put down the gun without pulling the trigger), and if she were to have performed that action, it might not have been the case that Jane is dead.

Lilith and Jane Part One is an instance of blame, a paradigmatic facet of moral responsibility. We are inclined to assign blame in this instance, and it is an instance where the conditions of MR obtain. Consider now an instance of praise-worthy action, another paradigmatic facet of moral responsibility.

---

[17] To say much more on this point would be to stray from my project. Readers interesting in the growing literature on PAP and ought-implies can principles may consult Widerker (1991), Schnall (2001), Copp (2003), and Speak (2005).

Lilith and Jane, Part Two

Lilith has always resented her twin sister Jane for entirely arbitrary and irrational reasons. Lilith knows that these reasons are arbitrary and irrational. But one day, Lilith decides (and she didn't *have* to decide this) to open up and confess to Jane the resentment she has actively harbored. Lilith carries out this intention, and over a four-hour-long, soul-searching conversation, the bad blood between them evaporates. The two are reconciled.

Knowing only these facts, it seems that Lilith is at least partly morally responsible for the proposition 'Jane and Lilith are reconciled.' More specifically, Lilith deserves *praise* (at least partly) for the fact that the two are friends again. Note that in Lilith and Jane Part Two, the conditions of MR are again met. It is true that 'Jane and Lilith are reconciled,' there is some action such that Lilith could have performed (continuing to actively harbor her irrational resentment of Jane), this is an action to which there are attached moral duties or obligations for Lilith (she has a duty or obligation to refrain from actively harboring her irrational resentment of Jane), and if she were to have perform the action, it might not have been the case that Jane and Lilith are reconciled.

I take it for granted that similar cases of supererogatory and suberogatory acts (if there are such things) involving moral responsibility can be generated, and can be successfully accounted for by MR. It is a consequence of MR that no one is morally responsible when acting 'neutrally.'[18] Someone must abide by or flaunt her duties to be eligible for ascriptions of moral responsibility under MR—and this is fitting, I think, given that it is *moral* responsibility we are concerned with (and not some other kind responsibility or responsibility *simpliciter*).

MR has some degree of initial plausibility as an analysis of what it is for someone to be morally responsible for a fact; that it fits a wide spectrum of cases nicely is evidence of this. I will

---

[18] This is somewhat controversial; some accounts allow for moral responsibility over morally *neutral* acts. See, for example, Fischer (1994): 165. I confess that this doesn't make much sense to me; it's hard to see how the responsibility in question could be of the *moral* variety if it didn't somehow involve duties, obligations, or something like them.

return to the success of MR as an analysis of moral responsibility shortly, but first let's see how

MR might be an ingredient in a provably valid transfer principle and thus inform an argument for

the incompatibility of determinism and moral responsibility.

Let Np still read 'p and S isn't even partly morally responsible for the fact that p.' Let us

stipulate the truth conditions of Np more carefully, however, taking into account the analysis of

moral responsibility given above: Np is true iff p & there is no action a (such that S could perform

a) & Oa, & if it were the case that Pa, it might be the case that not-p. We may symbolize the truth

conditions for Np as (where the quantification is over those actions that S can perform): $Np \equiv p$ &

$\sim\exists a(Oa$ & $(Pa \diamondsuit\!\!\rightarrow \sim p))$. Equivalently, $Np \equiv p$ & $\forall a(Oa \supset (Pa \:\square\!\!\rightarrow p))$.

Given a few uncontroversial rules of modal and counterfactual logic, this reading of N

suggests a provably valid iteration of the B transfer principle:

B∗: Np & Nq & $\square((p\&q) \supset r)$ $\;\vdash\;$ Nr

I supply a proof of B∗ in the Appendix. Let us now use B∗ in an argument for the incompatibility

of determinism and moral responsibility, using all terms as we have already defined them above:

<u>The Third Direct Argument</u>

1. $\square((P_o$ & L$) \supset p)$	(Assumed, consequence of determinism)
2. NP_o	(Assumed, no responsibility for the distant past)
3. NL	(Assumed, no responsibility for the laws)
4. Np	(1, 2, 3, B∗)

Note several things. As before, we may substitute any true proposition for p, and any (human)

subject for S; the generalized conclusion is the same as previous direct arguments, that no one is

morally responsible for anything. Furthermore, the transfer principle we have employed (B∗) is

provably valid; B∗ is a logical consequence of some rather innocuous assumptions about modal

and counterfactual logic. It is thus not subject to direct counterexample in the way that B and Beta

$\square$ were.

A principled disagreement with the validity of B∗ requires a principled disagreement with the modal and counterfactual rules that entail it, and such a project does not seem promising. Have we arrived, then, at a sound argument for the incompatibility of moral responsibility and determinism?

Perhaps not. Recall that B∗ makes use of the N operator, which in turn packs in the analysis of moral responsibility offered by MR. This suggests a response strategy to the third direct argument: to undermine MR as a successful analysis of moral responsibility via counterexample. Counterexamples to MR will not undermine the validity of B∗. They will suggest, however, that B∗ is a valid rule having little or *nothing to do* with moral responsibility. This response, if sound, would be sufficiently motivated (unlike a denial of Rule M, say) and would effectively neuter the third direct argument of its force.

Frankfurt-style cases are sufficient to accomplish this task. Recall Frankfurt-style Case 2. It directed our imagination to a scenario in which Green wasn't even partly morally responsible for the fact that p (scientists are ready to intervene), or for p ⊃ q (if the scientists are ready to intervene, the child is saved), but Green was at least partly responsible for the fact that q (the child is saved). The analysis of moral responsibility we have considered claims that:

> MR: S is at least partly morally responsible for the fact that p if and only if p & there is some action a (such that S could perform a) & Oa & if it were the case that Pa, it might have been the case that not-p.

But in Frankfurt-style Case 2, there *wasn't* an action that Green could perform such that if Green had performed it, it might have been the case that the child was not saved. No matter what Green did, the child would have been saved, and yet Green was *still* morally responsible for the fact that the child was saved. If this is all correct, MR fails as an analysis of what it is for someone to be morally responsible for some fact.

The compatibilist is then free to use this conclusion to undermine the third direct argument,

pointing out:

> Sure, your argument is valid (unlike the first and second direct arguments), but it has nothing to do with moral responsibility. The only thing that could give the third direct argument any teeth is by connecting it to a concept we care deeply about, like our personhood or sense of moral responsibility. But as Frankfurt-style cases show, the notion employed by the third direct argument (whatever it is) *isn't* coextensive with moral responsibility. The conclusion that 'no one is morally responsible' for anything isn't interesting anymore, given that this kind of 'moral responsibility' isn't the same as the one we employ in contemporary free will debates (and in everyday discourse).

Frankfurt-style cases have once again spelled trouble for arguments for the incompatibility of determinism and moral responsibility. They have presented prima facie counterexamples to the transfer principles employed by the arguments (B and Beta □) or to the analysis of moral responsibility required to give the argument its force (MR, as in the third direct argument).

To conclude. In this paper, I have shown the central role that Frankfurt-style cases play in several arguments for incompatibilism with respect to determinism and moral responsibility. I have thus suggested a plausible strategy to undermine and entire set of arguments for incompatibilism: to present and defend a Frankfurt-style case as a counterexample to the relevant transfer principle or analysis of moral responsibility.[19] For those interested in preserving moral responsibility in the face of determinism this is a pleasant prospect.[20]

---

[19] Due to space limitations, I do not interact with the ever-growing literature on the success of Frankfurt-style cases, though highly nuanced response strategies and cases have emerged in recent years. Widerker and McKenna's edited volume (2003) contains many contributions to this fascinating debate. For what are perhaps the most sophisticated Frankfurt-style cases on the market, see Mele & Robb (1998), Hunt (2000), and Hunt (2005). For discussion of some dialectical issues, see Haji & McKenna (2004).

[20] For this paper, I owe much to Tom Crisp; any virtues it may have are the result of conversations I have had with him. Any errors are, of course, my own.

APPENDIX: A PROVABLY VALID TRANSFER PRINCIPLE

N as defined in §4 makes use of a subjunctive conditional (counterfactual), expressed by the "if… were… then… it would be the case that…" locution. I thus introduce two rules of inference governing counterfactuals. On David Lewis' and Robert Stalnaker's counterfactual semantics, both are valid.

Rule∗: $p \:\square\!\!\rightarrow q \:\&\: p \:\square\!\!\rightarrow r \vdash p \:\square\!\!\rightarrow (q \:\&\: r)$
Rule∗∗: $p \:\square\!\!\rightarrow q \:\&\: \square(q \supset r) \vdash p \:\square\!\!\rightarrow r$

I briefly indicate why Rule∗ and Rule∗∗ are valid on Lewis' counterfactual semantics. For Lewis, $p \:\square\!\!\rightarrow q$ is non-vacuously true at W iff there is a $p \:\&\: q$ world closer to W than any $p \:\&\: \sim q$ world.[21]
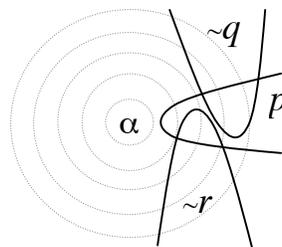
Proof-sketch for Rule*. Assume both $p \:\square\!\!\rightarrow q$ and $p \:\square\!\!\rightarrow r$. That is, there is a $p \:\&\: q$ world closer to the actual world than any $p \:\&\: \sim q$ world, and there is a $p \:\&\: r$ world closer to the actual world than any $p \:\&\: \sim r$ world. Consider now the (relevant) possible arrangements of the nearest $\sim q$ and $\sim r$ worlds (if there are any such worlds). We may diagram the three possible arrangements as follows, and on each, $p \:\square\!\!\rightarrow (q \:\&\: r)$ will be true:[22]
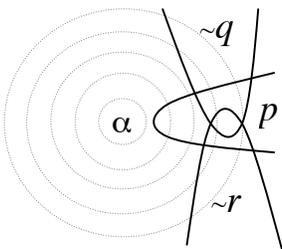
First, the nearest $p \:\&\: \sim q$ world could be equally near the actual world as the nearest $p \:\&\: \sim r$ world. But on this assumption, there is a still nearer $p \:\&\: (q \:\&\: r)$ world (somewhere in the logical space between the nearest $\sim q/\sim r$ worlds and the actual world), and thus a nearer $p \:\&\: (q \:\&\: r)$ world than any $p \:\&\: \sim(q \:\&\: r)$ world.
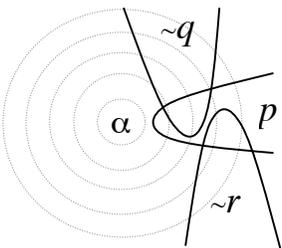
Second, the nearest $p \:\&\: \sim q$ world could be closer than the nearest $p \:\&\: \sim r$ world. But on this assumption, there is a still nearer $p \:\&\: (q \:\&\: r)$ world (somewhere in the logical space between the nearest $\sim q$ worlds and the actual world), and thus a nearer $p \:\&\: (q \:\&\: r)$ world than any $p \:\&\: \sim(q \:\&\: r)$ world.

Third, the nearest $p \:\&\: \sim r$ world could be closer than the nearest $p \:\&\: \sim q$ world. But on this assumption too, there is a still nearer $p \:\&\: (q \:\&\: r)$ world (somewhere in the logical space between the nearest $\sim r$ worlds and the actual world), and thus a nearer $p \:\&\: (q \:\&\: r)$ world than any $p \:\&\: \sim(q \:\&\: r)$ world.

Proof-sketch for Rule∗∗. There is a $p \:\&\: q$ world closer to the actual world than any $p \:\&\: \sim q$ world, and all q worlds are r worlds. So, there is a $p \:\&\: r$ world closer to the actual world than any $p \:\&\: \sim r$ world.

---

[21] See Lewis (1973): 13-21.

[22] α is the actual world, the dotted circles spheres around it, and the solid-lined shapes regions of logical space (sets of possible worlds) in which the noted proposition is true.

I shall also make use of 'Rule M' from the weak modal logic T: if it is necessary that p, then p. The version of B that I now derive with the help of Rule∗, Rule∗∗, and Rule M:

> B∗: Np & Nq & $\Box((p\&q) \supset r)$ $\vdash$ Nr

For the proof. The domain of discourse I quantify over is those actions such that S can perform them (leaving 'can' unanalyzed), while 'P' and 'O' remain as defined in §4.

<div align="center">

Proof for B∗

</div>

1.	Np	(assumed)
2.	Nq	(assumed)
3.	$\Box((p \& q) \supset r)$	(assumed)
4.	p & $\forall a(Oa \supset Pa \,\Box\!\!\rightarrow\! p)$	(1), definition of N
5.	q & $\forall a(Oa \supset Pa \,\Box\!\!\rightarrow\! q)$	(2), definition of N
6.	$\forall a(Oa \supset Pa \,\Box\!\!\rightarrow\! p)$	(4), conjunction elimination
7.	$\forall a(Oa \supset Pa \,\Box\!\!\rightarrow\! q)$	(5), conjunction elimination
8.	$Oa \supset Pa \,\Box\!\!\rightarrow\! p$	(6), universal instantiation
9.	$Oa \supset Pa \,\Box\!\!\rightarrow\! q$	(7), universal instantiation
10.	Oa	(assumed for conditional proof)
11.	$Pa \,\Box\!\!\rightarrow\! p$	(8), (10) modus ponens
12.	$Pa \,\Box\!\!\rightarrow\! q$	(9), (10) modus ponens
13.	$Pa \,\Box\!\!\rightarrow\! (p \& q)$	(11), (12), Rule∗
14.	$Pa \,\Box\!\!\rightarrow\! r$	(3), (13), Rule∗∗
15.	$Oa \supset Pa \,\Box\!\!\rightarrow\! r$	(14), conditional proof
16.	$\forall a(Oa \supset Pa \,\Box\!\!\rightarrow\! r)$	(15), universal generalization
17.	p	(4), conjunction elimination
18.	q	(5), conjunction elimination
19.	p & q	(17), (18), conjunction introduction
20.	$(p \& q) \supset r$	(3), Rule M
21.	r	(19), (20), modus ponens
22.	r & $\forall a(Oa \supset Pa \,\Box\!\!\rightarrow\! r)$	(21), (16), conjunction introduction
23.	Nr	(22), definition of N

REFERENCES

Chisholm, Roderick. 1966. "Freedom and Action." in *Freedom and Determinism*, Keith Lehrer (ed.): 11-44. New York: Random House.

Copp, David. 2003. "'Ought' Implies 'Can,' Blameworthiness, and the Principle of Alternate Possibilities." in *Moral Responsibility and Alternative Possibilities*, David Widerker & Michael McKenna (eds.): 265–99. Aldershot: Ashgate Press.

Crisp, Thomas M. & Ted A. Warfield. 2000. "The Irrelevance of Indeterministic Counterexamples to Principle Beta." *Philosophy and Phenomenological Research*. Vol. 61, No. 1: 173-84.

Fischer, John Martin. 1986. "Introduction: Responsibility and Freedom." in *Moral Responsibility* John Martin Fischer (ed.). Ithaca and London: Cornell University Press.

_____. 1994. *The Metaphysics of Free Will*. Oxford: Blackwell.

_____. 2002. "Frankfurt-Type Examples and Semi-Compatibilism," in *The Oxford Handbook of Free Will*. Robert Kane (ed.): 281–308. Oxford: Oxford University Press.

_____. 2004. "The Transfer of Nonresponsibility." in *Freedom and Determinism*, Michael O'Rourke & David Shier (eds.). Cambridge, MA: MIT Press.

Fischer, John Martin & Mark Ravizza. 1998. *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, Harry. 1969. "Alternative Possibilities and Moral Responsibility." *Journal of Philosophy* 66: 829–39.

Ginet, Carl. 1995. "Freedom, Responsibility, and Agency." *The Journal of Ethics* 1: 85-98.

Haji, Ishtiyaque & Michael McKenna. 2004. "Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities," *Journal of Philosophy* 101: 299–314.

Horgan, Terence. 1985. "Compatibilism and the Consequence Argument." *Philosophical Studies* 47: 339-56.

Huemer, Michael. 2000. "van Inwagen's Consequence Argument." *Philosophical Review* 109: 525-544.

Hunt, David. 2000. "Moral Responsibility and Unavoidable Action." *Philosophical Studies* 97: 195–227.

_____. 2005. "Moral Responsibility and Buffered Alternatives." *Midwest Studies in Philosophy* 29: 126-145.

Lewis, David. 1973. *Counterfactuals*. Oxford: Basil Blackwell.

_____. 1981. "Are We Free to Break the Laws?" *Theoria* 47: 113-21.

Mele, Alfred & David Robb. 1998. "Rescuing Frankfurt-Style Cases." *Philosophical Review* 107: 97–112.

McKay, Thomas & David Johnson. 1996. "A Reconsideration of an Argument Against Compatibilism." *Philosophical Topics* 24, 113-22.

O'Connor, Timothy. 1993a. "Alternative Possibilities and Moral Responsibility." *Southern Journal of Philosophy*, 31: 345-372.

_____. 1993b. "On the Transfer of Necessity." *Nous* 27: 204-18.

Ravizza, Mark. 1994. "Semi-Compatiblism and the Transfer of Nonresponsibility." *Philosophical Studies* 75: 61-93.

Schnall, Ira. 2001."The Principle of Alternate Possibilities and 'Ought' Implies 'Can.'" *Analysis* 61: 335–40.

Speak, Daniel. 2005. "PAPistry: Another Defense." *Midwest Studies in Philosophy* 29: 262-268.

Van Inwagen, Peter. 1980. "The Incompatibilisty of Responsibility and Determinism." M. Bradie and M. Brand (eds.) *Bowling Green Studies in Applied Philosophy*, Vol. 2. Bowling Green, Ohio: Bowling Green State University Press, 1980: 30-37. Also reprinted in Fischer (1986).

_____, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

_____, Peter. 2002. "Free Will Remains a Mystery." in *The Oxford Handbook of Free Will*, Robert Kane (ed.): 281–308. Oxford: Oxford University Press.

Warfield, Ted. 1996. "Determinism and Moral Responsibility are Incompatible." *Philosophical Topics*, Vol. 24, No. 2: 215-226.

Widerker, David. 1991. "Frankfurt on 'Ought' Implies 'Can' and Alternative Possibilities."*Analysis* 49: 222–24.

Widerker, David & Michael McKenna. (eds.) 2003. *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate Press.