

In Defense of the Direct Argument for Incompatibilism

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Paul Roger Turner
May 2014

To Anna

ACKNOWLEDGMENTS

I would like to thank, initially, my dissertation committee members, E. J. Coffman, David Palmer, Markus Kohl, John Martin Fischer, and Jeffrey Kovac. Their time and invaluable input have made this project what it is, and I am certain that without their help I would have never gotten this thing off the ground floor. Special thanks go to E. J. Coffman, my director, and David Palmer (my *de facto* co-director!) for providing thorough and penetrating comments on at least two full drafts of the dissertation, and, perhaps more importantly, for their friendship and general mentorship throughout my time in the program.

Secondly, I would like to thank my wife, Anna, for, in an important sense, sacrificing the last seven years of her life in order to support me as I pursued this degree. She has patiently, and graciously, taken on the burden of pulling double duty: providing for our family, as well as managing the home. I only hope that the way the rest of our lives plays out will make it so that her sacrifice will have been worth it.

Finally, but preeminently, I must thank Jesus, who is called Christ. He is my Master; all that I have, and all that I do is for Him.

ABSTRACT

Is moral responsibility compatible with the truth of causal determinism? One of the most influential arguments that moral responsibility is *incompatible* with causal determinism is the so-called ‘Direct Argument,’ developed by Peter van Inwagen in his *An Essay on Free Will*.

Informally put, the Direct Argument goes as follows:

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But we are not responsible for what went on before we were born, and neither are we responsible for what the laws of nature are. Therefore, we are not responsible for the consequences of these things (including our present acts).

The Direct Argument is highly significant. If it is successful, we have an argument for incompatibilism about responsibility and determinism that does not make use of two controversial claims typically invoked by incompatibilists: (i) a person is morally responsible for what she has done only if she could have done otherwise, and (ii) if the person’s action is causally determined, then she could *not* have done otherwise. Since compatibilists typically deny one or the other of these claims, the Direct Argument offers an intriguing way to argue for incompatibilism about responsibility and determinism that sidesteps many of the traditional battlegrounds between compatibilists and incompatibilists.

The Direct Argument relies on two rules of inference, both of which have been questioned by the Argument’s opponents. In my dissertation, I defend the Direct Argument from some of the most pressing recent attacks against these rules. But, there is a further objection, an objection called the No Past Objection, that I argue successfully undermines the Direct Argument. So, I go on to revise the Direct Argument in light of the No Past Objection, and I do so in a way gets around this objection without sacrificing the Argument’s inference rules, or the

spirit of its metaphysical assumptions. The result, what I call the Direct Argument*, is a successful argument for incompatibilism about moral responsibility and causal determinism.

TABLE OF CONTENTS

CHAPTER ONE: Introduction.....	1
II. Why this is Important.....	5
III. In it to Win it.....	6
CHAPTER TWO: Criticisms of Rule B (I): Alleged Counterexamples.....	9
II. Fischer and Ravizza’s Alleged Counterexample to Rule B.....	10
III. Widerker’s (2002) Attempt at a Counterexample to Rule B.....	19
IV. Haji’s (2008) Attempt at a Counterexample to Rule B.....	22
V. Haji’s (2010) Attempt at a Counterexample to Rule B.....	31
VI. Shabo’s Attempt to Provide a Counterexample to (Logical Versions of) Rule B.....	46
VII. Schnall and Widerker’s Attempt to Provide a Counterexample to Rule B.....	54
VIII. Conclusions.....	57
CHAPTER THREE: Criticisms of Rule B (II): Dialectical Objections.....	59
II. Fischer and Dialectical Stalemates.....	60
III. McKenna’s Attempt to Say ‘Goodbye’ to the Direct Argument.....	70
IV. Shabo and the Fate of the Direct Argument.....	76
V. Conclusions.....	84
CHAPTER FOUR: An Objection to Rule A.....	87
II. Kearns’s First Alleged Counterexample to Rule A: Murder!.....	88
III. Kearns’s Second Alleged Counterexample to Rule A: Hey Jude.....	94
IV. Kearns’s Third Alleged Counterexample to Rule A: Torturing Babies.....	99
V. Conclusions.....	102

CHAPTER FIVE: The No Past Objection and a Revised Direct Argument: The Direct Argument*	104
II. The No Past Objection.....	106
III. The Direct Argument*.....	108
IV. Objections, Round 1: A Potential Counterexample to (1) of the Direct Argument*.....	115
V. Objections, Round 2: <i>Ex Nihilo Nihil Fit?</i>	117
VI. Conclusions.....	121
CHAPTER SIX: Looking Forward: Ramifications of a Successful Direct Argument*.....	124
II. Counterexamples to Rule B Are Impossible.....	126
III. Is Truth Dependence _{MORAL} True?: Objections to TDM.....	133
IV. Incompatibilism is True, and What This Might Say About Our World.....	141
V. Conclusions.....	154
REFERENCES.....	159
VITA.....	164

CHAPTER ONE

In Defense of the Direct Argument for Incompatibilism

Is moral responsibility compatible with the truth of causal determinism?¹ One of the most influential arguments that moral responsibility is *incompatible* with causal determinism is the so-called ‘Direct Argument,’ developed by Peter van Inwagen (1983). The Direct Argument rests on the following two rules of inference (where, ‘ $\Box$ ’ stands for broadly logical necessity; ‘ $\supset$ ’ stands for material implication; and ‘NRp’ stands for ‘p and no-one is now or ever has been even partly morally responsible for p’):

Rule A: From $\Box p$, we may infer NRp

Rule B: From NRp and NR($p \supset q$), we may infer NRq

To illustrate the Direct Argument, consider an individual, Colin, who decides to donate some money to charity. Van Inwagen argues that, with these two rules of inference in hand and two very plausible premises, we can show that *if* Colin’s decision to donate to charity is causally determined, then it’s *not* something for which he can be morally responsible. Here are the details of his argument. Assume, for conditional proof, that causal determinism is true. From this assumption, we can reason as follows (where ‘C’ stands for Colin’s decision to donate money, ‘P’ labels a complete description of the world prior to the existence of any human person, and ‘L’ stands for a conjunction of the laws of nature):

(1) $\Box (P \ \& \ L \supset C)$ By definition of ‘determinism’

¹ Here, and throughout, I’ll understand these terms to mean the following:

Moral responsibility: Praiseworthiness or blameworthiness for some morally significant action, A.

Causal Determinism: The thesis that there is at any instant exactly one physically possible future. (van Inwagen, 1983, p. 3)

(2) $\square (P \supset (L \supset C))$	1, and logic
(3) NR $(P \supset (L \supset C))$	2, and Rule A
(4) NR P	Premise
(5) NR $(L \supset C)$	From 3, 4, and Rule B
(6) NR L	Premise
(7) NR C	From 5, 6, and Rule B

In other words, if Colin's decision to donate to charity is causally determined, then the past and the laws of nature jointly entail Colin's decision at that time. But since Colin is not morally responsible for the past prior to the existence of any human person and since he is not morally responsible for the laws of nature, then—with Rules A and B in hand—we can conclude that he is not morally responsible for his present decision to donate to charity.

The Direct Argument is highly significant. If it is successful, we have an argument for incompatibilism about responsibility and determinism that does not make use of two controversial claims typically invoked by incompatibilists: (i) a person is morally responsible for what she has done only if she could have done otherwise, and (ii) if the person's action is causally determined, then she could *not* have done otherwise. Since compatibilists typically deny one or the other of these claims, the Direct Argument offers an intriguing way to argue for incompatibilism about responsibility and determinism that sidesteps many of the traditional battlegrounds between compatibilists and incompatibilists.

In recent years, the Direct Argument has received a lot of critical attention, most of it paying special attention to the Argument's inference rules, as well as some key metaphysical assumptions. In particular, Rule B has come under the most fire from compatibilists wishing to

disarm the Direct Argument. This has happened in (at least) two ways. First, and most prominently, compatibilists have leveled alleged counterexamples to Rule B in an attempt to show that the rule is invalid. Second, compatibilists have leveled charges of dialectical impropriety with respect to Rule B: they've either said that the rule is straightforwardly question-begging or, more subtly, that it lacks any non-question-begging antecedent support.

Rule A, too, is not without its detractors. Though most philosophers agree with van Inwagen that the validity of Rule A is "beyond dispute" (1983, p. 184), there is some recent literature that tries to argue that someone is (or could be) morally responsible for a necessary truth. And if this sort of objection to Rule A is successful, then (obviously enough) Rule A is invalid and the Direct Argument fails.

I think, then, that the recent challenges to Rules A and B lead to at least two important questions that must be answered by the incompatibilist. (i) are the inference rules, upon which the Direct Argument rests, valid? And, (ii) are the inference rules, upon which the Direct Argument rests, admissible in the relevant dialectical context? A negative answer to either (i) or (ii) undermines the Direct Argument more generally.

Finally, there is the recent, so-called, No Past Objection to the Direct Argument.² The upshot of the No Past Objection is that a remote past isn't a necessary condition of the

² To be clear, the No Past Objection is raised, in particular, against the so-called Consequence Argument, an *indirect* argument for the incompatibility of moral responsibility and determinism. Van Inwagen (1983, p. 94), puts the argument (or, more exactly, argument *form*) as follows. Assume the following two principles:

Rule α : $\Box p \vdash Np$, and

Rule β : $Np, N(p \supset q) \vdash Nq$, then

where 'N' is a modal operator that means 'it is not now, nor has it ever been within anyone's power to render it false that', and 'p' and 'q' are any propositions whatever, what follows is that, if determinism is true, then

(1*) $\Box (P \& L \supset p)$ By definition of 'determinism'

‘determinism’ thesis. And since the premises of the Direct Argument aren’t necessary truths—for, the Direct Argument asserts that, if determinism is true, then, necessarily, any proposition about the remote past, conjoined with the laws of nature, entails a unique future—the Direct Argument fails to support incompatibilism since incompatibilism, if true, is necessarily true. Or, another way to think about the No Past Objection is this. The No Past Objection’s proponent points out that, at best, the Direct Argument can establish only this: necessarily, no deterministic universe *with a remote past* contains morally responsible agents. But *that* thesis leave open the following thesis: possibly, *some* deterministic universe that *lacks* a remote past contains morally responsible agents. And if this last thesis is true, then compatibilism is true. So, even if the Direct Argument is sound, it does not rule out compatibilism, and so is not a valid argument for incompatibilism. What is important about this objection is that it *grants* the truth of the inference rules upon which the Direct Argument rests, so even if the incompatibilist can answer (i) – (ii) in the affirmative, this will have no effect on the No Past Objection. Thus, the No Past Objection generates yet another important question that the incompatibilist must answer: (iii) if there needn’t be a remote past, and there are possible universes that lack a remote past but

(2*) $\Box (P \supset (L \supset p))$	1*, and logic
(3*) $N(P \supset (L \supset p))$	2*, Rule α
(4*) NP	No one (ever) has the power to render it false that P
(5*) $N(L \supset p)$	3*, 4*, Rule β
(6*) NL	No one (ever) has the power to render it false that L
(7*) Np	5*, 6*, Rule β

As with the Direct Argument, let ‘P’ stand for ‘the remote past’ and ‘L’ stand for ‘the laws of nature, and ‘p’ stand for any proposition whatever. What follows, if the argument is successful, is that, if determinism is true, then no one has any choice about anything.

Now, what’s important to notice about the Consequence Argument is that it has the same first premise, and (basically, though the operator is different) same fourth premise. Each of these premises—in either argument—rely on there being a remote past. And this is what the No Past Objection aims to undermine.

include morally responsible agents, then doesn't the Direct Argument fail as an argument for incompatibilism?

Thus, there seem to me to be three crucial questions that an incompatibilist must answer. A negative answer to any of the three will prove to be fatal for the Direct Argument.³

II. Why This is Important

The overarching issue of my dissertation is whether or not moral responsibility is compatible with causal determinism. But, before I go into more details as to why this is important, let me address a key worry someone might raise here. Perhaps someone might object to this issue by arguing that it's just obvious that we're morally responsible for some stuff whether or not causal determinism is true. So, what's all the fuss about?

In reply, let me first say what I mean by moral responsibility. I don't mean whether *we have* moral responsibilities—clearly we can *have* moral responsibilities whether or not causal determinism is true. Instead, by 'being morally responsible for some stuff' I mean 'being blameworthy or praiseworthy for what one does'. But even with this in hand, the critic might still say, "isn't it just obvious that we're sometimes blameworthy or praiseworthy for what we do, and this is the case whether or not causal determinism is true?" In reply, I think the critic is making a clear mistake. What *is* obvious is that we do sometimes *regard* one another as praiseworthy or blameworthy for what we do. This much, I grant, is true. But of course, it's one thing to *regard* or *take* someone to be praiseworthy or blameworthy; it's another thing

³ One might wonder whether or not a successful No Past Objection really would be *fatal* to the Direct Argument. After all, even if the No Past Objection is successful, it might be that a sound Direct Argument shows what incompatibilists generally want, anyway, viz., the conclusion that *in worlds like ours* causal determinism rules out moral responsibility. Fair enough. But, the stronger view (i.e. the view that says incompatibilism is necessarily true if true at all), I take it, is the classical incompatibilist view. And, inasmuch as the Direct Argument was launched in an effort to argue for *this* claim, a negative answer to (iii) *will* prove fatal. Nevertheless, I invite the reader to skip chapter five if he/she doesn't find it necessary to argue for the modally strong incompatibilist claim.

entirely for them to actually *be* praiseworthy or blameworthy. And, *pace* the critic, it's not at all obvious that people are *in fact* blameworthy or praiseworthy whether or not causal determinism is true. In fact, there are some well-known skeptical arguments that purport to show that if determinism is true, then people can't be morally responsible (e.g. the Consequence Argument, and the Direct Argument). Moreover, there are also some well-known arguments that purport to show that if determinism is *false* then people can't be morally responsible (e.g., the so-called 'luck objection', and the so-called 'Mind' argument). My dissertation is a critical examination of one of the arguments—the Direct Argument—that purports to show that if determinism is true, then people are not morally responsible for anything they do.

III. In it to Win it

My aim in this dissertation, then, is to demonstrate (minimally) these three things: (a) that the inference rules upon which the Direct Argument rests are both valid and permissible in the relevant dialectical context; (b) that the No Past Objection *really does* undermine the Direct Argument as it is currently stated, but that it does not undermine the revised version of the Direct Argument that I'll present; and (c) that the success of the revised Direct Argument provides clarity for moving forward in the relevant and related debates. Thus, the central purpose of this dissertation is to defend the inference rules of the Direct Argument in light of recent objections, and moreover, to revise the Direct Argument in light of the No Past Objection.

To demonstrate (a), I intend to defend Rules A and B from various recent objections. In particular, I intend to defend Rule B's validity from alleged counterexamples offered in the recent literature by philosophers such as John Martin Fischer and Mark Ravizza, Ishtiyaque Haji, Seth Shabo, and David Widerker and Ira Schnall. I'll do this in chapter two where each alleged

counterexample will be taken in turn, each being shown to fail outright, or to be inconclusive and so of no help to the Rule B opponent.

Moreover, in chapter three, I intend to defend Rule B's propriety within the relevant dialectical context from recent objections to the contrary by Fischer, and Michael McKenna. Thus, chapters two and three will conclude that Rule B is safe from counterexample and that it is dialectically appropriate, respectively.

Lastly, I will defend Rule A's validity from recent objections made by Stephen Kearns. While most philosophers think it's obviously true that no one is (or can be) morally responsible for a necessary truth—which is just what Rule A says—Kearns has recently raised an objection to the contrary. So, the central aim of chapter four, then, is to analyze, and argue against, Kearns's attempts to provide a counterexample to Rule A. Put another way, in chapter four I will show that *contra* Kearns, nobody is (or can be) morally responsible for a necessary truth; thus, Rule A is valid. What's more, given that chapters two and three will have successfully defended Rule B from salient objections, chapter four will conclude that the key inference rules upon which the Direct Argument rests are both valid.

To demonstrate (b), I will, in chapter five, show that the No Past Objection, as recently raised by Joseph Keim Campbell, *is* successful at undermining the Direct Argument as it is currently stated. Even so, I intend to argue that there is a way to revise the Direct Argument in light of the No Past Objection; a way that is consistent with my defenses of the Direct Argument's key inference rules, and is such that it not only circumvents the No Past Objection, it also rules out the possibility of any relevantly similar objections.

After due reflection on the success of the revised version of the Direct Argument that I'll offer, we'll see that (c) is true. That is, we'll see that a successfully revised Direct Argument

will provide a guiding light into the future of the discussion(s) surrounding moral responsibility and free will. I will discuss this much, in chapter six. But this is only part of what I'll do in chapter six. In the main, I argue that by reflecting on the nature of *truth*, we can see that counterexamples to Rule B are impossible. This conclusion, of course, strengthens my defense of Rule B in chapter two.

CHAPTER TWO:

Criticisms of Rule B (I): Alleged Counterexamples

The subject of this project is the Direct Argument, which goes as follows. Assume the following two principles

Rule A: $\Box p \vdash \text{NR}p$, and

Rule B: $\text{NR} (p \supset q), \text{NR}p \vdash \text{NR}q$,⁴

where ‘NR’ is an operator that means ‘nobody is now, or ever has been, even partly morally responsible for the fact that,’ and ‘p’ and ‘q’ are any propositions whatever. With these two rules in hand, proponents of the Direct Argument argue that if determinism is true, then (where ‘P’ stands for ‘the remote past’, ‘L’ for the ‘laws of nature’, and ‘p’ for ‘any true proposition whatever’)

- | | |
|---|--|
| (1) $\Box (P \& L \supset p)$ | By definition of ‘determinism’ |
| (2) $\Box (P \supset (L \supset p))$ | 1, exportation |
| (3) $\text{NR} (P \supset (L \supset p))$ | 2, Rule A |
| (4) $\text{NR} P$ | NR for what went on in the remote past |
| (5) $\text{NR} (L \supset p)$ | 3, 4, Rule B |
| (6) $\text{NR} L$ | NR for the laws of nature |
| (7) $\text{NR}p$ | 5, 6, Rule B. |

⁴ Originally from Peter van Inwagen’s *An Essay on Free Will* (Oxford University Press, 1983), p. 184. The operator in the original is ‘N’. However, it’s become standard parlance to express the operator in this argument as ‘NR’, as above.

The upshot of the Direct Argument is that, if determinism is true, then nobody is now, or ever has been even partly morally responsible for any fact whatever (including, of course, facts about our behavior). If so, then moral responsibility and determinism are incompatible.

Now, incompatibilists take it that Rule B is a valid rule of inference—at any rate, Rule B is *prima facie* valid. And Rule A seems to be an undeniable truth.⁵ Compatibilists, then, generally pick up their fight with the Direct Argument at Rule B, usually by attempting to offer a counterexample. Thus, the central point of this chapter: to investigate alleged counterexamples to Rule B.

There are two ways in which a compatibilist might try to call Rule B into question. She can try either to “produce some set of propositions intuitively more plausible than the validity of [Rule B]...or [come up with] a counterexample [to Rule B] that can be evaluated independently of the question whether moral responsibility and determinism are compatible” (van Inwagen, 1983, p. 188). In this chapter, I deal only with the second option, viz., attempts to construct counterexamples to Rule B. Thus, in this chapter, I will defend Rule B from some of the most recent and forceful of these attempts.

II. Fischer and Ravizza’s Alleged Counterexample to Rule B

In chapter six of their influential book *Responsibility and Control* (1998), John Martin Fischer and Mark Ravizza set out to show that the Direct Argument fails. In particular, Fischer and Ravizza set out to show that the Direct Argument’s Rule B is invalid. They do this by positing certain Frankfurt-style thought experiments,⁶ thought experiments in which an outcome is, in some sense, overdetermined, i.e., is such that if the outcome was not brought about by the

⁵ But see Stephen Kearns’s “Responsibility for Necessities,” *Philosophical Studies* 155 (2011), 307 – 324. I deal with Kearns’s objection to Rule A in chapter four.

⁶ These types of thought experiments find their origin in Harry Frankfurt’s (1969).

agent herself, then it would be brought about by some *other* fact. And it's these thought experiments that they allege are counterexamples to Rule B. I think their allegations are mistaken; their proposed counterexamples to Rule B are not counterexamples at all. In what follows, I intend to show why I think this is the case.

In order to see how my objections to Fischer and Ravizza's counterexamples will go, let me first restate their case and present their proposed counterexamples to Rule B. Consider:

Erosion: Imagine that Betty [a soldier charged with destroying an enemy fortress] 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 at T2, starting a natural avalanche that would have crushed the enemy camp at T3. (Fischer and Ravizza, 1998, p. 157)

Erosion is alleged to be a counterexample to Rule B because

1. The glacier is eroding and no one is, or ever has been, even partly responsible for the fact that it is eroding; and
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 [for the avalanche crushing the enemy base at T3], 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. (Ibid., my insertion)

So, it appears that Rule B is invalid. For, there are two paths that suffice for the enemy camp's having been crushed by the glacier: one that in fact obtains; the other counterfactual since that *path* didn't obtain (though it *would* have). But, since this counterfactual *natural* intervener—the erosion of the glacier—does not *actually* cause the avalanche, it does not remove Betty's moral responsibility for the enemy camp's having been crushed by the glacier. So, even though the

enemy camp's being crushed by the glacier is inevitable, given the circumstances, it doesn't follow that Betty isn't morally responsible for its having been so crushed. Rule B is invalid.

To put the point a bit more clearly, notice that Erosion contains two paths. The first path passes through Betty, a normally functioning agent. The second path, however, does not pass through Betty (or any other normally functioning agent). The second path is merely a counterfactual path that Fischer and Ravizza call the 'Ensuring Path'. The Ensuring Path, obviously enough, ensures that the consequence—in this case, the enemy's being crushed by the glacier—obtains. So, even though,

4. 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

5. 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 is false that

6. the outcome is reached and no one is, or ever has been, even partly responsible for this fact.

6 is false because, since the outcome (the camp's being destroyed by the avalanche) was not caused by the natural intervention of ice and rocks breaking free (but, rather, by Betty's placing the dynamite), Betty is responsible for the enemy camp's having been crushed by the glacier even though this would have happened even if she had never detonated her explosives at T1.

Rule B is invalid.

But I think that this argument fails, for 6 isn't clearly false. There are two reasons I think this. First, there might be a plausible way in which a proponent of Rule B could modify the rule so that the *spirit* of Rule B is weaker than what Fischer and Ravizza give it credit. This is the move that Michael McKenna (2008) suggests for the Direct Argument's proponent. Second—

and I think this is the stronger of the two reasons—I think Erosion simply misses the salient sort of responsibility that Rule B addresses, namely, *direct* responsibility; and any successful counterexample to Rule B will have to include an instance of *direct* moral responsibility.⁷ I'll now defend these two reasons for thinking that 6 isn't clearly false.

McKenna argues that Frankfurt-style counterexamples (such as Erosion) do, technically speaking, undermine Rule B. But that's only if we read Rule B as Fischer and Ravizza read it. They read it as follows:

Transfer of Non-responsibility (TNR):

- i. p, and no one is, or ever has been, even partly responsible for the fact that p.
- ii. If p, then q, and no one is, or ever has been, even partly responsible for the fact that if p, then q.
- iii. Therefore, no one is, or ever has been, even partly responsible for the fact that q.

If TNR is left as is, it's easy to see why Fischer and Ravizza think a case like Erosion undermines Rule B. But McKenna thinks that such an objection doesn't get at the heart of the Direct Argument; that is, such a counterexample doesn't get at the heart of Rule B. Rule B, and, so, the Direct Argument, can be saved from a case like Erosion if we simply amend Rule B to read as follows:

Transfer NR** (TNR**):

- i*. p at time T1, and no one is, or ever has been, even partly responsible for the fact that p;
- ii**. a. p is part of the actual sequence of events E that gives rise to q at T2 (where T2 is later than T1);
b. p is causally sufficient for the obtaining of q at T2, and any other part of E that is causally sufficient for q either causes or is caused by p;

⁷ What is 'direct responsibility'? And why must we understand Rule B in terms of such responsibility? I answer these questions below (pp. 16 – 17).

c. no one is, or ever has been, even partly responsible for ii**.a. and ii**.b.;

iii**. Therefore, no one is, or ever has been, even partly responsible for the fact that q obtains at T2. (McKenna, 2008, p. 364)

The advantage of TNR** is that it rules out cases like Erosion; for, it rules out two-path cases.

McKenna's "central objection" to Fischer and Ravizza's use of two-path cases like Erosion is that

the manner in which the source incompatibilist [that is, one who holds the view that for an agent to be morally responsible for her actions, she must be the ultimate source of her actions, and that this sort of 'sourcehood' is incompatible with the truth of causal determinism] takes determinism to rule out moral responsibility and the manner in which the Direct Argument can be employed on behalf of the source incompatibilist is *not* relevantly like a two-path case of the sort Fischer and Ravizza enlist to reject [TNR]. If determinism is true, it is not as if there is one path to a certain result in virtue of which determinism is true, and then there is some other *distinct* path, unhinged from the former, that a morally responsible agent initiates so that he or she bears responsibility for the object of responsibility in question... The facts settled by the truth of determinism *include* the facts pertaining to the "action path" in virtue of which the agent is alleged to be responsible.⁸ (Ibid., my insertions)

This certainly seems plausible to me. Fischer and Ravizza argue 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

does not follow from

1. The glacier is eroding and no one is, or ever has been, even partly responsible for the fact that it is eroding; and
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, how are 1 and 2, in actual fact, related to 3? As Erosion tells the story, 1 and 2 are only counterfactually related; that is, the *actual* facts do *not* include the facts pertaining to the

⁸ Here, McKenna's central objection to Fischer and Ravizza seems to suggest that the reply on offer is available only to source incompatibilists. It is not clear to me whether or not the envisaged reply is available to incompatibilists generally, or whether or not he thinks the Direct Argument is an argument for source incompatibilism, but not incompatibilism generally. For what it is worth, I think that the Direct Argument is silent on the question of whether or not there is a 'sourcehood' requirement on moral responsibility.

avalanche's crushing the enemy base. But, according to source incompatibilism, if determinism is true and the past and laws are such and so, in order for Betty to be morally responsible for the destruction of the enemy base, 1 and 2 must be such that they *include* the action path containing Betty's having planted the dynamite in the crevices of the glacier given that the past and the laws are such and so. To suppose that the erosion *would*, counterfactually, result in 3 from an, as a matter of fact, unrelated 1 and 2 is to replace 3 with some other event that does *not* include the aforementioned action path, and to replace p (where 'p' is *Betty's having planted the dynamite in the crevices of the glacier*) with some, as a matter of fact, unrelated r. And what the Direct Argument is concerned with is whether or not anybody is, or could be, responsible for q when p implies q (where 'p' is *Betty's having planted the dynamite in the crevices of the glacier*). The Direct Argument (and its proponent) isn't concerned with counterfactual two-path scenarios.

But is limiting the Direct Argument's 'punch' to one-path scenarios really all that defensible? Perhaps not. I said, above, that 3 in actual fact doesn't (indeed, if determinism is true and the past and laws are such and so, 3 *can't*) follow from 1 and 2 because it's not related to 1 and 2 in any relevant way. But if not, this must mean that 3 can only follow—indeed, 3 necessarily follows—from:

1*. The past is such that it includes Betty's planting dynamite in the crevices of a glacier above the enemy's base at T1, and the laws are such and so, and no one is, or ever has been, even partly responsible for the fact that the past is such that it includes Betty's planting dynamite in the crevices of a glacier above the enemy's base at T1, and the laws are such and so; and,

2*. If the past is such that it includes Betty's planting dynamite in the crevices of a glacier above the enemy's camp, and the laws are such and so at T1, 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 1* and 2* are straightforwardly question begging against the compatibilist, the opponent of the Direct Argument. 1* and 2* are question begging because they assume that no one is, or

ever has been, responsible for the fact that the past includes Betty's having planted dynamite in the crevices of a glacier above the enemy's base camp. But why should the compatibilist agree to a thing like that? Of course the compatibilist will not agree with such a thing. She won't agree because whether or not *Betty* is responsible for her having so acted is exactly what's at issue. So, limiting Rule B to one-path cases seems, at least in cases like a modified Erosion, to beg the question.

Now it looks as if Erosion is either a counterexample to Rule B, or it generates a need to revise Rule B (that is, to read it as TNR**) in a way that ends up begging the question. What's a proponent of the Direct Argument to do?

It seems to me that the proponent of the Direct Argument ought to agree with McKenna that cases like Erosion miss the point of the Direct Argument, but for different reasons than what McKenna gives. To see what I mean, notice that Erosion would have us believe that what's at issue is whether or not Betty is responsible for

3'. There is an avalanche that crushes the enemy base at T3.

But I think the proponent of the Direct Argument should *deny* that what's at issue in Erosion is whether or not Betty is responsible for 3' since all should agree that 3' isn't something Betty could be directly morally responsible for. And it's precisely *direct* moral responsibility that is the relevant sort of responsibility to which Rule B refers.⁹ Following David Widerker, we can think of direct moral responsibility in this way: "S is directly responsible for p just in case S is

⁹ This seems to me to be what van Inwagen is gesturing at in the long parenthetical remark straddling pages 161 – 162 of *An Essay on Free Will*. And since the discussion, there, is leading into his discussion of the Direct Argument (starting at page 182), I think we should read 'NR' in the Direct Argument's inference rules as follows:

'NRp' means *p is true and no one is even partly directly morally responsible for the fact that p*.

David Widerker (2002) makes a similar point since Rule B's conclusion (where Rule B is viewed as an argument form) is with respect to direct moral responsibility. And if Rule B's conclusion is with respect to direct moral responsibility, it must be that its premises are with respect to direct moral responsibility.

responsible for p, but *not* in virtue of being responsible for some other fact” (2002, pp. 118 – 119). In other words, direct moral responsibility—the relevant sort of responsibility to which Rule B applies—is the sort of responsibility one bears for the truth of some morally significant fact, but *not* because one bears responsibility for it in virtue of some *other* fact. Thus, even if Betty is directly responsible for causing the avalanche, she is not directly responsible for 3’. But, since Erosion needs Betty to be directly responsible for 3’, Erosion fails as a counterexample to Rule B.

If I’m right about this, then I think Fischer and Ravizza are faced with a dilemma: Given Erosion, either Betty is directly morally responsible for something or she isn’t. If she *is* directly morally responsible for something, then, at most, what she’s directly morally responsible for is *causing* the avalanche,¹⁰ something that is compatible with Rule B. If she’s not directly morally responsible for something, then Rule B is confirmed, not refuted. To be clear, this dilemma defeats Fischer and Ravizza’s alleged counterexample as follows: if Betty is directly morally responsible for something, then she’s directly responsible for:

3’’. Betty causes there to be an avalanche that crushes the enemy base at T3.

But, if 3’’ is what Betty is directly morally responsible for, then Erosion fails as a counterexample to Rule B since 3’’ doesn’t follow from 1 and 2. Moreover, if Betty is *not* directly morally responsible for something, then that’s just what Rule B says.

Now, Fischer and Ravizza might object that, even if we *should* narrow the scope of 3’ to include Betty’s *having caused*, it still follows that 3’ is entailed by any such narrowing. And if so, it’s still true that Betty is directly morally responsible for 3’. But in order for an objection like this to go through, the following *prima facie* plausible principle has to be true:

¹⁰ I say “at most.” As a matter of fact, however, I think that if there is anything for which Betty is *directly* morally responsible, it’s her intention to destroy the enemy base. So, really, I think it’s highly dubious that Betty could be directly morally responsible for causing the avalanche.

ER: If S is directly morally responsible for p, and $p \supset q$, then S is directly morally responsible for q.

To be clear, the envisaged objection relies on ER because the claim is that since 3'' entails 3', and Betty is directly morally responsible for the truth of 3'', she's also directly morally responsible for the truth of 3'.

But this objection fails because ER is false. To see why ER is false, consider the following. Butch Jones is (we may assume) directly morally responsible for the fact that he is the head football coach at the University of Tennessee. But Butch Jones's being directly morally responsible for his being the head football coach at the University of Tennessee implies that there is such a place as the University of Tennessee. Even so, it's obvious that, even if Butch Jones is directly morally responsible for his being the football coach at the University of Tennessee, he is *not* responsible (directly or otherwise) for the fact that there is such a place as the University of Tennessee. So, ER is false.

Now, if I'm right about the need to narrow the scope of 3' (to 3'', or something like it), then from

1.** The past and the laws are such and so, and no one is, or ever has been, even partly morally responsible for the fact that the past and the laws are such and so; and,

2.** If the past and the laws are such and so, then Betty will plant dynamite in the crevices of a glacier above the enemy's base at T1 and this will cause an avalanche to crush said base at T3, and no one is, or ever has been, even partly morally responsible for this fact;

it follows that

3.** Betty will plant dynamite in the crevices of a glacier above the enemy's base at T1 and this will cause an avalanche to crush said base at T3, and no one is, or ever has been, even partly morally responsible for this fact.

And if so, then we can see that there's a problem with 4 – 6: there is an equivocation on 'outcome'. Recall 4 – 6:

4. 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

5. 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;

but it is allegedly false that

6. the outcome is reached and no one is, or ever has been, even partly responsible for this fact.

But it seems to me that, in 4 and 5, ‘outcome’ means something like 3’. In 6, however, ‘outcome’ means something like 3’’. Clearly, the Ensuring Path cannot ensure that 3’’ is reached; only the past and the laws being such and so can do that. So 4 – 6, which is supposed to follow from Erosion, does *not* undermine Rule B. For, if the past and the laws being such that they entail that 3’’, and no one is, or ever has been, even partly responsible for this fact, it seems that 3** is true. And this is just what the Direct Argument says. So, it looks as if Rule B is safe from Fischer and Ravizza’s objections, and their alleged counterexample is just that: alleged.

III. Widerker’s (2002) Attempt at a Counterexample to Rule B

While Widerker agrees that Fischer and Ravizza’s Erosion case fails as a counterexample to Rule B, he thinks that there *is* such a counterexample in the offing. Consider the following case from Widerker:

Fate: Suppose that Jones murders Smith at T_0 for some selfish reason, and that later on he murders another person, Green, at T_3 . Suppose also that the second murder is made possible by the first murder. That is, in the circumstances, the second murder requires the first murder as a causally necessary condition. Finally, suppose that Jones could have avoided murdering both Smith and Green, and believed that he could have avoided so to act. (Widerker, 2002, p. 319)

Now, given Fate, assume that it is T_1 (where T_1 is earlier than T_3 but later than T_0). Widerker thinks that Fate generates the following invalid substitution instance of Rule B:

- A. NR (Jones murders Green at T_3)
- B. NR (Jones murders Green at $T_3 \supset$ Jones murders Smith at T_0)
- C. Therefore, NR (Jones murders Smith at T_0)

If A and B are true while C is false, then Rule B is false. Widerker thinks that A and B *are* true, but that C is false for the following reasons.

Premise [A] is true, since no one is now ($=T_1$), or ever has been, directly morally responsible for a murder that has not been committed yet. And since we assumed that Jones's murdering Smith at T_0 is a causally necessary condition of his murdering Green at T_3 , [B] is true as well. But notice that the conclusion is false, since we assumed that Jones could have avoided killing Smith. (Ibid., p. 320)

So, the idea is that A – C provides a counterexample to Rule B because, assuming the time is T_1 , Jones's murdering of Green hasn't happened yet. And since Widerker thinks that moral responsibility is "a backward-looking notion" (Ibid., p. 320, footnote 8),¹¹ it's not possible to be morally responsible for something that hasn't happened yet. Thus, A is true. But if A is true, then B must be true as well since there doesn't seem to be any plausible way to be responsible *now* for some future fact's implying some past fact. Thus, by *modus ponens*, C is true. But, C is false; so, Rule B is invalid.

But I don't think that Fate provides a successful counterexample to Rule B. To see why not, I am going to borrow a move from a later paper by Widerker with Ira M. Schnall (Widerker and Schall, forthcoming). Here's how the move goes. Grant, with Widerker, that A is true; that is, that no one now is (at T_1), or ever has been, even partly directly morally responsible for the

¹¹ Widerker doesn't say in the text what, exactly, he means by moral responsibility's being a 'backward-looking notion'. But, I think that he must mean something like this. 'Being morally responsible for' is an attribute (or a relation?) one has with respect to an action that one *has done*. That is, moral responsibility is only attributable to an agent *after* she acts. Thus, it's by 'looking back', as it were, on an action that's occurred that we can see whether or not an agent is morally responsible for an action.

fact that Jones murders Green at T_3 . Now, notice that we can restate the conditional fact expressed in B can as follows:

B*. $\sim (\text{Jones murders Green at } T_3) \vee (\text{Jones murders Smith at } T_0)$

But since the first disjunct of B* is, *ex hypothesi*, false, what makes B* true is its second disjunct. That is, B* is true *because* Jones murders Smith at T_0 . Thus, Jones is at least partly directly morally responsible for B*. Moreover, B* reveals what I think is a plausible principle about moral responsibility:

Truth Dependence_{MORAL} [TDM]: For all agents, S, and all propositions, p, if S is directly morally responsible for that which p's truth depends on (in the sense of 'depends on' in which truth depends on the world), then S is at least partly directly morally responsible for p's truth.¹²

If TDM is true, then it follows not only that Jones is at least partly directly morally responsible for B*; it also follows that he's at least partly directly morally responsible for the conditional fact expressed in B of the argument generated by Fate. And if so, then Fate is no counterexample to Rule B. I think TDM *is* true; thus, I conclude that Fate is no counterexample to Rule B.¹³

¹² I have essentially taken TDM from Trenton Merricks (2011, p. 66) and his notion of what it means to 'have a choice about' the truth of a proposition. Very briefly, what I mean when I say (as I do in TDM) that truth 'depends on the world' is just the trivial notion that truth depends on the way things are and not the other way around. It's true, for example, *that dogs bark* because dogs bark. It's true *that there are no hobbits* because there are no hobbits. And so on. This is what I mean when I say that 'truth depends on the world'; I mean it in just this trivial way, setting aside questions as to whether or not there are more *substantive* things to say about how truth depends on (or is otherwise related to) the world. More about this in chapter six where I give sustained motivation for, and a defense of, TDM.

¹³ Objection: TDM isn't obviously true. A reason to think that it's not obviously true is because—granting what you've said—it entails that Jones is responsible for the conditional fact expressed in B, i.e., that he's responsible for the fact that $(\text{If Jones murders Green at } T_3 \supset \text{Jones murders Smith at } T_0)$. And this is surely unintuitive. The upshot, then, is this: isn't the seeming truth of B in Widerker's argument—that NR $(\text{If Jones murders Green at } T_3 \supset \text{Jones murders Smith at } T_0)$ —a counterexample to TDM?

Reply: This objection is a concern only if it's true that NR $(\text{If Jones murders Green at } T_3 \supset \text{Jones murders Smith at } T_0)$, but it isn't. As I argue above, and following Widerker and Schnall (forthcoming), since B* is true just in virtue of the fact that Jones murders Smith when he does, Jones is at least partly directly morally responsible for B*. That is, Jones is at least partly directly morally responsible for the conditional fact housed under the 'NR' operator in B

IV. Haji's (2008) Attempt at a Counterexample to Rule B¹⁴

Here is how any counterexample to Rule B will have to work. Rule B, as I've stated it, is a kind of argument form; so, what we'll need from a potential counterexample to Rule B is to be given substitution instances of 'p' and 'q' in the premises of Rule B such that they render both of the premises true, but the conclusion false. I think Haji's (2008) attempt to do this fails because the argument Haji offers on the basis of his alleged counterexample is question begging. It's question begging, so I say, because it presupposes that compatibilism is true. Or, even if it doesn't, strictly speaking, beg the question, a more generous way to put the problem is this: I think that Haji's alleged counterexample to Rule B is dialectically unhelpful. In this section, then, I'll attempt to show why I say this.

To help us get clear on how his alleged counterexample to Rule B is going to work, Haji does a little stage setting. Among the requirements for moral responsibility, he thinks, are a freedom requirement (i.e. free will is required for moral responsibility), and an epistemic requirement. He thinks the requisite epistemic requirement might, plausibly, go like this:

ECONDITION-2 [EC2]: Agent, S, is morally blameworthy for action, A, only if S does A at least partly on the basis of the belief that A is morally wrong; and S is praiseworthy for A only if S does A at least partly on the basis of the belief that A is morally obligatory or permissible (Haji, 2008, p. 13).

According to Haji, EC2 entails that

blameworthiness requires, loosely, only belief in what is wrong and not...that the object for which one is blameworthy is in fact wrong. Analogously, this condition implies that praiseworthiness requires only belief in what is obligatory or permissible (Ibid.).

because he is responsible for the truth of the consequent of that conditional. So, B is false. Because B is false, its seeming truth is no counterexample to TDM. Indeed, I argue that the reason B is false (and that Jones is at least partly directly morally responsible for the truth of B*) *reveals* TDM. (Thanks to David Palmer for the objection.)

¹⁴ To be clear, Haji's (2008) counterexample to Rule B also appears in his (2009). Even so, I've chosen to deal with the version that appears in his (2008), so any citations will be with respect to that piece.

There is also, thinks Haji, a plausible condition concerning the moral status of *actions*, one that isn't entailed by EC2, but can be consistently endorsed by one who thinks that EC2 is a necessary condition for moral responsibility. This plausible condition is

ACTION-STATUS [AS]: Agent, S, is morally blameworthy for action, A, only if A is morally wrong; and S is praiseworthy for A only if A is morally obligatory or permissible (Ibid.).

So, we have two plausibly necessary conditions for an agent S's being morally responsible for an action A.

In order to meet the necessary conditions for praiseworthiness or blameworthiness laid out in either EC2 or AS, an agent must be obliged to act, it must be permissible for an agent to act, or an agent must be obliged to refrain from acting, in the way for which she's allegedly praiseworthy or blameworthy. But these requirements on praiseworthiness or blameworthiness, according to Haji, come with a requirement for alternative possibilities. Why think alternative possibilities are required for obligation (to act, or refrain from acting), in particular? Haji thinks the following two principles help to illuminate the reasons why.

(K): If it is overall obligatory for one to do something, then one can do it, and if it is (overall) obligatory for one to refrain from doing something, then one can refrain from doing it [i.e. 'ought' implies 'can'].

(OW): It is (overall) obligatory for one to do something if and only if it is (overall) wrong for one to refrain from doing it. (Ibid., p. 14. My insertion.)

Haji thinks that the conjunction of K and OW entails "that an agent can do moral wrong only if the agent had an alternative" (Ibid.). In other words, the conjunction of K and OW entails

(WAP): If it is wrong for one to do something, then one can refrain from doing it (Ibid.).

So, according to WAP, an agent can do moral wrong only if said agent had an alternative.

What's more, there is an analogue of K that bears the idea that 'wrong' implies 'can' (where K is the principle that 'ought' implies 'can'). And if so, thinks Haji, then there's a straightforward

argument to the conclusion that moral obligation requires alternative possibilities. The argument goes as follows:

(P1) If it is obligatory for one to refrain from doing A, then it is wrong for one to do A.
[From OW]

(P2) If it is wrong for one to do A, then one can do A. [From ‘wrong’ implies ‘can’ analogue of K]

(C1) Therefore, if it is obligatory for one to refrain from doing A, then one can do A
(Ibid.).¹⁵

Thus, the argument that moral obligation requires alternative possibilities.

What’s needed, now, is an argument whose conclusion is such that moral *rightness* (as opposed to obligation) requires alternative possibilities. That is, we need an argument from

(P1*) If it is permissible (i.e. morally right) for one to do A, then one can refrain from doing A,

to

(C1*) Therefore, if it is permissible (i.e. morally right) for one to do A, then one can do A.

¹⁵ For what it’s worth, though I agree with the conclusion of this argument—i.e., I think it’s true that ‘ought’ implies ‘can’—I think that this argument fails to show such a conclusion. For, I think that P2 is false. Haji, here, and in his (2012) seems simply to assert that moral obligation and moral rightness/wrongness are equivalent. That is, morally obligatory actions *just are* morally right actions. Though I agree that an action’s being morally obligatory implies that the action is morally right, I deny that the implication works the other way. That is, I deny that what is morally right is thereby morally obligatory.

To begin to see why I say this, note that, and as I’ll describe below, Haji thinks that what’s morally obligatory, permissible (i.e. morally right), or impermissible (i.e. morally wrong) partly depends on what possible worlds are ‘accessible’ for one. I am happy to grant this with respect to moral obligation (though, I have my doubts), but I think that this *must* be false with respect to moral rightness or wrongness. Here is why: if moral rightness or wrongness—i.e., the notions of ‘moral’ or ‘immoral’—are partially dependent on the worlds that are accessible to an agent, then it turns out that what’s moral or immoral partially depends on the agent and her circumstances. But this is just a version of moral relativism. This view implies that, if the only worlds accessible to me are worlds in which I torture a little baby for fun, then torturing a little baby for fun *isn’t* immoral. But I think that whether or not it’s immoral to torture a little baby for fun doesn’t at all depend on me or my situation; rather, I think it’s necessarily true that it’s immoral to torture a little baby for fun. And this is because I think that morality isn’t relative to anything about me or any other agent, which is just to say that morality isn’t relative. But this argument for ‘ought implies can’ relies on such a notion. Since I think that such a notion is false, I think that P2 is false, and the argument fails.

Initially, Haji claims that we *cannot* give such an argument because “[t]here is no similar way [i.e. no argument like P1 – C1] to derive the proposition that rightness, likewise, requires alternative possibilities. This is because even if it is agreed that ‘right’ implies ‘can’, there is no principle like (OW) that will allow us to infer ‘right’ implies ‘can refrain’” (Haji, 2008, p.14). “Nevertheless,” he says, “it is very plausible that ‘right’ does imply ‘can refrain’” (Ibid.). It’s “very plausible” that ‘right’ implies ‘can refrain’, thinks Haji, because, if we think about a typical Frankfurt-style situation, we can see that there are only two plausible moral judgments that we can make with respect to the situation-relevant agent. The agent’s action is either amoral (i.e. neither being right, nor wrong, nor obligatory), or it is morally permissible. For, suppose that Jones is in a typical Frankfurt-style situation in which he kills Smith but couldn’t refrain from doing so because Black was waiting in the wings to ‘make’ him do it. Either Jones’s killing Smith is amoral for Jones, or it’s permissible. But it’s not plausible that Jones’s killing Smith is permissible, so it must be amoral for Jones. Thus, Haji concludes that moral rightness (or permissibility) requires alternative possibilities.

In his most recent work, however, Haji (2012) gives reason to think that we *can* give an argument that goes from P1* to C1*. Assuming that moral responsibility requires ‘can’ (and I am happy to assume this),

[o]ne is morally praiseworthy or blameworthy (whatever the case may be) for doing something only if one *could have* done it. This principle highlights a link between moral responsibility and freedom that holds of conceptual necessity. Again, the link is simply that praiseworthiness and blameworthiness both require control [i.e. both require a ‘can’ condition]. Furthermore, if we think that praiseworthiness requires a certain variety of control, then without good reason to believe the contrary, blameworthiness, also requires this variety of control. Similarly, it would seem that control or freedom requirements of moral obligation, *unless we have strong reason to think otherwise* (and I know of no such reason), should also be the very ones of moral wrong and moral right. If [K] expresses just one more incarnation of the association between morality and freedom, then, again, in the absence of special reason to believe otherwise, it should also be the case that the principle that each of “wrong” and “right” implies “can” expresses two other instance of

this association. In brief, if obligation requires control, so do right and wrong; and if obligation requires a certain sort of control—that we can do the things we are obligated to do—right and wrong require this sort of control as well. (Haji, 2012, pp. 25 – 26, my insertions)

So, the idea is that [K] shows us that morality and freedom (‘freedom’ in the sense of ‘can’) are closely associated. And if we accept that OW is true, then unless we have some special reason to deny that moral obligation and rightness (or wrongness) come apart, then we should conclude that moral rightness (and wrongness) are subject to a ‘can’ condition. Thus, if ought implies can, then moral rightness implies can.

Haji gives further motivation for this belief that there is a symmetrical relationship between the freedom requirements on moral oughts and the freedom requirements on moral rightness (and moral ought *nots* and moral wrongness). Following the work of Fred Feldman (1986) and Michael Zimmerman (1996; 2008), he thinks of moral obligations as resulting from the possible worlds that are “accessible” to one. The idea, here, is that “a possible world is accessible to a person at a time if and only if it is still possible, at that time, for the person to see to the world is actual” (Haji, 2012, p. 27). Here’s an illustration. Suppose that, right now, I can’t, alone, get to my doctor’s appointment on time, but I could if you were to give me a lift in your car. If, right now, I can call you on the phone and convince you to give me a ride to my doctor’s office, then there is a world accessible to me in which I make it to my doctor’s appointment on time.

On such an analysis of accessible possible worlds, an agent’s moral obligations turn out to depend on which worlds are accessible to that agent. And each accessible world is given a sort of deontic ‘ranking’, a “deontic value” (Ibid.). This suggests the following principle:

(MO): A person, S, ought, as of, t, to see to the occurrence of a state of affairs, p, if and only if p occurs in some world, w, accessible to S at t, and it’s not the case that p’s

negation (not-p) occurs in any accessible world deontically as good as or as deontically better than w. (Ibid.)¹⁶

For Haji, MO results in the following definitions:

Morally obligatory: A is morally obligatory for S, at t, if and only if S can do A, and A occurs in all the best worlds accessible to S at t.

Morally permissible (i.e. morally right): A is morally permissible for S, at t, if and only if S can do A and A occurs in some but not all of the best worlds accessible to S at t.

Morally impermissible (i.e. morally right): A is morally impermissible for S, at t, if and only if S can do A and A does not occur in any of the best worlds accessible to S at t.¹⁷

So, what is morally obligatory, morally right, or morally wrong for an agent to do depends on the situation in which she finds herself; that is, these things depend on which possible worlds are open to her at a time, t. Thus, since what is morally obligatory, morally right, or morally wrong for an agent depends on what an agent can do, it follows trivially that what is morally obligatory implies ‘can’.

That’s the set up. Now, the alleged counterexample.

Hal’s Creation(1): Consider...Hal, who creates a best world. Hal cannot do otherwise than create this world, W1, owing to an unknown (to Hal) condition of his brain...He has no alternative possibilities to his actual conduct. Hence, given the principle that obligation, right, and wrong require alternative possibilities, Hal’s act of creating W1 does not violate a moral obligation of his; nor is this creative act right or obligatory. Suppose, also, that Hal is an amoral agent as he lacks any understanding of moral right, wrong, and obligation...If praiseworthiness requires only belief in what is obligatory or right...then again Hal is not praiseworthy for his creative act because Hal is amoral: being amoral he does not create W1 on the basis of the belief that he is doing something morally obligatory or right. Now imagine that Hal creates a deterministic world in which a person, Michael performs an act [at time t]—Michael eats noodles. And suppose that Michael’s doing so is [the sort of act for which free agents are] morally praiseworthy (Ibid., p. 15).

From Hal’s Creation(1), Haji generates the following argument which he thinks impugns Rule B, call it ‘Hal’s Argument(1)’.

¹⁶ But see, initially, (Feldman, 1986, p. 37).

¹⁷ See (Haji, 2012, p. 28).

Hal's Argument(1): (1H): NR (Hal creates W1)

(2H): NR (Hal creates W1 $\supset$ Michael decides, at ts, to eat noodles)

(3H): Therefore, NR (Michael decides, at ts, to eat noodles) (Ibid., p. 16)

Where 'ts' is sometime prior to t, Hal's Argument(1) impugns Rule B because "there is nothing in Hal's Creation[(1)] to preclude its being true that Michael's decision, at ts, to eat noodles is free in the compatibilist sense of 'free'...[u]nless the proponent of the Direct Argument assumes the falsity of compatibilism" (Ibid, p. 17). So, 3H is either false—given the truth of compatibilism—or the proponent of Rule B confronted with Hal's Argument(1) presupposes that compatibilism is false; and *that's* question-begging.

What's more, on a compatibilist notion of freedom, both 1H and 2H could be true while 3H is false. 1H is true, thinks Haji, for one or both of the following reasons. First, if AS is true, then it follows that if we conjoin AS with the fact that Hal's creative act is amoral, Hal is not morally responsible for his creative act. Second, if EC2 is true, then it follows that Hal is not morally responsible for his creative act since he, *ex hypothesi*, lacks any understanding of right, wrong, or moral obligation.

2H is true because, according to EC2, if Michael is to be responsible for the truth of the material conditional portion of 2H, he'd have to at least believe that it's permissible or obligatory to make it the case that (Hal creates $\supset$ Michael decides, at ts, to eat noodles). But, of course, how could Michael be expected to believe a thing like that? So, Haji concludes that Hal's Creation(1) is a counterexample to Rule B.

But I think that Hal's Creation(1) fails as a counterexample to Rule B because I think that Haji's set-up of the case has begged the question against the incompatibilist. Or, more

generously, I think that Hal's Creation(1) proves to be dialectically unhelpful given the context of the debate between compatibilists and incompatibilists.

To see why I say that Hal's Creation(1) begs the question, recall Hal's Argument(1). Haji thinks that Hal's Argument(1) tells *against* Rule B. But why think a thing like that since Haji thinks that 3H just straightforwardly begs the question against the compatibilist? For, Haji thinks that "compatibilists are well within their rights to claim that, at W1, when Michael eats noodles, he is praiseworthy for doing so" (Ibid., p. 17). Exactly what rights are those to which Haji refers? The answer to this question isn't at all clear to me. Indeed, I doubt the answer is clear at all. Whatever the answer, it *can't* be anything like

Non-starter: Take Michael's eating noodles at t. Obviously, we compatibilists think that he does this freely; and, moreover, our compatibilist theory is a highly plausible, well-motivated theory, that's backed by many great historical and contemporary philosophers. So, put the sort of action that our plausible theory suggests is consistent with moral responsibility—like Michael's eating noodles at t, for example—in the conclusion of the argument generated by Rule B, and Rule B will be false.

The answer can't be anything like Non-starter, because Non-starter is a non-starter; it begs the question against Rule B's proponent. The point of a principle like Rule B is that its *prima facie* validity is enough to show that Michael's eating noodles, at t (for example), to be something he's not morally responsible for.¹⁸ And this because it follows necessarily from the conjunction of the past and the laws of nature. So, if Rule B is *prima facie* valid, then (unless we're given some argument as to why we should think that Rule B is false) in order to conclude that Michael is directly morally responsible for his eating noodles at t, we have to assume that incompatibilism is false (i.e. that compatibilism is true). Thus, Hal's Creation(1) begs the question.

¹⁸ A similar point about Rule B's *prima facie* validity is made by Schnall and Widerker (2012, and forthcoming). I say more about this in the next chapter.

What the Rule B proponent demands from the compatibilist is a counterexample that shows *why* she should think Michael is morally responsible for his having eaten the noodles when he did. And the counterexample cannot presuppose the truth of compatibilism.¹⁹ Hal's Creation(1) fails to do any such thing. I conclude, then, that Haji's (2008) attempt at a counterexample to Rule B fails.

An objector, though, might think I've been too harsh. I've been too harsh, the objector might claim, because

if a compatibilist has good reasons she believes she can marshal for why an agent is free and responsible in acting as she does, then she is entitled to call attention to those and ask an interlocutor to weigh those as reasons against an opposing thesis. Since the incompatibilist is the one advancing an argument here with the Direct Argument, and she wishes to say that it follows from her principles that a person like Michael is not responsible, the compatibilist should not be prohibited from just offering the reasons underwriting her theory for the opposite thesis.²⁰

So, according to the envisaged objector, Hal's Creation(1) doesn't beg the question; for, Haji is within his rights to put forth his well-motivated compatibilist theory as a reason to think the Direct Argument fails; thus, Hal's Creation(1) doesn't beg the question against the Direct Argument's proponent.

Let's grant that this is so: Hal's Creation(1) doesn't, strictly speaking, beg the question. Even if we concede that this is true, there is still reason to think that Hal's Creation(1) fails to provide a convincing counterexample to Rule B, and here is why. In order to accept that Hal's Creation(1) is a legitimate counterexample to Rule B, one has to be antecedently committed to the truth of compatibilism. It would be a far more effective strategy for undermining Rule B—

¹⁹ Note: I do not say that this counterexample *depends on* the truth of compatibilism. I only say that it, *in fact*, *presupposes* the truth of compatibilism, given its setup. If this counterexample is set in an *indeterministic* context, then it will fall, *mutatis mutandis*, to what I say about the counterexample under consideration in the next section.

²⁰ Thanks to a referee from *Philosophical Explorations* for raising this objection.

and with it the Direct Argument—if one could find counterexamples that would be appealing to fair and neutral-minded inquirers. If this is true, then it is better not to make use of examples at all in which determinism is assumed, and the putative counterexample to Rule B gets its force simply by appealing to an agent who *would* satisfy relevant compatibilist conditions of freedom and responsibility. But Hal’s Creation(1) fails to do any such thing; so, I conclude that Hal’s Creation(1) fails to provide a convincing counterexample to Rule B.²¹

V. Haji’s (2010) Attempt at a Counterexample to Rule B

In this section, we’ll turn to Haji’s (2010) attempt at a counterexample to Rule B. But, before we do, I wish to note this bit of curiosity. Both of Haji’s alleged counterexamples, as we’re about to see, go by the name ‘Hal’s Creation’; yet, they’re different thought experiments. This is curious because Haji, in his (2010) doesn’t mention his (2008) attempt at a counterexample. Even so, there are crucial differences between Haji’s two recent attempts at providing a counterexample to Rule B. The most salient difference, it seems to me, is that the moral agent in Haji’s (2010) is free in the libertarian sense whereas the moral agent in his (2008) is free in the compatibilist sense. This difference, I think, makes his (2010) a more formidable challenge to the Rule B proponent.²²

²¹ Thanks to a referee from *Philosophical Explorations* for help on this point.

²² I think it’s worth pointing out that Haji’s (2010) attempt at a counterexample is a more formidable opponent (than his (2008)) for the Rule B proponent only if Rule B is as I have it in the paper. But, Rule B might be better stated in a way that mirrors Ted Warfield’s (1996) reconstruction of Rule β of the so-called Consequence Argument. Where Rule β has the modal operator ‘N’ that means ‘it is not now, nor has it ever been, within anyone’s power to render it false that’, it goes as follows:

$$\text{Rule } \beta: N(p), N(p \supset q) \vdash Nq.$$

Warfield’s reconstruction, however, went like so:

$$\text{Rule } (\beta \square): N(p), \square(p \supset q) \vdash Nq.$$

Here, then, is Haji's (2010) attempt at a counterexample to Rule B. Consider:

Hal's Creation(2): Hal-2 [an essentially omniscient, sempiternal, amoral—that is, lacks knowledge of moral right, wrong, or obligatory—being] has the ability to create (or actualize) any one of an infinite number of possible worlds... Suppose Hal-2 creates a world, W1, in which, after due reflection, Yasmin [in a libertarianly free way] donates a large sum of money (at some time, ts) to a credible charity, UNICEF. Yasmin really cares about the plight of the needy children; she donates because she wishes to help the kids and not, for instance, because she wants a big tax break. We safely suppose that she is morally praiseworthy for her bountiful donation. Under appropriate circumstances normal agents would be deserving of praise for such an act (Haji, 2010, p. 125, my insertions).

Given Hal's Creation(2), Haji generates the following argument:

Hal's Argument(2):

(4H): NR (Hal-2 creates W1).

(5H): NR (Hal-2 creates W1 $\supset$ Yasmin donates (at ts in W1)).

(6H): Therefore, NR (Yasmin donates (at ts in W1) (Ibid.).

The upshot of Hal's Argument(2) is that, since Hal is an amoral agent, 4H is true because neither he nor anyone else is morally responsible for the creation of W1, 6H is false because, *ex hypothesi*, Yasmin (libertarianly) freely donates the money at ts in W1 in the right sort of way for moral responsibility, but 5H is true because neither Hal nor Yasmin is morally responsible for the fact that (Hal-2 creates W1 $\supset$ Yasmin donates (at ts in W1)).

Of course, one might object that Yasmin *is* morally responsible for the truth of 5H. But, if so, then Yasmin would have to be morally responsible for

So, suppose the Direct Argument is better developed with the following principle in place of Rule B:

Rule (B $\square$): NR(p), $\square(p \supset q) \vdash$ NRq.

Where q occurs later than p (since the Direct Argument is concerned with past to future determinism), if the Direct Argument is better developed with Rule (B $\square$), then Haji's (2010) attempt at a counterexample doesn't even get off the ground. It doesn't get off the ground because it's not possible for a thought experiment set in an indeterministic world (which Haji's (2010) is) to be a substitution instance of Rule (B $\square$). Also, see Thomas Crisp and Ted Warfield's (2000). Thanks to E. J. Coffman for the pointer.

YC: Hal-2 creates $W1 \supset$ at ts Yasmin donates to UNICEF (Ibid., p. 127).

But, thinks Haji, for Yasmin to be morally responsible for YC, she'd have to meet the following epistemic condition on moral responsibility, namely:

E3: S is morally praiseworthy [blameworthy] for seeing to the occurrence of state of affairs, a, only if S morally ought to have known or believed that a is morally obligatory [morally wrong] or, as some might prefer, morally good [morally bad] (in some specified sense of 'good' ['bad']) (Ibid.).²³

And how *could* Yasmin be expected to meet an epistemic condition like E3 with respect to YC?

Indeed, "there are no compelling grounds to require that in Hal's creation, Yasmin ought to have known or believed that YC is obligatory or morally good" (Ibid.). So, Yasmin is not morally responsible for YC even though she's responsible for donating, at ts, to UNICEF. But, since Hal isn't responsible for YC either, 5H is true. Thus, thinks Haji, we have a successful counterexample to Rule B.

But Haji is mistaken: Hal's Creation(2) fails as a counterexample to Rule B. To better see why I say this, let's, again, consider Hal's Argument(2).

Hal's Argument(2):

(4H): NR (Hal-2 creates $W1$).

(5H): NR (Hal-2 creates $W1 \supset$ Yasmin donates (at ts in $W1$)).

(6H): Therefore, NR (Yasmin donates (at ts in $W1$)) (Ibid., p. 125).

I think I can challenge both 4H and 5H. Let's start with 5H. 5H depends on the truth of E3; for, how could it be the case that Yasmin *morally* ought to have known or believed that YC is

²³ More curiosity. In his (2010), Haji doesn't mention EC2 (from above). E3 and EC2 are not the same; so, does Haji think that E3 is an epistemic principle that's more sound than EC2? If so, why so? If not, why the change? One thing to note, however, is that E3 allows for culpable ignorance. But, if (P2) is false—where (P2) is the claim that if it's wrong for one to do A, then one can do A—then it's plausible that Hal-2 meets the epistemic condition for moral responsibility after all. Hal-2 might meet the epistemic condition because, plausibly, he *should* have known that creating $W1$ was good. If so, then the argument that Hal's Creation(2) generates—Hal's Argument(2), below—fails to go through. Thanks to David Palmer for pointing this out.

morally good or obligatory? Indeed it does seem strange that Yasmin *morally* ought to have believed a thing like YC. But, then, it seems strange to say that anyone *morally* ought to believe *anything*; for, as Robert Audi points out in his (2001), it's highly plausible that "neither believing nor forming beliefs is a case of action" (Audi, 2001, p. 105). But if neither believing nor forming beliefs is a case of action, and actions are the only things for which anyone could plausibly be held morally responsible, then it isn't the case that anyone *morally* ought to believe anything.²⁴ And if this is right, then there's a problem with E3: it's too strict. If E3 is too strict

²⁴ Two objections (and two replies):

Objection #1: There is a false hidden premise in your argument. Call it:

Q: S morally ought to A only if S can be morally responsible for A.

Here is why the hidden premise is false: suppose 'ought implies can' is false, and that you ought to do something that you cannot do. Since you cannot do it, you obviously cannot be morally responsible for it. But it's still true that you ought to do it; so it's false that S ought to A only if S can be morally responsible for A.

Reply #1: But I deny that Q—where Q is understood as the premise upon which *my* argument depends—is false if 'ought implies can' is false; for, this objection trades on an equivocation on the word 'can'. As I understand it, in the 'ought implies can' thesis, 'can' denotes an ability to do the act in question *given the relevant agent's actual powers*. For, suppose someone told me that I morally ought to end world hunger. If I'm a proponent of the 'ought implies can' thesis, my response will be that I'm under no such obligation because I can't, given my actual powers, do any such thing. My response would *not* be that I'm under no such obligation because *there are no possible worlds* in which I have the power to do so. Surely there are possible worlds where I *do* have said powers. But the fact of the matter is that I do not; so, I do not—supposing that 'ought implies can' is true—have any such obligation. And it is here where the problem in the objection emerges: my claim that 'it isn't the case that anyone morally ought to believe anything' relies on Q only if 'can' in the consequent is read in this logically weaker way; for, the only things for which anyone could *possibly* (in the logical sense of 'possibility') be held directly morally responsible are actions. But, in the 'ought implies can' thesis, 'can' is read in the logically stronger way I just described. So, I deny that my hidden premise is false, and this is because I deny that the falsity of 'ought implies can' implies the falsity of my hidden premise. (Thanks to David Palmer for the objection.)

Objection #2: There are *states of being* that a person can be morally responsible for. For example, suppose that I am in the state of being drunk. It's perfectly plausible that I'm *morally responsible* for my being in such a state, and being in such a state is no action. Thus, there are things for which we can be morally responsible that aren't actions at all.

Reply #2: I'm happy to grant, for the sake of the argument, that we can be morally responsible for being in certain states, where being in a certain state is not an action. So, I grant that I can be morally responsible for my being in a state of drunkenness. Even if this is right, it doesn't do much to help Haji's case. It seems to me that being morally responsible for being in a certain state is really just a *derivative* sort of moral responsibility. Plausibly, we can only be *directly* morally responsible for actions. So, in the case of my being morally responsible for being drunk, what I'm *directly* morally responsible for is my acts of drinking. What's important about this is that, with respect to Rule

an epistemic requirement on moral responsibility, then Yasmin's failing to meet E3 with respect to YC isn't yet a problem for her being morally responsible for the fact expressed in 5H.

Surely, though, there is *some* epistemic requirement for moral responsibility. Here, I think, is a plausible candidate for such a requirement:

E4: S is morally praiseworthy [blameworthy] for seeing to the occurrence of state of affairs, a, only if S had epistemic justification to believe that a is morally obligatory [morally wrong] or, as some might prefer, morally good [morally bad] (in some specified sense of 'good' ['bad']).²⁵

Now, E4 merely requires that Yasmin have (on balance) *good reason* to believe that YC is morally obligatory or morally good. What reason could Yasmin have to believe such a thing? Given that she's *ex hypothesi* morally responsible for the consequent of YC, she must meet whatever the epistemic requirement is for moral responsibility. So, Yasmin has good reason to believe that the consequent of YC is morally good or obligatory. But, if she has good reason to believe *that*, then since the consequent of YC (obviously) materially implies YC, itself, Yasmin has good reason to believe that YC itself is morally good or obligatory.²⁶ And if so, then,

B—and the Direct Argument more generally—what's at issue is *direct* moral responsibility. So, I think E3 must mean the sort of moral obligation that one can be *directly* responsible for following (or failing to follow). Moreover, there's still the problem of being *morally obligated* to believe (or disbelieve, I suppose) certain things. There is a whole host of literature devoted to the topic of the ethics of belief, so a discussion about this topic is well beyond the scope of this paper. Suffice it to say, however, that there is good reason to doubt (as Audi argues) that beliefs are the sort of thing for which we can be *directly* morally responsible (or for which we have some moral obligation to hold, or refrain from holding). Such doubt, I think, is reason enough to doubt that E3 is the correct epistemic principle. I'll offer another epistemic principle on moral responsibility, just below. It, *too*, might be subject to serious doubt. Even so, I think it's a plausible competitor to E3. I'll let the reader decide which she thinks is the more plausible. But the point is that Haji's counterexample depends on E3, and we have reason to doubt E3; so, we have reason to doubt that the counterexample is successful. (Thanks to an anonymous referee from *Erkenntnis* for this objection.)

²⁵ I owe E4 to E. J. Coffman.

²⁶ Objection: Suppose that Haji supplements the counterexample to include this piece of information: Yasmin has no conception of W1. Doesn't this imply that she'll have no conception of YC? If so, then it doesn't seem as if Yasmin could have 'good reason' to believe anything about W1 and its actualization. Thus, Yasmin wouldn't meet E4, and isn't morally responsible for YC.

according to E4, it's plausible that Yasmin is morally responsible for YC. Thus, we have a plausible reason to think that 5H is false.

Here, I can think of two potential objections to my claim that Yasmin is morally responsible for YC. Here is the first:

5H's conditional is a necessary truth since it's a fact about what happens in W1 and what's true about W1 is true in all possible worlds. That is, it's true in all possible worlds, that, in W1, Yasmin donates at *ts*. If that's so, and Yasmin is, as you've argued, partly morally responsible for the truth of 5H's conditional, then it looks like you are committed to the claim that a human person can be (at least in part) morally responsible for a necessary truth. If this is your argument, then while you may have saved Rule B, you've undermined Rule A and have, therefore, not saved the Direct Argument more generally. What a curious upshot of your argument!²⁷

But I reply that I've claimed only that Yasmin is, at least in part, morally responsible for the truth of YC. Notice that YC is *not* world-indexed; yet, Haji claims that YC is the fact expressed in 5H's conditional. So, either 5H's conditional is world-indexed (and, so, a necessary truth), or it isn't. If it isn't, then there's no problem with my argument thus far; Yasmin is, at least in part, morally responsible for the truth of YC, and is, thereby, at least in part, morally responsible for the truth of 5H's conditional. If 5H's conditional *is* world-indexed (and, so, a necessary truth), then YC is *also* supposed to be world-indexed. If that's right, then I'm forced to concede, on pain of giving up Rule A, that Yasmin is *not* (at least partly) morally responsible

Reply: Yes, Haji could supplement his counterexample in this way, but I don't think it would help his case. Notice that I don't say that Yasmin *reasonably believes* anything about W1. All I say is that she has good reason to believe something about W1 given that she has good reason to believe something about YC. What do I mean when I say she 'has good reason to believe' that YC, e.g., is morally permissible? All I mean is that she *has evidence* that YC is morally permissible, whether or not she *realizes* she has this evidence or not. And here I'm thinking of what Earl Conee and Richard Feldman (2004, p. 84) call 'scientific evidence', where 'scientific evidence' =_{df} E is scientific evidence for P provided that E is publicly available and E reliably indicates the truth of P. So, for example, suppose that I am running a criminal investigation; Jones allegedly shot Smith. The fingerprints on the gun Jones allegedly used to shoot Smith are part of my evidence *even if* I never consider checking for Jones's fingerprints. This evidence, it seems to me, gives me *good reason*—even if I never consider it—to believe that Jones used the gun that killed Smith. Something similar is true about Yasmin with respect to W1 and YC.

²⁷ Thanks to Thomas Reed for this objection.

for the truth of 5H's conditional. But, even if so, my attempt to undermine Haji's alleged counterexample hasn't yet been shown to be unsuccessful. For one thing, there's still the matter of 4H and whether or not it's true (which, I'll argue just below, it isn't). For another, if 5H's conditional is world-indexed (and, so, a necessary truth), then it looks as if the fact under the scope of the 'NR' operator in 6H is world-indexed, too. And if so, then contra what Haji says, 6H is *true* since the fact under the scope of the 'NR' operator is a necessary truth, and Hal's Argument(2) is no counterexample to Rule B. Thus, unless Haji wishes to attack Rule A—and we're given no reason to think that he *does*—it's best to read 5H's conditional is a *non*-world-indexed way. If I'm right about this, then it looks like my argument against the truth of 5H is successful, after all.

Here is the second objection:

W1 is a maximal state of affairs; so, Yasmin's donating to UNICEF when she does is a necessary condition for W1's being actual. That is, it's a necessary truth that Hal's creating W1 implies that Yasmin donates to UNICEF when she does. Thus, if Yasmin is in any part directly morally responsible for the truth of YC, then she's at least partly directly morally responsible for a necessary truth. Thus, either Hal's Creation(2) is a counterexample to Rule B, or Rule A is false. Either way, the Direct Argument fails.²⁸

But I reply that this objection, too, says more against Haji's alleged counterexample than it does against my attempt to refute it; for, if YC is a necessary truth, then Hal's Creation(2) is really a deterministic set-up. I say that Hal's Creation(2) is a deterministic set-up because, given the fact that Yasmin's donating to UNICEF is a necessary condition on W1, it just follows that if Hal-2 creates W1, then this determines that Yasmin will donate to UNICEF when she does. By hypothesis, Yasmin isn't in any part directly morally responsible for the fact that Hal-2 actualizes W1; thus, Yasmin has no influence as to whether or not Hal-2 actualizes W1. But according to the set-up, Yasmin's donating to UNICEF when she does is supposed to be a

²⁸ Thanks to a referee from *Philosophical Explorations* for this objection.

libertarian free action, which means, among other things, that her action is supposed to be such that it *isn't* determined to happen by antecedent conditions over which she has no influence. So, if YC is a necessary truth, and Yasmin has no influence over the antecedent conditions, then Hal's Creation(2) is just an instance of determinism and Yasmin's act is *not* a libertarian free action. Thus, if YC is a necessary truth, then Hal's Creation(2) fails to be motivated the way Haji claims that it is; for it simply asserts that someone can be morally responsible for an action that is determined by antecedent conditions over which they have no influence. This envisaged objection, then—even if it undermines my argument against 5H's truth—fails both to save Hal's Creation(2) and undermine the Direct Argument more generally.

Moreover, the above objection helps motivate a further reason to think that Hal's Argument(2) fails to provide a counterexample to Rule B; for the above objection helps illuminate reasons to think that 4H is false. To see why 4H is plausibly false, let's take a look at that premise in some detail. Just what does 4H say? 4H says, at least, that Hal-2 creates, or, better, *actualizes*,²⁹ W1. Now, just what does W1 include? Well, according to Hal's Creation(2), W1 *includes* Yasmin's donating, at ts, to UNICEF (this is one lesson from the second objection considered above). Let's call Yasmin's so donating 'A'. So, W1, I take it, includes A.

Before moving on to say just why it is that I think 4H is problematic (indeed, I think 4H is false), let's take a step back for a moment and think about what it means for Hal-2 to 'actualize' W1. According to Alvin Plantinga, in his well-known *The Nature of Necessity*, we

²⁹ I say 'actualize' is better than 'create' because I assume that possible worlds exist necessarily, but aren't necessarily actual. Thus, since they exist necessarily, they are in no need of being 'created'. Rather, they are in need of being made actual.

can think of ‘actualizing’ in, at least, two ways.³⁰ There’s the *weak* sense of actualization and the *strong* sense of actualization. What does it mean to *strongly* actualize some state of affairs (A, say)? To actualize some state of affairs in the *strong* sense, Plantinga says, is to “*cause* [A, or whatever] to be actual” (Plantinga, 1974a, p. 173).³¹ So, if Yasmin A’s, she *strongly* actualizes A since she *causes* it to be the case that A is actual.³² To actualize something, S, in the *weak* sense, is to perform some action A such that, if you were to perform A, S would obtain.³³ Here’s a way to think about it. Suppose Jones kills Smith with his revolver, and call this action ‘R’. Now, prior to Jones’s killing Smith, there is another man, Black, a revolver salesman, who wants Smith dead. Unfortunately for Black, however, he’s incapable of firing revolvers because he’s a quadriplegic. Even so, Black knows that his friend Jones wants Smith dead, too. And he knows, further, than if he were to sell Jones a revolver, Jones *would* kill Smith. So, Black sells Jones a

³⁰ See, especially, pp. 169 – 174.

³¹ On pages 169 – 174 of the cited volume, Plantinga’s discussion is with respect to which (possible) worlds God can create (or, actualize). His argument about which (possible) worlds God can create hinges on whether or not there are any libertarianly free creatures and whether or not they have any creative function. If they do, it looks like which (possible) worlds God can create will depend heavily on which states of affairs libertarianly free creatures cause to be actual; that is, which states of affairs libertarianly free creatures *strongly* actualize. Thinking about this is quite instructive for the case at hand.

³² Here I can imagine an objection by philosophers in the ‘non-causal’ camp (e.g. Carl Ginet). The non-causalist I imagine might say something like the following: “Suppose non-causalism about free will is true. I act freely but don’t cause my action. Then I don’t *strongly* actualize my action (because to strongly actualize something is to cause its obtaining). Do I, then, merely *weakly* actualize my action?” But I don’t think this question should worry the proponent of the distinction between strong and weak actualization. I don’t think this question should worry the proponent of the distinction between strong and weak actualization because I think this distinction can be thought of—though, in an admittedly murky way—non-causally. If non-causalism about free will is true, then all we’d have to say is something like: to strongly actualize a state-of-affairs, S, is to ‘bring it about’ that S is the case (where ‘bring it about’ is non-causal; cf. Ginet’s (2007), e.g.). To *weakly* actualize a state-of-affairs, S, is to ‘bring it about’ that one performs some action A such that, if one were to ‘bring it about that’ A, S would obtain. I think any notion of actualization is much clearer on causal theories, but, even so, if any sense of actualizing can be made on the non-causal theories of free will, making a distinction between strong and weak forms of actualization provides no further difficulty. Thanks to David Palmer for pressing me, here.

³³ Here is how Plantinga puts it. If thinking about how God can actualize certain worlds with creatures who are free to act, then God’s weakly actualizing some world W is “(roughly) whether for each world W there is something he [i.e. God] could have done—some series of actions he could have taken—such that, if he had, W would have been actual” (Plantinga, 1974a, p. 173.).

revolver; Jones, then, kills Smith. Now, does Black *strongly* actualize R? No; how could he? Black doesn't *cause* R to be actual; he's incapable of doing any such thing. But what Black *does* do is actualize R in the *weak* sense. Black weakly actualizes R by strongly actualizing Jones's being in the circumstances in which Black knew he'd kill Smith. However, Black does *not* cause it to be the case that R is actual. Only Smith is responsible for *that*.

Now, what's all this got to do with Hal-2 and Yasmin? Quite a lot, I say. To see why, let's zoom in on A and have a closer look at it. What do we know about A? We know at least this much: it is supposed to be libertarian free action, one for which Yasmin is directly morally responsible. Okay; fair enough. Now, can Hal-2 *actualize* A? I guess that depends on what Hal's Creation(2) means by 'actualize'. Hal-2 can't *strongly* actualize A. He can't strongly actualize A because, if he did, then Yasmin wouldn't be directly morally responsible for actualizing A. To be clear, the reason Hal-2 can't actualize A in the *strong* sense is because, if he did, then he'd be A's cause. And if Hal-2 were A's cause, then Yasmin wouldn't be directly morally responsible for Aing (since she wouldn't be A's cause). So, if Hal-2 causes A to become actual, then Yasmin wouldn't be directly morally responsible for actualizing A since she wouldn't actualize A; Hal-2 would. That is, Hal-2 would actualize A, and not Yasmin, if Hal-2 actualizes A in the *strong* sense. But W1 *includes* A. So, if Hal-2 strongly actualizes W1, he would strongly actualize A. And since Yasmin is, *ex hypothesi*, directly morally responsible for A, Hal's Creation(2) *can't* mean that Hal-2 strongly actualizes A; and, if not, then Hal's Creation(2) can't mean that Hal-2 strongly actualizes W1.

So, Hal's Creation(2) doesn't mean to assert that Hal-2 strongly actualizes W1. This means, then, that Hal's Creation(2) asserts that Hal-2 *weakly* actualizes W1. Let's see what happens when we read Hal's Creation(2) in light of Hal-2's weakly actualizing W1. Now, we

said that W1 *includes* A, the act whereby Yasmin (libertarianly) freely donates to UNICEF at ts. But, again, this means that A is a necessary condition of W1; for, if W1 had lacked A, then W1 would have been some other possible world, W2, say. So, A's occurring is a necessary condition for W1's obtaining, and since Yasmin is directly morally responsible for A, it is Yasmin who strongly actualizes A. This means that it is (we can safely assume) Yasmin who, at least in part,³⁴ strongly actualizes W1, too. And, if so, then, since A is a necessary condition of W1, and W1 owes its obtaining (at least in part) to Yasmin's having strongly actualized A, a necessary condition of Hal-2's weakly actualizing W1 is Yasmin's having strongly actualized A.

Moreover, if E4 (or some condition like it) is right, then, since Yasmin has epistemic justification to believe that Aing is morally good or permissible, she has epistemic justification for believing that actualizing W1, at least in part, is morally good or permissible. And if so, then she has epistemic justification to believe that Hal-2's weakly actualizing W1 is, at least in part, morally good or permissible (for she believes that actualizing A is morally good or permissible). The upshot, then, is that when we read 4H as NR(Hal-2 weakly actualizes W1), we have good reason to think that 4H is false. We have good reason to think that 4H is false because Yasmin is at least partly directly morally responsible for the fact that Hal-2 weakly actualizes W1. She's at least partly directly morally responsible for the fact that Hal-2 weakly actualizes W1 because her strong actualization of A is a necessary condition for Hal-2's weakly actualizing W1. Or, put another way, Yasmin is partly directly morally responsible for W1's actualization because she's directly morally responsible for a part of W1, namely, A. Indeed, given that Yasmin's donation is a libertarian free action, it *must* be the case that she has some influence over the factors that

³⁴ I say 'at least in part' because, plausibly, there are other free agents who are morally responsible for other actions that are necessary for the existence of W1. Insofar as these other agents, at least in part, strongly actualize W1 (by Bing, or Cing, or whatever), then these agents will bear some moral burden for W1's being actual. This conclusion, it seems to me, is unsurprising if one views free agents as world-portion actualizers.

determine or entail her having so donated. And, given the foregoing argument, it must be the case that if Yasmin's donating is a libertarian free action for which she is directly morally responsible, it follows that she bears at least some direct moral responsibility for the conditions which necessitate her action, viz., Hal-2's weakly actualizing W1.

Now, my claim that Yasmin is partly directly morally responsible for W1's actualization because she's morally responsible for a part of W1, commits me to something like the following:

PR: If a person, S, bears direct responsibility for a part of the way things are, then S bears part of the direct responsibility for the way things are.

But Haji might be able to challenge a principle like PR. Haji might be able to challenge a principle like PR by arguing that in order for a person to be even partly directly morally responsible for something (e.g. an act, a state of affairs, etc.), that person would have to at least be capable of grasping, or comprehending, that thing. So, in the given context, for Yasmin to bear any part of the direct moral responsibility with respect to the actualization of W1, she'd have to at least be capable of grasping, or comprehending, W1. But W1 is a maximal state of affairs; plausibly, no finite mind is capable of grasping, or comprehending, a thing like that. And, if not, then Yasmin can't be in any way morally responsible for W1's actualization.³⁵

But I think this objection to PR doesn't work; its epistemic requirements are too strict. For, suppose that I am a mason, tasked with the making, and laying, of a single brick that's part of the foundation of a beautiful mansion. And suppose that my creating and laying this brick is, other things equal, a morally praiseworthy action. Now, suppose, also, that all I know how to do, at least when it comes to building things, is make and lay bricks; I don't know anything about engineering, blue-prints, or anything else relevant to the task of building a house. Moreover, I'm not so much as *capable* of knowing such things (for whatever reason). According to PR, I'm at

³⁵ Thanks to E. J. Coffman for raising this objection and pressing me here.

least partly directly morally responsible for the building of the house; I'm partly directly morally responsible for the fact that things are such that a beautiful mansion is built.

Now, suppose that the homeowner, the person who designed the house and commissioned me to build and lay the single brick, came to me and thanked me for the fact that her house is built. That is, suppose the homeowner praised me on the basis of her home's being built. Is the homeowner out of line? Has she gotten her wires crossed with respect to whom she ought to praise for the building of her house? I don't see any obvious reason to think the answer to either of these questions is 'yes'. Indeed, it seems perfectly natural for the homeowner to thank me since her home's being built the way it was (we may assume) wasn't so much as possible without my having constructed the relevant brick. And this seems true to me regardless of whether or not I understood (or could understand) the larger state of affairs of which my constructing the brick was a part. So, this objection to PR rests, I think, on an epistemic requirement for moral responsibility that's too strict; thus, I think this objection to PR fails.

But, perhaps there's a better objection to PR in the offing. Consider the following potential counterexample to PR:

Sara does cutting edge scientific research. She knows her craft well. She fulfills the requisite requirements to be morally responsible for her research, and reports her research to her boss, Ted. Unbeknownst to Sara, however, Ted is a mastermind controlling a large conglomerate of labs. Ted uses Sara's work, along with the work of many other scientists (whose work we can safely assume Sara would not grasp), to create a virus that, when released, turns half of the world into flesh-eating zombies. Sara bears moral responsibility for a part of the way things are: namely, that her research took place. But Sara doesn't bear part of the moral responsibility for the way things are. It would be inappropriate to blame Sara for the zombie outbreak. Plausibly, Sara is non-culpably ignorant, and so gets off the hook.³⁶

³⁶ This case is from Kyle Fritz.

I'll call this the 'Zombie Case'. I'm not sure whether or not this is a genuine counterexample to PR;³⁷ even so, it certainly makes PR seem much less plausible than it did before considering the case. So, allow me to amend PR in a way that, I think, gets around the Zombie Case but still allows for Yasmin to have been, at least in part, morally responsible for the truth of 4H.

Consider:

PR*: If a person, S, bears direct responsibility for a part of the way things are and has epistemic justification to believe [borrowing from E4, a plausible E-condition on MR] that the part of the way things are for which she bears direct responsibility will contribute to the relevant larger whole, then S bears part of the direct responsibility for the relevant larger whole.

The Zombie Case doesn't undermine PR* because, according to the case, it is "unbeknownst" to Sara that anything untoward will come of her research. Moreover, I think we can assume, by the details of the case, that Sara doesn't have epistemic justification to believe that anything untoward will come of her research, either. But *Yasmin*, from the Hal's Creation(2) case, *does* meet PR*—she has justification to believe that her donating to UNICEF contributes to the fact that the actual world is W1 and not some other world. So, Yasmin is still partly directly morally responsible for Hal-2's having weakly actualized W1.

Now someone might object as follows.

Wait. You say that Yasmin is, on the basis of PR*, partly morally responsible for the fact that the actual world is W1 and not some other world. But, suppose that the zombie outbreak from the Zombie Case is a part of W1. Doesn't what you've just argued commit you to the claim that Yasmin is, at least in part, morally responsible for things being such that there's a zombie outbreak?

My reply to an objection like this is, in a word, 'No'. I say that I am *not* committed to the claim that Yasmin is, at least in part, directly morally responsible for things being such that there's a

³⁷ And even if it is a genuine counterexample to PR, notice that I hedge a bit about PR. I don't say that my argument against 4H's truth depends on PR; what I say is that it depends on something *like* PR. So, even if the Zombie Case is a genuine counterexample to PR, it still seems to me there's a *successful* principle, like PR, in the offing, upon which my argument for 4H rests. Just below, I try to give such an account of an amended PR.

zombie outbreak because things being such that there's a zombie outbreak isn't the *relevant* larger whole of which Yasmin's donating to UNICEF plays a part. While I grant that Yasmin is, at least in part, directly morally responsible for W1's being actual, I do not grant that she bears direct responsibility for *every one* of the states of affairs that make up W1's being actual; or that W1 includes all of the states of affairs that it, in fact, includes. So, if we combine both the Zombie Case and Hal's Creation(2), I argue only that both Sara and Yasmin are partly directly morally responsible for the fact that W1 is actual (and, so, they're partly directly morally responsible for the fact that Hal-2 weakly actualizes W1). I do *not* argue that *either* person—Yasmin *or* Sara—is in any way directly morally responsible for the fact that there's a zombie outbreak.³⁸

³⁸ Readers might have noticed that the Zombie Case seems a potential counterexample to TDM. I handle this potential objection to TDM in chapter six. What I say, there, in defense of TDM, is very similar to what I say, here, in defense of PR*. For now however, I'll handle the following objection:

Objection: Your criticism of 4H appears to depend on the notion of *partial responsibility for a whole* that includes these features:

1. One can be partially responsible for a whole by performing an action that is part of that whole even though one has no grasp of the whole.
2. One can be partially responsible for the existence of a whole even though one does not have partial responsibility for *all the parts* of that whole.

1 and 2 are jointly problematic. 1 seems to lead to the odd result that by doing something one is thereby partly responsible for a world that contains horrible atrocities (e.g. the Zombie outbreak), even though one has no knowledge of these atrocities and no way of knowing one's acts contribute to such a world. That seems wrong. 2 seems to be employed in order to take away the sting from this by insisting that partial responsibility for the whole does not entail responsibility for its parts. The problem here is that no explication of 'responsibility for a whole' is given that allows us to understand how one can be responsible for a whole without being responsible for its parts. Hence, the crucial notion is neither properly explicated nor motivated in such a way as to afford a compelling critique of 4H.

Reply: I grant that my critique of 4H relies on both 1 and 2, but I find neither to be especially problematic. With respect to the alleged problem with 1, consider one possible world, the actual world. I think that possible worlds are maximal states of affairs, which are, themselves, extremely large conjunctions. So, call the conjunction that is the actual world '(x & y)'. Suppose that I am morally responsible for the fact that x is true. And suppose that y includes, among other things, Hitler's having slaughtered millions of innocent people. Thus, (x & y), the actual world, includes something I'm responsible for, and Hitler's having slaughtered millions of innocent people. I think I am at least partly directly morally responsible for the fact that (x & y) is actual, that it's *true* that (x & y). Why do I

So, it seems to me that PR*, or something like it, is true. At any rate, PR* is *plausibly* true. And if PR* is plausibly true, then we have good reason to think it's at least plausible that Yasmin is partly morally responsible for W1's actualization because she's morally responsible for a part of W1 (namely, A). But our having good reason to believe *this* means that we have good reason to think it's plausible that 4H is false. Thus, I conclude that we have good reason to think it's plausible that 4H is false.

We have good reason to think it's plausible that 4H is false. And if 4H is false, then, even if we grant that 6H is false, Haji has failed to give a successful counterexample to Rule B. To have a successful counterexample, we need all *true* premises and a false conclusion. But, here, we have reason to believe that Hal's Argument(2) gives us two *false* premises (both 5H and 4H), and a false conclusion (6H). Thus, I conclude that Haji's Hal's Creation(2) fails as a counterexample to Rule B.

VI. Shabo's Attempt to Provide a Counterexample to (Logical Versions of) Rule B

In his "Against Logical Versions of the Direct Argument: A New Counterexample," Seth Shabo sets out to show that *noncausal* views of Rule B fail. Where a 'noncausal' view of Rule B depends on the claim that non-responsibility transfers over the material conditional *as a matter of*

think this? I think this because the truth of (x & y) depends on x's being true, and I'm directly morally responsible for the fact that x is true. And I think that if I'm directly morally responsible for something that the truth of a proposition depends on, then I'm at least partly directly morally responsible for the truth of the proposition. Now, I grant that this latter claim is no doubt contentious. But I think many will agree that it's at least *prima facie* plausible, and this is enough to motivate 1.

With respect to 2, I think the brick-layer example should provide sufficient explication of how one might be partially directly morally responsible for the whole, but not *all* of the parts of the whole. I'm fully responsible for the brick that I make and partially responsible for the fact that the beautiful mansion is built. I am *not*, however, responsible for the brick that Jones makes, or the brick that Smith makes, etc. Or, more to the point, I am (we may assume) fully morally responsible for the writing of this dissertation and partially morally responsible for the fact that the actual world is actual. I am *not* morally responsible for the fact that Hitler slaughtered millions of innocent people even though that's a fact about the actual world, a world I'm partially morally responsible for. (Thanks, again, to an anonymous referee for *Erkenntnis*.)

logic, Shabo thinks that he has devised a counterexample that shows noncausal views of Rule B fail. In what follows, I'll attempt to analyze and, ultimately, undermine Shabo's alleged counterexample to the noncausal view of Rule B.

To begin, recall Rule B:

Rule B: $NR(p \supset q), NRp \vdash NRq$

Now, consider:

Bad Angle [BA]: Suppose that Jed sees a perfect opportunity to dispatch Kenny, a nettlesome business rival. Jed knows that Kenny regularly hikes on a secluded trail, and he plans to dislodge a large boulder from a height of thirty feet when Kenny reaches the switchback below. Unbeknownst to Jed, he and Kenny are both visible to a park ranger, who watches intently through high-powered binoculars as Jed swiftly works to dislodge the boulder, which comes loose at $t1$. As it happens, the ranger's vantage point is skewed, and at $t2$ he believes—incorrectly, as it happens—that Kenny is not in the boulder's path at $t2$. (Shabo, 2010a, p. 243)

Let 'r' stand for the proposition 'that, at $t2$, the ranger forms the false belief that at $t2$ Kenny is not in the boulder's path'. Let 's' stand for the proposition 'that Kenny *is* in the boulder's path at $t2$ '. Let 'q' stand for the proposition 'that the boulder hits Kenny at $t3$ ', and 'L' stand for 'a complete statement of the laws of nature. Now, given BA, the conjunction of L and s materially implies that the boulder hits Kenny at $t3$. Moreover, r implies s. Thus, the conjunction of L and r materially implies q. More formally, BA gives us:

[*]: $r \supset s, L \& s \supset q \vdash L \& r \supset q$.

Where 'L & r' is 'p', [*] concludes that p implies q. Now, consider:

4*. $NR p$

5*. $NR(p \supset q)$

6*. $NR q$

4* - 6*, is supposed to be a substitution instance of Rule B, one that is generated by BA.

If this a successful substitution instance, then it will follow that Rule B is false, and BA is a

successful counterexample to noncausal, or logical, forms of Rule B. Shabo believes that 4* - 6* gives a successful counterexample to Rule B because, clearly, someone *is* responsible for q. *Jed* is responsible for q. This is true, Shabo thinks, because even if nobody is even partly responsible for the conjunction the laws of nature and the proposition ‘that, at *t*₂, the ranger forms the false belief that, at *t*₂, Kenny is not in the boulder’s path’—that is, *p*—and nobody is even partly morally responsible for the fact that *p* materially implies q, it surely doesn’t follow that nobody is even partly morally responsible for q. So, Shabo concludes that BA is a counterexample to Rule B.

I’d like to give three responses to this alleged counterexample. Here is the first. Following Widerker and Schnall (forthcoming), I argue that it’s plausible to think that 4* is false. For, given that van Inwagen understands the notion of someone’s being ‘partly responsible’ for something as to include being responsible for a *logical part* of something,³⁹ there’s good reason to think that Jed is partly responsible for the fact expressed in 4*. To see why, recall that 4* says that no one is now, or ever has been, even partly morally responsible for the fact that (*L* & *r*)—that is, *p*—where ‘*L*’ stands for the laws of nature, and ‘*r*’ stands for the proposition ‘that, at *t*₂, the ranger forms the false belief that at *t*₂ Kenny is not in the boulder’s path’. So, as Widerker and Schnall note, it’s true that “[*L* & *r*] partially in virtue of the ranger’s mistakenly believing that the boulder will *not* hit Kenny” (Widerker and Schall, p. 15, their emphasis, my insertion). They continue:

But mistakenly to believe that the boulder will not hit Kenny means to believe that the boulder will not hit Kenny, when in fact the boulder *will* hit Kenny. Thus [*L* & *r*] is true partially in virtue of its being the case that the boulder will hit Kenny. But since Jed is responsible for the boulder hitting Kenny, it follows that Jed is partially responsible for [*L* & *r*]. (Ibid., their emphasis, my insertion)

³⁹ See van Inwagen, p. 243 n. 28.

So, the idea is that, since Jed is responsible for a part of r —namely, he’s responsible for the fact that there’s a particular event about which the ranger forms a false belief—and r is a part of $(L \ \& \ r)$, it follows that Jed is partially responsible for $(L \ \& \ r)$. Thus, they conclude that 4* is false and BA is not a counterexample to Rule B.

I like this response, but I think Shabo has already provided an attempt to deflect this sort of move. To be clear about how Shabo has anticipated the above sort of objection, notice that Jed’s shoving the boulder in Kenny’s path is a logical consequence of the ranger’s forming a false belief about the boulder and whether or not it will hit Kenny. The ranger’s mistaken belief, after all, implies that there’s a boulder in Kenny’s path; so, the boulder’s being in Kenny’s path is a logical consequence of the ranger’s having a false belief about the about the boulder and whether or not it will hit Kenny. But, Shabo argues that, if it’s in virtue of *this* fact—the fact that Jed’s hitting Kenny with the boulder is a logical consequence of the ranger’s having the relevant false belief—that Jed is partially responsible for the r , and so $(L \ \& \ r)$, then the notion of responsibility being invoked “encompasses too much” (Shabo, 2010a, p. 246).

Shabo thinks the notion of responsibility being invoked in Widerker and Schnall’s response (or, anyway, this *type* of response) encompasses too much for the following reason. For, suppose that determinism is true. If determinism is true, then, necessarily, $(P \supset (L \supset p))$. And, if this is right, then the move to establish NR $(L \supset p)$ from NR P by Rule B is unnecessary. To be clear, the idea, here, is that the move from premises 3 and 4 to 5 of the Direct Argument via Rule B, is unnecessary. It’s unnecessary, Shabo claims, because $(L \supset p)$ is a logical consequence of P , the remote past. And if it’s unnecessary to establish NR $(L \supset p)$ by using Rule B, then NR $(L \supset p)$ is a mere assertion on the part of the incompatibilist. This, Shabo thinks, is (or would be) an illicit move. “Whatever else,” Shabo writes, “the argument’s

defender is not entitled simply to assert $[NR (L \supset p)]$ at this stage in the argument, for that is something she is supposed to *derive*” (Ibid., p. 245, my insertion). Thus, I think Shabo would conclude that Widerker and Schnall’s response doesn’t work.

So, Shabo thinks that, whatever else it might mean, *being partially responsible for something* can’t mean *being responsible for p or any of p’s logical consequences*. The key, I think, is the idea that being responsible for *just any* of p’s logical consequences says too much. I think we should grant Shabo this point. For, I think that what Shabo must mean is that, to be at least partly responsible for p, a person has to be responsible for p and some *relevant* logical consequence of p. The tricky part, here, is to figure out what counts as ‘relevant’. Obvious cases of the appropriate sort of ‘relevancy’ aren’t hard to pick out. For, p, itself, is a relevant logical consequence of p. No problem there. But what about less obvious cases? Can we give plausible examples of *those* (whatever they may be)? It seems to me that we can.

I think a person is responsible for a conjunction if she’s responsible for one of the conjuncts.⁴⁰ Y, for example, is a logical consequence of (X & Y). And if someone is responsible for Y, it seems plausible to me that she should be at least partly responsible for the fact that (X & Y). So, obviously enough, being responsible for *some* of p’s consequences is sufficient for being at least partly responsible for p. Even so, it is very difficult to see what *all* counts as a relevant consequence of p, though I do think we have some intuitive notions about what counts and what doesn’t.

I think it’s fairly intuitive when something is *obviously* a relevant logical consequence of p (though I don’t claim that it’s obvious what counts as a relevant logical consequence of p). I said, just above, that if a person is responsible for Y, and (X & Y) is true, then that person is at

⁴⁰ And I think something similar holds, *mutatis mutandis*, for responsibility with respect to disjunctive facts.

least partly responsible for the fact that $(X \ \& \ Y)$ is true. I think Y is obviously a relevant logical consequence of $(X \ \& \ Y)$. To see that this is so, consider a case from van Inwagen (1983). Suppose that Smith kills the older of the two Jones twins. And suppose, further, that the younger of the two is killed by a random lightning strike. Let the conjunction expressed by the fact *that the oldest Jones twin is dead and the youngest Jones twin is dead* be expressed as the fact *that both Jones twins are dead*. Now, it's a logical consequence of the conjunction expressed as the fact *that both Jones twins are dead* that *that the oldest Jones twin is dead*. But since Smith kills the oldest of the Jones twins, surely it's plausible to claim that he's at least partly responsible for the fact *that both Jones twins are dead*. In other words, surely it's at least plausible to claim that Smith is at least partly responsible for the relevant conjunction's being true because he's responsible for the truth of one of its conjuncts.

If this is right, then I think we now have reason to side with Widerker and Schnall. To see why, recall

[*]: $r \supset s, L \ \& \ s \supset q \vdash L \ \& \ r \supset q$.

[*] says that, among other things, $r \supset s$. But, given the Bad Angle case, its being true that $r \supset s$ *depends on* s 's being true—since, *ex hypothesi*, r is true. And since s is a necessary condition for r , and Jed is responsible for s , it looks very plausible that he's at least partly responsible for the fact that r obtains. I say that it's very plausible that Jed is at least partly responsible for the truth of $r \supset s$ because I think the principle TDM (from above, p. 21) is true. Recall:

Truth Dependence_{MORAL} [TDM]: For all agents, S , and all propositions, p , if S is directly morally responsible for that which p 's truth depends on (in the sense of 'depends on' in which truth depends on the world), then S is at least partly directly morally responsible for p 's truth.

If TDM is true, and if Jed is at least partly directly responsible the fact that r obtains, then it appears plausible to claim that Jed is at least partly directly responsible for p from premise 4*.

To see why, notice what p expresses. p expresses the conjunction ($L \ \& \ r$); and since we've already said that being directly morally responsible for a conjunct is sufficient for partial direct moral responsibility with respect to the conjunction, it follows that Jed is at least partly directly morally responsible for ($L \ \& \ r$). Jed is, after all, plausibly partially directly morally responsible for r given his direct moral responsibility for s ; so, it looks as if Widerker and Schnall are right after all: Jed is at least partly directly morally responsible for ($L \ \& \ r$), that is, p . Thus, they're right that 4^* is false, and BA is no counterexample to Rule B.

I think similar reasoning applies to 5^* , as well. That is, I think that premise 5^* must be false. I think that 5^* is false because, in the Bad Angle case, $p \supset q$'s truth *depends on* q 's being true, since, *ex hypothesi*, p is true. Moreover, q 's truth depends on Jed. And since Jed is, *ex hypothesi*, responsible for the truth of q (that is, Jed is responsible for Kenny's being hit by the boulder), (and given TDM) he's at least partly responsible for the fact that $p \supset q$. Therefore, Jed is morally responsible for the fact that $p \supset q$ is true since he is responsible for the thing upon which the truth of the conditional proposition depends. Thus, I conclude that 5^* is false and BA is no counterexample to Rule B.

There are further reasons to doubt that Bad Angle is a counterexample to Rule B. To see what further reasons there are to doubt that Bad Angle is a counterexample to Rule B, notice that following follows trivially from the truth of $[*]$:

$$[**]: s \supset r, L \ \& \ r \supset q \vdash L \ \& \ s \supset q.$$

Now, let 'u' stand for ' $L \ \& \ s$ '. Now, consider:

4.** NR u

5.** NR ($u \supset q$)

6.** NR q

4** - 6** is supposed to be a substitution instance of Rule B, and, thus, a counterexample to Rule B given that Jed is morally responsible for q . But I think there's good reason to doubt that 4** - 6** is a successful counterexample to Rule B. To see why, let's first zoom in on 4**. Is 4** really true? It's plausible to think that it isn't; for, it's plausible to think that it isn't true that nobody is now, or ever has been, even partly morally responsible for the fact that u . This is because 'u', we'll recall, stands for the 'the conjunction of L and s'. Now, 's' stands for the proposition 'that Kenny is in the boulder's path at t_2 '. And it's at least plausibly true that Jed is directly morally responsible for the fact that Kenny is in the boulder's path at t_2 . That is, it's plausible that Jed is directly morally responsible for the fact that there's a boulder-and-Kenny-containing path at all. Thus, if Jed *is* directly morally responsible for this fact, then, given TD, it follows that Jed is at least partly directly morally responsible for the conjunction of L and s. That is, it follows that Jed is at least partly directly morally responsible for u . So, plausibly, 4** is false.

I conclude that 4** is plausibly false because it's plausible that Jed is at least partly directly morally responsible for the fact that u . If I'm right about this, then I think there is good reason to doubt the truth of 5**, as well. To see why, recall that s is such that, when conjoined with the laws of nature, it materially implies q , the proposition 'that the boulder hits Kenny at t_3 '. But, given that Jed *intended* for it to be the case that there being a boulder-and-Kenny-containing path at t_2 materially implies that Kenny is hit by the boulder at t_3 , it's plausible that Jed is at least partly directly morally responsible for the fact that $u \supset q$. So, plausibly, both 4** and 5** are false. Thus, 4** - 6** fails to provide a counterexample to Rule B.

If I'm right that 4** - 6** fails to provide a counterexample to Rule B, we have further reason to think that 4* - 6* fails to provide a counterexample to Rule B. To see why, notice that since we get [*] and [**] from BA, we also get:

$$[***]: s \leftrightarrow r, (((L \& r) \leftrightarrow (L \& s)) \supset q) \vdash (((L \& r) \leftrightarrow (L \& s)) \supset q)$$

What's crucial to note about [***] is that it shows us that s and r are material equivalents, and, so, (L & s) and (L & r) are material equivalents. Thus, if what I've just argued is true, and Jed is at least partly directly morally responsible for the conjunction of L and s, then it follows that he's also at least partly directly morally responsible for the conjunction of L and r. That is, if Jed is at least partly directly morally responsible for u, then he's at least partly directly morally responsible for p. Thus, if 4** is false, then 4* is false (and similar reasoning applies with respect to 5** and 5*). I think that 4** is false, so I conclude that 4* is false. Thus, Shabo's Bad Angle case fails to provide a successful counterexample to Rule B.

VII. Schnall and Widerker's Attempt to Provide a Counterexample to Rule B

I'd like to discuss one final attempt to give a counterexample to Rule B, this time provided by Widerker and Schnall (forthcoming). Here is the alleged counterexample:

Molecules: Suppose that Jones fires a bullet at Smith at T1, a fact which, in the circumstances, provides a causally necessary and sufficient condition for the movement of some air-molecules at T3 in his vicinity. (Ibid., p. 15).

From Molecules, we get the following substitution instance of Rule B:

7*. NR (Some air-molecules move at T3)

8*. NR (Some air-molecules move at T3 $\supset$ Jones fires a bullet at Smith at T1)

9*. Therefore, NR (Jones first fires a bullet at Smith at T1).

If 7* - 9* is invalid, then so is Rule B. Widerker and Schnall argue that 7* - 9* is invalid, so they conclude that Rule B is invalid.

Here is why Widerker and Schnall conclude that 7* - 9* is invalid. Premise 7* is true, they claim, because “(in the circumstances) the fact that some air-molecules are moving at T3 is morally neutral, and so is not a fact for which one can be *morally* responsible” (Ibid., p. 16, their emphasis). Moreover, they claim that premise 8* is true because Jones’s firing a bullet at Smith when he does is a necessary causal condition for the morally neutral fact that some air-molecules at T3, and this is just a nomological necessity given the setup of the case (and no one is, or can be, even partly directly morally responsible for a nomological necessity). But, they conclude that 9* is false since Jones is responsible for the non-morally-neutral fact that he fires a bullet at Smith at T1. Thus, 7* - 9* is invalid, and so is Rule B.

As Widerker and Schnall admit, Molecules is a variation on Widerker’s Fate case, from above (§III). The salient differences between the two cases, I think, must be these: the fact that falls under the scope of the ‘NR’ operator in premise 7* is morally neutral (whereas, in Fate, all of the facts that fall under the scope of the ‘NR’ operator are morally significant), and 8* is a nomological necessity. It is *these* facts that are supposed to lead us to the conclusion that 7* and 8* are true (since nobody can be morally responsible for something that’s morally neutral, and nobody can be morally responsible for a nomological necessity), but 9* is false (since Jones is responsible for the morally significant fact that he fires a bullet at Smith when he does). Even so—that is, even given this apparent difference between the cases—I think they suffer a similar fate; that is, they both fail successfully to provide a counterexample to Rule B. For, I think that, as in Fate, Jones is morally responsible for the fact expressed in 7*.

To begin my defense of the claim that Molecules fails to provide a counterexample to Rule B, recall that Widerker and Schall claim that the proposition that falls under the scope of the ‘NR’ operator in 7* is a morally neutral fact. Here is my first reason for thinking that

Molecules fails to provide a counterexample to Rule B: there's reason to think that 7* is plausibly false.

To see why, note that if we take a very *fine-grained* view of the fact that falls under the scope of the 'NR' operator in 7*, we don't merely get *that some air-molecules move at T3*. Rather, what we get is something more like *that some air-molecules move out of the way of a bullet intended for Smith at T3*. This more fine-grained view, I think, is less obviously morally neutral; for this isn't *just any* movement of molecules; it's a movement of molecules as part of a larger causal chain of a morally significant action. And if this is right, then if Jones is responsible for the fact *that some air-molecules move out of the way of a bullet intended for Smith at T3*, then Jones is responsible for (the more fine-grained view of) the fact expressed in 7*. Thus, on this view, 7* is false. So, I think it's plausible to conclude that 7* is false, and Molecules fails to provide a counterexample to Rule B.

Here is another reason to think that Molecules fails to provide a counterexample to Rule B: it misses the point of its intended target. The Direct Argument (and so Rule B) is concerned with *past to future* determinism. That is, the Direct Argument rests on an inference rule (i.e. Rule B) about the transfer of non-responsibility for the past to non-responsibility for the future. But 8* is a premise that has the past determined by the future.⁴¹ Perhaps Molecules is a counterexample to a transfer principle about non-responsibility for the future to non-responsibility for the past, but this is irrelevant to the case at hand. Molecules, thus, fails as a counterexample to Rule B.

⁴¹ I get this style of objection from Ted A. Warfield (1996, footnote 2).

VIII. Conclusions

In this chapter I've argued that the alleged counterexamples to Rule B considered herein are merely that: alleged. I've argued that, in each case, except for Fischer and Ravizza's Erosion case and Haji's (2008), the moral agent in the story—anyway, the moral agent whose responsibility we're considering—is morally responsible for the truth of the fact expressed in one or more of the premises of the given substitution instance of Rule B that each case generates. If the foregoing arguments have been correct on that score, then it follows that those cases have failed to provide a counterexample to Rule B.

With respect to Fischer and Ravizza's Erosion case, I argued, following Widerker (2002), that the case misses the point of Rule B, to wit, that Rule B deals with *direct* moral responsibility. And given the Erosion case, it's plausible to conclude—contra Fischer and Ravizza—that *no one* is even partly *directly* morally responsible for the fact *that there is an avalanche that crushes the enemy base at T3*. Thus, I concluded that no one—and, in particular, not Betty—is *directly* morally responsible for the fact *that there is an avalanche that crushes the enemy base at T3*. And since Fischer and Ravizza need it to be true that Betty (or someone) *is* directly morally responsible for this fact, I concluded their Erosion case fails as a counterexample to Rule B.

With Respect to Haji's (2008) iteration of Hal's Creation, I concluded that the case begs the question against the incompatibilist. I concluded that the case is question begging because it assumes, at the outset, that someone could be morally responsible for an action done in a deterministic universe. But since this is the very question up for debate, then it follows that the case is based on a question begging assumption. So, I concluded that Haji's (2008) Hal's Creation case fails to provide a (non-question begging) counterexample to Rule B.

If the foregoing arguments are successful, it follows that Rule B is safe from the considered alleged counterexamples. I don't know of any stronger attempts to provide a counterexample to Rule B than those I've so far considered; so, I conclude something stronger, namely, that Rule B is safe (for now) from counterexample.

Now, given that Rule B is safe from counterexample, we might be tempted to conclude that Rule B is safe, *full stop*. But this would be too hasty. Rule B is (or so I say) safe from counterexample, but there are other objections to Rule B lurking nearby. In the next chapter I consider these objections, objections that question Rule B's dialectical propriety. It is to that task I now turn.

CHAPTER THREE

Criticisms of Rule B (II): Dialectical Objections

In the last chapter I concluded that Rule B is safe (so far) from counterexample. But this is not the end of the story for Rule B; for, even if I'm correct that the counterexamples I took on in the last chapter are unsuccessful, there might be other reasons for denying the rule, and so, rejecting the Direct Argument. In this chapter, then, I wish to analyze three recent attempts to provide reasons for rejecting the Direct Argument based on the criticism that Rule B is *dialectically inappropriate*.⁴²

In particular, I'll focus, in the second section, on John Martin Fischer's contention that Rule B (or, anyway, various of its revisions) lead to a dialectical stalemate. In the third section, I'll focus on Michael McKenna's recent attempt to say "good-bye" to the Direct Argument, an argument he makes that concludes the Direct Argument is dialectically infelicitous because Rule B hasn't been properly established. Lastly, I'll consider Seth Shabo's recent argument that, though McKenna is right about *one particular reading* of the Direct Argument, there is another reading of the Argument that doesn't succumb to his objections. But even so, Shabo concludes that such findings will reduce the role that the Direct Argument should play in the relevant dialectic. In what follows, I'll argue that each of these objections to Rule B (and the Direct Argument more generally) fails.

⁴² I do not wish to suggest that Rule B's alleged dialectical inappropriateness is claimed as a reason to reject the Direct Argument because such inappropriateness shows that Rule B is *false*. Rather, what's being claimed in the objections I consider, below, is that Rule B's dialectical inappropriateness is a reason to reject the Direct Argument as being of any *use* for the relevant debate. In other words, if Rule B is inappropriate for some reason or another, we should conclude that the Direct Argument is dialectically unhelpful.

II. Fischer and Dialectical Stalemates

John Martin Fischer, in his *My Way: Essays on Moral Responsibility*, devotes an entire chapter (viz., chapter 8) to principles regarding the transfer of non-responsibility. In particular, Fischer is concerned with showing that Rule B is invalid by counterexample, and furthermore, that any attempt to revise Rule B in light of successful counterexamples leads to, what he calls, a “dialectical stalemate” (Fischer, 2006, p. 164). Put another way, Fischer argues that Rule B is invalid, and no (as yet devised) revision of Rule B is such that it is dialectically helpful to the debate concerning the compatibility of moral responsibility and causal determinism. The upshot, then, is that the Direct Argument is infected with dialectical impropriety; thus, the Direct Argument is of no (or very little) use to the debate about whether or not determinism is compatible with moral responsibility.

But, I think that Fischer is mistaken; for, I think that despite whatever the case may be for *revisions* of Rule B, Rule B, itself does not lead to dialectical stalemates. And since, as I argued in the previous chapter, Rule B is safe from counterexample⁴³, it follows that there is no need for a revision to the Rule, and so no need to think there is any hint of dialectical impropriety within the Direct Argument. So, in this section, I intend to argue that Fischer’s assertion that there is dialectical impropriety within the Direct Argument rests on a mistake, viz., that Rule B is invalid by counterexample.

I begin with a summary of Fischer’s case against Rule B and its revisions. According to Fischer, Rule B (or, as he calls it, ‘Transfer NR’) is invalid; it’s subject to counterexample. To see that this is so, note Fischer’s rendering of Rule B:

Transfer NR: If

1. p obtains and no one is even partly morally responsible for p; and

⁴³ In chapter six I’ll argue something stronger: counterexamples to Rule B are *impossible*.

2. if p obtains, then q obtains, and no one is even partly morally responsible for the fact that if p obtains, then q obtains; then
3. q obtains, and no one is even partly morally responsible for q. (Ibid., p. 160)

Thus, Transfer NR. To see that Transfer NR is invalid, recall Erosion:

Erosion: Imagine that Betty [a soldier charged with destroying an enemy fortress] 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 at T2, starting a natural avalanche that would have crushed the enemy camp at T3. (Fischer and Ravizza, 1998, p. 157)

Erosion is purported to be a counterexample to Transfer NR because while it's true that no one is even partly morally responsible for the fact that the glacier is eroding, and no one is even partly morally responsible for the fact that, if the glacier is eroding, then there will be an avalanche that crushes the enemy base at T3, it's *not* similarly true that no one is even partly morally responsible for the fact that there is an avalanche that crushes the enemy base at T3. For, *Betty* is thought to be morally responsible for this fact. Thus, Transfer NR is invalid.

But if Transfer NR is invalid, then what is the proponent of the Direct Argument to do?

One thing she might do is amend Transfer NR as follows:

Transfer NR': If

1. p obtains and no one is even partly morally responsible for p; and
2. (i) p is part of the actual sequence of events *e* that gives rise to q at T3,
(ii) p is a part of *e* that is causally sufficient for the obtaining of q at T3, and
(iii) no one is or ever has been even partly responsible for (2i and ii);
then
3. q obtains at T3, and no one is or ever has been even partly morally responsible for q. (Fisher, 2006, p. 161)

But as Fischer points out, Transfer NR' won't do. Transfer NR' won't do because there is a counterexample in the offing, one that is an almost exact mirror of Erosion. Consider:

Erosion*: [The details are the same as Erosion except that, in this case,] the conditions of the glacier *do* actually cause the ice and rocks to break free, triggering an avalanche that arrives at the fortress precisely at the same time as the independent avalanche triggered freely by Betty. Each avalanche is sufficient for the destruction of the enemy fortress. (Ibid.)⁴⁴

We can see that Erosion* is a counterexample to Transfer NR' because, even though no one is even partly morally responsible for the fact that the glacier is eroding, and no one is even partly morally responsible for the fact that the glacier is eroding and its erosion is, in fact, causally sufficient for the fact that an avalanche crushes the enemy base at T3, it does not follow that no one is even partly morally responsible for the fact that an avalanche crushes the enemy base at T3. This is because *Betty* is morally responsible for the fact that her free actions are, as a matter of fact, causally sufficient for the fact that an avalanche crushes the enemy base at T3. Thus, Transfer NR' is invalid.

Next, Fischer considers an ingenious attempt by Michael McKenna (2001) to give yet another revision of Transfer NR (i.e. Rule B). In this revision, McKenna draws special attention to the fact that what's at stake in the incompatibilism/compatibilism debate is *not* whether or not agents are (or can be) morally responsible for states of affairs that come about through overdetermination (or counterfactual overdetermination). These are what McKenna calls 'two path' cases, and he concedes that determinism does not rule out cases like these, cases of overdetermination. But, as McKenna notes, van Inwagen, when attempting to establish Transfer NR (i.e. Rule B), gives examples that proceed along only 'one path'. For example, in his *An Essay on Free Will*, van Inwagen gives us the following case. No one is even partly morally responsible for the fact that Plato died in antiquity; no one is even partly morally responsible for

⁴⁴ But, originally formulated in Fisher and Ravizza (1998, pp. 160 – 161).

the fact that, if Plato died in antiquity, then Plato never met Hume; therefore, no one is even partly morally responsible for the fact that Plato never met Hume.⁴⁵ Thus, the thought is that Transfer NR is *prima facie* established on the bases of cases like this, cases that go through ‘one path’. And, as McKenna goes on to explain,

if determinism is true, then the manner in which the facts of the past and the laws of nature entail one unique future is *not* analogous to the manner in which one set of independently existing causally sufficient conditions (for example, erosion) ensures a subsequent event also ensured by *some distinct set* of independently sufficient conditions (i.e., Betty’s action). Assuming determinism, the pertinent facts (consisting in the deterministic order of things) are not *independent* of an agent’s reasons for action, *they constitute them!* Therefore, at a deterministic world involving a typical case regarding a judgment of moral responsibility, the case is relevantly like a one-path, *not* a two path. (McKenna, 2001, p. 45)⁴⁶

Thus, McKenna thinks that the only cases relevant to causal determinism are one-path cases, and that two-path cases (i.e. cases like Erosion, and Erosion*) are irrelevant to the debate.

So, in order to protect Transfer NR from what he takes to be the *relevant* sorts of potential counterexamples (counterexamples that proceed via one-path cases), McKenna gives the following revision to Transfer NR:

Transfer NR’’: If

1. p obtains and no one is or ever has been even partly morally responsible for p; and
2. (i) p is part of the actual sequence of events *e* that gives rise to q at T3;
(ii) p is causally sufficient for the obtaining of q at T3; and any other part of *e* that is causally sufficient for q either causes or is caused by p; and
(iii) no one is or ever has been even partly responsible for (2i and ii);
then
3. q obtains, and no one is or ever has been even partly morally

⁴⁵ From p. 187 of the relevant text.

⁴⁶ Also quoted in (Fischer, 2006, p. 162).

responsible for the fact that q obtains at T3. (Fischer, p. 163)⁴⁷

But, for his part, Fischer doesn't think that Transfer NR'' fares any better than the other revisions to Transfer NR. For, he thinks that Transfer NR'' can't be invoked to *establish* the incompatibility of causal determinism and moral responsibility. Moreover, there is *this* preliminary worry:

Note what is happening with Transfer NR'': the principle entails that when behavior is the result of one causally deterministic sequence, there cannot be moral responsibility, but when the behavior is the result of two or more such sequences, there maybe moral responsibility. But if one causally deterministic sequence rules out moral responsibility, it would seem that two or more would be even worse. Transfer NR'', however, gives us precisely the opposite result. Transfer NR'' seems to entail, for the relevant sorts of normative notions, that two wrongs may well make a right. (Fischer, 2006, p. 164)

The idea, here, is that McKenna's concession that causal determinism does not rule out moral responsibility in cases of overdetermination is quite curious. In cases like Erosion or Erosion*, it's admitted that the moral agent in the story might very well be morally responsible for the consequence universal (e.g. that an avalanche crushes the enemy fortress) even though that very consequence universal is determined to happen. Yet, somehow, if Betty *alone* (i.e. without an ensuring condition) were the only potential cause of the relevant consequence universal, then she wouldn't be morally responsible for it. This seems quite implausible. Thus, there's good reason to doubt the validity of Transfer NR''.

The main worry for Fischer, however, is that Transfer NR'' leads to a "dialectical stalemate." This is, in part, because there are no non-question-begging counterexamples that can be marshaled against a principle like Transfer NR'' (for, any scenario one could describe that meets all of the conditions spelled out in Transfer NR'' could not render a false conclusion if set in a deterministic context). In other words, Transfer NR'' rules out, just by *fiat*, counterexamples set in a deterministic world. Moreover, Fischer thinks that there is a similarly

⁴⁷ Originally from (McKenna, 2001, p. 45). McKenna calls it 'Transfer NR***'.

plausible principle about the transfer of non-responsibility that is *consistent with* compatibilism.

Consider one such principle:

Transfer NRC: If

1. p obtains and no one is even partly morally responsible for p; and
2. if p obtains, then q obtains, and no one is even partly morally responsible for the fact that if p obtains, then q obtains; and
3. on the actual path that leads from p's obtaining to q's obtaining, either this is no factor that at least *prima facie* could be thought to ground moral responsibility, or there is some factor that uncontroversially undermines moral responsibility (e.g., a factor that distorts or impairs the distinctive process of human practical reason); then
4. q obtains, and no one is even partly morally responsible for q. (Ibid., p. 169)

Fischer thinks that Transfer NRC is at least as attractive, and plausible, as Transfer NR'' because it, too, is consistent with one-path cases (cases like van Inwagen's Plato case), but without entailing that Betty is *not* morally responsible for the consequence in cases like Erosion or Erosion*. Moreover, the thought is that Transfer NRC is logically weaker than Transfer NR'' because it *doesn't* rule out compatibilism—indeed, it's perfectly silent on whether or not compatibilism is true. Transfer NR'', however, is not silent about this; Transfer NR'' rules out compatibilism altogether. But, if Transfer NRC can just as easily explain the cases that van Inwagen uses to *establish* Transfer NR (and its revisions), then we seem to be at an impasse. For, both Transfer NR (and its revisions), and Transfer NRC explain the uncontroversial cases, while each pulls in the opposite direction of the other when it comes to the controversial cases. And if we have two competing principles, one weaker than the other (say, P and P*), that both explain some claim C (that no one is morally responsible for the fact that Plato never met Hume, given that no one is morally responsible for Plato's dying in antiquity, and no one is morally

responsible for that fact's implying that Plato never met Hume), then it will be very difficult to establish that P , but not P^* , is the principle we should follow without begging the question against proponents of P^* . And the same is true the other way around. Thus, Fischer concludes that we have a dialectical stalemate.

If Fischer is right that Transfer NR is invalid, and its revisions are either invalid or lead to a dialectical stalemate, then this spells doom for the Direct Argument. For, even if McKenna's principle, Transfer NR'', is valid, it's very difficult to see that it establishes incompatibilism since Fischer's competing principle, Transfer NRC, seems to be just as plausible, and is consistent with compatibilism. Thus, if Fischer is right, the Direct Argument is infected with a principle that leads to a dialectical stalemate. And if so, then the argument is of no (or not much) use in the relevant dialectical context.

But, I say that Fischer is *not* right. That is, irrespective of the validity or dialectical usefulness of revised versions of Transfer NR, Transfer NR, itself, is valid. So, since there is no need to make adjustments to Transfer NR, there is no need to give a reason to think that Transfer NR'' (or whatever) is more plausible (or whatever) than Transfer NRC. These questions are irrelevant because Transfer NR is, itself, valid.

To begin to see why I say this is so, recall what I argued in the previous chapter. Fischer thinks that Erosion is a counterexample to Transfer NR (or, as I'll now go back to calling it: Rule B). But, I argued in chapter two that this conclusion is mistaken. So, let me, now, say more about why I think such a conclusion is mistaken. What I say, here, will also help us see why Erosion* (a purported counterexample to Rule B I *didn't* consider in the last chapter) fails to be a counterexample to Rule B, as well.

Recall, again, Erosion and Erosion*:

Erosion: Imagine that Betty [a soldier charged with destroying an enemy fortress] 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 at T2, starting a natural avalanche that would have crushed the enemy camp at T3.

Erosion*: [The details are the same as Erosion except that, in this case,] the conditions of the glacier *do* actually cause the ice and rocks to break free, triggering an avalanche that arrives at the fortress precisely at the same time as the independent avalanche triggered freely by Betty. Each avalanche is sufficient for the destruction of the enemy fortress.

Now, consider the substitution instance of Rule B that these examples are supposed to generate:

7. NR (The glacier is eroding)
8. NR (If the glacier is eroding, then there is an avalanche that crushes the enemy base at T3)
9. NR (There is an avalanche that crushes the enemy base at T3)

As Fischer has argued, 7 – 9 is supposed to be a counterexample to Rule B because, even though 7 and 8 are true, it's false that no one is even partly morally responsible for the fact that there is an avalanche that crushes the enemy base at T3. I argued in chapter two that this conclusion is mistaken, in part, because he's got Rule B (or Transfer NR) wrong. Notice that in his rendering of Transfer NR, there's no mention of *direct* moral responsibility. But Rule B is *about* what we're *directly* morally responsible for. So, if we think in terms of direct moral responsibility, then it's much less clear that Betty is *directly* morally responsible for the obtaining of the consequence universal in Erosion or Erosion*. It's more likely, I claimed, that Betty is morally responsible for the fact that *she causes* there to be an avalanche that crushes the enemy base at T3.⁴⁸

⁴⁸ And even *more* likely (I say) that what Betty is directly morally responsible for is her *intending to destroy the enemy base*. In any case, she's not directly responsible for the fact that the enemy base is crushed.

But I wish to take a different tack, here; for I wish to make it doubly clear that Betty is not directly morally responsible for the fact that an avalanche crushes the enemy base at T3, even if she's causally responsible for it (as in Erosion and Erosion*). I think that in order for a moral agent to be even partly directly morally responsible for some fact, that fact's being true has to depend on the agent in the relevant way. What *is* the relevant way? This is difficult to say, exactly, but I think the idea I have in mind is fairly intuitive, and has to do, in large measure, with what the fact in question is *about*. That is, I think what a proposition's truth depends on is closely related to what that proposition is about. Here's an example of 'aboutness' with respect to true propositions. Suppose that God exists. If God exists, then it's true, necessarily, *that God exists*. In other words, *that God exists*—this proposition—is necessarily true. Now, upon what does this proposition depend? I think the answer to this question is pretty simple: its truth depends on whether or not God exists. That is, this proposition is *about* God, and depends on God and whether he exists.

Something similar holds for contingent truths, as well. Take my dog Rylie, a Golden Retriever. The truth of the proposition *that Rylie is a Golden Retriever* depends on whether or not Rylie is a Golden Retriever. That is, this proposition is about Rylie and whether or not he's a Golden Retriever. Thus, very trivially, the truth of the proposition *that Rylie is a Golden Retriever* depends on Rylie and what breed of dog he is.⁴⁹

Now, take the proposition *that Betty causes an avalanche to destroy the enemy base at T3*. Upon what does the truth of this proposition depend? I think it must be that the truth of this proposition depends on Betty and whether or not she causes the relevant avalanche. That is, I

⁴⁹ This isn't to say that there's nothing more substantive that can be said about the way in which contingent truths 'depend on' the world. Rather, it's just to say that, regardless of whatever *else* might be said about how truth 'depends on' the world, it's just an obvious fact about truth that truth 'depends on' the world in this very trivial way. That is, truth depends on the way things are; the way things are doesn't depend on what's true. For more on this issue, see Merricks's (2007, 2009, 2011a, 2011b). Also, I say more about this in chapter six.

think that this proposition is about Betty and what she does. So, Betty is directly morally responsible for the truth of *that Betty causes an avalanche to destroy the enemy base at T3* because the truth of the proposition depends, in the right way (i.e. the proposition is *about* Betty and what she does) on Betty and what she does.

Can we say the same for the fact expressed in 3, above? I think that we cannot. For upon what does the truth of *that there is an avalanche that crushes the enemy base at T3* depend? Does it depend on Betty? I don't see how, since, given Erosion and Erosion*, this proposition would have been true even if Betty hadn't acted as she did.⁵⁰ Instead, it seems to me that the truth of this proposition depends on the avalanche, and whether or not it crushes the enemy base at T3. But, in order for Betty to be *directly* morally responsible for some fact, I think it must be that this fact's truth depends on Betty and what she does. Thus, I conclude that Betty is not directly morally responsible for the fact *that there is an avalanche that crushes the enemy base at T3*.⁵¹ So, Fischer is mistaken: Rule B does *not* lead to a dialectical stalemate; thus, the Direct Argument is not infected with dialectical impropriety (or unhelpfulness).

⁵⁰ Objection: You're here confusing two different kinds of dependence, viz., *ontological* dependence, and *counterfactual* dependence. You suggest that a proposition P's truth ontologically depends on an object O-involving state of affairs only if P wouldn't have been true had the O-involving state of affairs in question not obtained. But here's a potential counterexample. Suppose that you commit suicide. Then *that Roger is dead* is true and you'll want to say that this proposition's truth ontologically depends on you and your behavior. Now we can add that if you hadn't committed suicide precisely when you did, an independent process would have killed you instead. But then *that Roger is dead* isn't *counterfactually* dependent on your relevant behavior (by stipulation, that proposition would be true even if you hadn't committed suicide), even though that proposition's truth is *ontologically* dependent on your behavior. If all of this is correct, then what you're here assuming about the connection between ontological and counterfactual dependence is mistaken. (Thanks to EJ Coffman for the objection.)

Reply: But I do *not* want to say that the truth of *that Roger is dead* depends on me and my behavior; rather, it depends on me and whether or not I'm dead (but not how I got that way). In this particular story, if anything depends on me and my behavior, I think it must be the truth of *that Roger commits suicide*. Surely *that* truth ontologically depends on me and wouldn't have been true had I not killed myself. But, if we replace the proposition *that Roger is dead* with the proposition *that Roger commits suicide* in the story now under consideration, then it doesn't at all follow that I'm mistaken about the connection between ontological and counterfactual dependence. Rather, such connection is confirmed.

III. McKenna's Attempt to Say 'Good-bye' to the Direct Argument

In "Saying Good-bye to the Direct Argument the Right Way," Michael McKenna argues, among other things, that Peter van Inwagen's Direct Argument should never have gotten off the ground. That is, McKenna thinks the Direct Argument is dialectically uncharitable, especially "in light of the dialectical context with the compatibilist" (McKenna, 2008, p. 350). The Direct Argument is uncharitable, he thinks, because the cases van Inwagen uses in order to establish the incompatibility of determinism and moral responsibility rely on incompatibilist presuppositions. And this is just to beg the question against the compatibilist. But I do not think the Direct Argument *does* rely on such presuppositions. And, further, I deny that the cases to which McKenna refers are meant to *establish* the incompatibility of determinism and moral responsibility. Rather, I think the results of these cases are *upshots* of applying an established principle (namely, Rule B) to various thought experiments. In this section, I'll argue that if I'm right, then McKenna hasn't said 'good-bye' to the Direct Argument at all, much less 'the right way'.

To begin, McKenna wants to challenge Rule B in a different way from Fischer and Ravizza. Rather than show that Rule B is false by counterexample, he wants to suggest that it's not properly established in the dialectical context with the compatibilist. To properly understand McKenna's criticisms, let's remind ourselves how van Inwagen went about establishing Rule B in the first place. He did this by appealing to two examples, call them 'Snakebite' and 'Plato'. They went as follows:

Snakebite: NR John was bitten by a cobra on his thirtieth birthday,

NR (John was bitten by a cobra on his thirtieth birthday $\supset$ John died on his

⁵¹ If the reader isn't satisfied with this line of reasoning, in chapter six I argue that *even if* the truth of *that there is an avalanche that crushes the enemy base at T3* depends on Betty in the relevant way, Erosion and Erosion* still fail to provide a counterexample to Rule B.

thirtieth birthday), hence

NR John died on his thirtieth birthday.

Plato: NR Plato died in antiquity,

NR (Plato died in antiquity $\supset$ Plato never met Hume), hence

NR Plato never met Hume. (van Inwagen, 1983, p. 187)

But McKenna thinks these cases *do not* establish Rule B's validity in the way that van Inwagen supposes. These cases fail to establish Rule B's validity because they do not "pass through any normally functioning agent who exercises unimpaired deliberative capacities in the production of (allegedly) free action for which he or she is morally responsible" (McKenna, 2008, p. 376). To make the point clearer, McKenna asks us to imagine the following case about former President George W. Bush and his deciding to 'get down' (that is, dance a jig) on national television, something he promised his daughters that he'd do. Now, if Rule B is valid, then we can get the following case:

Bush Deliberates: NR (Bush deliberates about getting down & L)

NR (Bush deliberates about getting down & L $\supset$ Bush gets down),
hence

NR (Bush gets down). (Ibid.)

Or, put another way (where 's' stands for 'Bush gets down'):

NR (P & L)

NR (P & L $\supset$ s)

NR s

Clearly, though, Bush Deliberates cannot be used to *establish* a principle like Rule B; it's just a matter of controversy whether or not Bush, even if he's determined to do it, is morally

responsible for his getting down. But, now consider the following propositions about Bush's being determined to deliberate:

P = The facts long before George Bush was born were thus and so;
g = a causal chain led to the existence of Bush and his deliberative capacities;
r = Bush deliberates about getting down;
s = Bush gets down at a certain time.

From these propositions we get the following conjunction:

DBD: $\square (P \ \& \ L \supset g) \ \& \ \square (g \ \& \ L \supset r) \ \& \ \square (r \ \& \ L \supset s)$

And, again, by Rule A we get:

DBD': $NR (P \ \& \ L \supset g) \ \& \ NR (g \ \& \ L \supset r) \ \& \ NR (r \ \& \ L \supset s)$. (Ibid., pp. 374 – 375)

Now, if Rule B is valid, then we get the following argument from DBD':

- | | |
|---|-------------------------------------|
| 1* . $NR (P \ \& \ L)$ | Given |
| 2* . $NR (P \ \& \ L \supset g)$ | DBD' |
| 3* . $NR \ g$ | 1*, 2*, and Rule B |
| 4* . $NR (g \ \& \ L)$ | 1*, 3* |
| 5* . $NR (g \ \& \ L \supset r)$ | DBD' |
| 6* . $NR \ r$ | 4*, 5*, and Rule B |
| 7* . $NR (r \ \& \ L)$ | 1*, 6* |
| 8 . $NR (r \ \& \ L \supset s)$ | DBD' |
| 9 . $NR \ s$ | 7*, 8*, and Rule B. (Ibid., p. 375) |

It is just here that McKenna thinks there is a problem. Steps 7* through 9 *just are* Bush Deliberates; and we already said that Bush Deliberates cannot be used to establish a principle like Rule B. We cannot use Bush Deliberates to establish a principle like Rule B because it relies on a controversial premise, namely, that if Bush is determined to get down, then he's not

responsible for his getting down. And this is exactly the sort of thing a compatibilist would reject.

So, cases like Snakebite and Plato cannot be used to establish a principle like Rule B. They can't be used to establish Rule B, thinks McKenna, because they are uncontroversial cases in which it's *obvious* that nobody is morally responsible for the consequences that take place. Thus, in order for the proponent of the Direct Argument to succeed in establishing a principle like Rule B—at any rate, in order for the proponent of the Direct Argument to establish Rule B without begging the question against the compatibilist—she will have to give an example that includes a causal chain that passes through a normally functioning agent in the right sort of way.⁵²

And by McKenna's lights, this just isn't possible. Any potential example will have to be like the argument from DBD'; and this argument *can't* be used to establish Rule B since it makes use of Bush Deliberates, a scenario that presupposes that compatibilism is false. And if this is true, then Rule B doesn't have anything going for it. What's more, it *is* true, so Rule B has nothing going for it. The Direct Argument, therefore, should never have gotten off the ground.

I say that the Direct Argument has plenty going for it. But that's because I say McKenna's objection to Rule B fails. In fact I think it is *he* that begs the question. Let me explain.

McKenna argues that, in order to get Rule B off the ground, what's needed is an example that "passes through" the appropriate sort of agent. But why should we think a thing like that? One of the (perceived) strengths of the Direct Argument is that it is supposed to be an appeal to intuition. We see that, if determinism is true, then any event whatever follows necessarily from

⁵² Where 'right sort of way' means that the causal chain passes through a normally functioning agent "who exercises unimpaired deliberative capacities in the production of an (allegedly) free action for which he or she is morally responsible" (McKenna, 2008, p. 376).

the conjunction of the remote past and the laws of nature. Suppose that determinism is true, and the remote past conjoined with the laws of nature entails that it will rain tomorrow at 5pm. Since you are not responsible for the conjunction of the remote past and the laws of nature, and you are not responsible for the fact that this conjunction entails that it will rain tomorrow at 5pm, it seems to follow easily that you are not responsible for its raining tomorrow at 5pm. Now suppose that determinism is true, and the conjunction of the remote past and the laws of nature entails that you will wash your car tomorrow at noon. Since you are not responsible for the conjunction of the remote past and the laws of nature, and you are not responsible for the fact that this conjunction entails that you will wash your car tomorrow at noon, it seems by parity of reasoning that you are not responsible for the fact that you will wash your car tomorrow at noon.

As we've seen, however, McKenna thinks this latter step is question begging. It's question begging because I'm assuming, in the set-up, that compatibilism is false. But, am I? If I am, it's difficult for me to see exactly why I should think I am. McKenna says that the latter argument passes through a normally functioning agent; this is what makes the difference for the validity of the argument since it's precisely at this step where the compatibilist will balk. But, so what? Isn't it up to the compatibilist to tell me *why* the argument's "passing through" a normally functioning agent makes a difference? It can't make a difference to the argument merely because compatibilists think it's possible for a person to be responsible for something she's been determined to do. As we saw with Haji's (2008) attempt to provide a counterexample to Rule B (above, p. 29), this sort of proposal is question begging.

McKenna claims that "buried deep inside these seemingly innocuous claims are claims of non-responsibility that have it that one is not even partly responsible for the fact that one's own deliberations led to one's own actions" (Ibid., p. 377). And it's this that "the compatibilist

should not grant just because the relation is deterministic” (Ibid.). But here I’m puzzled, and for two reasons. First, it’s not the case that such claims are buried deep inside Rule B; rather, it’s that such claims are *an upshot* of Rule B; they’re a result of Rule B’s having been *prima facie* established by uncontroversial cases like Snakebite and Plato. Second, *why* shouldn’t the compatibilist grant that one isn’t responsible for one’s own deliberations if they follow necessarily from the conjunction of the remote past and the laws of nature (something for which no one is even partly directly morally responsible)? I can’t see what the answer to this question might be; but, whatever the answer is, it *can’t* be anything like “because compatibilism is true.” And this, it seems to me, is the answer McKenna gives. Such an answer begs the question against the proponent of the Direct Argument.

To put the point another way, following Ira Schnall and David Widerker (2012; forthcoming), van Inwagen uses cases like Snakebite and Plato in order to show that Rule B has been “*prima facie* established on the basis of uncontroversial cases” (Schnall and Widerker, 2012, p. 34), and once it’s been so established, “the burden of proof rests on the person who opposes its application to the controversial cases, i.e. the compatibilist” (Ibid.). Thus, the challenge for McKenna is to explain *why* Rule B doesn’t apply to controversial cases like DBD’. But, so far as I can tell—and Schnall and Widerker agree—McKenna hasn’t done any such thing; he simply wants to assert that Rule B is not *ultima facie* valid because an example passing through a “normally functioning agent” won’t confirm Rule B without entailing the falsity of compatibilism. But, presumably, at this point, *that* says more about the truth of compatibilism than it does the truth of Rule B.

IV. Shabo and the Fate of the Direct Argument

Seth Shabo (2010b) tends to agree with McKenna that *some versions* of Direct Argument are dialectically improper. That is, Shabo thinks that McKenna is right about *logical* versions of the Direct Argument, where the locution ‘logical versions’ is intended to denote versions of the argument that “focus on the *logical* relationship between earlier states and later states of a deterministic world” (Shabo, 2010b, p. 406). This is because Shabo thinks that these logical versions of the Direct Argument are the versions that rely on Rule B. Thus, Shabo thinks that van Inwagen’s original formulation—the one I’ve been defending—is a “case in point” (Ibid.). And, if so, then the argument I’ve been defending is dialectically improper for the reasons McKenna gives.

What Shabo thinks is *less* clear, is that *causal* versions (that is, versions that focus on the causal relationship between the earlier states and later states of a deterministic world) of the Direct Argument are dialectically inappropriate. Indeed, Shabo thinks that causal versions *are not* dialectically inappropriate; and this because causal versions aren’t susceptible to the sorts of counterexamples to the Direct Argument that McKenna gives. Even so, Shabo thinks that the Direct Argument—thought of in this *causal* way—can do only so much work for the incompatibilist. While Shabo thinks that causal versions of the Direct Argument aren’t subject to counterexample in the way that logical versions of the argument are, he thinks that causal versions are too tendentious to play a “starring role” in the relevant dialectic.⁵³ Thus, while the Direct Argument may not win the day for incompatibilism, it ought not be thought irrelevant to the dialectic.

⁵³ What sort of role can causal versions play in the relevant dialectic? Shabo concludes that, while these versions can’t play a “starring role”, they can play a serious role in (what Shabo calls) the ‘ultimacy argument’. That is, they can play a role in support of arguments that purport to show that humans aren’t the *ultimate source* of their actions done in a deterministic world. And this sort of argument might undermine direct moral responsibility. See, especially, pp. 419 – 423 of Shabo (2010b).

But I think that the Direct Argument, however one wants to think about its notion of the relationship between the earlier states and the later states of a deterministic world, is *not* subject to counterexample. And, moreover, I think that the reasons Shabo gives for thinking that *causal* versions of the argument aren't subject to the same sorts of counterexample that logical versions are, are sufficient to show that, after all, logical versions aren't subject to counterexample either. So, in this section, I wish to analyze and challenge Shabo's claims. For, I think that the Direct Argument remains as powerful as ever and, indeed, should play a starring role in the contemporary debate about the compatibility of moral responsibility and causal determinism.

To begin to see why Shabo thinks that McKenna is right about (at least) logical versions of the Direct Argument, recall:

Bush Deliberates: NR (Bush deliberates about getting down & L)

NR (Bush deliberates about getting down & L $\supset$ Bush gets down),
hence

NR (Bush gets down).

What Bush Deliberates purports to show is that such a substitution instance of Rule B cannot be used to *establish* Rule B. Thus, when thinking about (at least) the logical versions of the Direct Argument (those that rely on Rule B as I've been defending it), it's just a matter of controversy as to whether or not Bush is responsible for his getting down in a deterministic world. Thus, Shabo concludes that McKenna is right about logical versions of the direct argument.

But *causal* versions of the argument don't rely on Rule B; they rely on a variation of Rule B. One plausible variation that Shabo gives is as follows:

Rule B-co: (i) NR_p,

(ii-c) NR (it is the q-states that are actually deterministically produced [i.e. 'caused'] by the p-states) and

(ii-o) The q-states aren't overdetermined by distinct causes; therefore,

(iii) NRq (Ibid., 409. My insertion.).⁵⁴

Shabo thinks that Rule B-co is better than Rule B because it doesn't succumb to counterexamples as easily as Rule B allegedly does. For example, Shabo thinks that Fischer and Ravizza's Erosion case is a counterexample to Rule B. Of course, I've already shown (above in chapter two, and then again in §II of the present chapter) that such a conclusion is mistaken, but we can ignore that for now. For, even if Erosion doesn't provide a counterexample to Rule B, Shabo thinks that any version of the Direct Argument that *relies on* Rule B (so, essentially, any form of the Direct Argument that is merely *logical*) is dialectically improper.

Just as McKenna argues, Shabo thinks that Bush Deliberates helps us to see why logical versions of the Direct Argument are dialectically improper. Recall that Bush Deliberates helps generate the following (where 'P' stands for 'the facts long before George Bush was born were thus and so'; 'g' stands for 'a causal chain led to the existence of Bush and his deliberative capacities'; 'r' stands for 'Bush deliberates about whether to get down at t'; and 's' stands for 'Bush gets down at t'):

DBD: $\Box (P \ \& \ L \supset g) \ \& \ \Box (g \ \& \ L \supset r) \ \& \ \Box (r \ \& \ L \supset s)$.

And by applying Rule A to DBD, we get:

DBD': $NR (P \ \& \ L \supset g) \ \& \ NR (g \ \& \ L \supset r) \ \& \ NR (r \ \& \ L \supset s)$.

Thus, from DBD and DBD', we get the following:

- | | |
|-----------------------------------|--------------------|
| 1*. NR (P & L) | Given |
| 2*. NR (P & L $\supset$ g) | DBD' |
| 3*. NR g | 1*, 2*, and Rule B |

⁵⁴ As with Fischer and McKenna, Shabo refers to Rule B as 'Transfer NR'; so, his causal rendering of Transfer NR is 'Transfer NR-co'. I've changed it to 'Rule B-co' for the sake of expediency.

4*. NR (g & L)	1*, 3*
5*. NR (g & L $\supset$ r)	DBD'
6*. NR r	4*, 5*, and Rule B
7*. NR (r & L)	1*, 6*
8. NR (r & L $\supset$ s)	DBD'
9. NR s	7*, 8*, and Rule B.

But, as you will recall, McKenna's charge is that applying Rule B to steps 7* – 9 is an illicit move since 7* – 9 just is Bush Deliberates, and the conclusion was that Bush Deliberates can't be used to establish Rule B.

So, Shabo agrees that 1* – 9 is enough to show that Rule B is dialectically improper. But, what about Rule B-co? Shabo thinks that 1* – 9 does *not* show that Rule B-co is dialectically improper. To see why not, notice that if we change Rule B to Rule B-co, instead of getting Bush Deliberates, we get:

Bush Deliberates-co: (i) NR (L & Bush deliberates about whether to get down at t);

(ii-c) NR (It is Bush's getting down at t that is actually deterministically produced by his deliberations about whether to get down at t); and

(ii-o) Bush's getting own at t isn't overdetermined by distinct events; therefore,

(iii) NR (Bush gets down at t). (Ibid., 415)

Now, as with Bush Deliberates, the compatibilist will ask the Direct Argument's defender to justify the appeal to Rule B-co for steps 7* – 9 since the compatibilist will just deny that no one is even partly morally responsible for the fact that Bush gets down when he does. Bush, the compatibilist will think, is responsible for that.

But Shabo thinks that if the incompatibilist defends a *causal* reading of the Direct Argument—and, so, relies on Rule B-co—then, she has resources available to her that the defender of Rule B doesn't have. To see why, Shabo fills in the sub-stages of the causal sequence described in 1* – 9. To wit:

- (a) NR (Bush finds himself in his actual deliberative situation—a situation in which his salient choice concerns whether or not to get down at t—shortly before t).
- (b) NR (It is the initial stages of Bush's deliberations that is actually deterministically produced by his finding himself in the deliberative situation he does).
- (c) NR (The initial stages of Bush's deliberations are exactly as they are).
- (d) NR (It is the subsequent stages of his deliberations that are actually deterministically produced by these initial stages).
- (e) NR (Bush's subsequent deliberative states are exactly as they are).
- (f) NR (It is his decision, immediately before t, to get down at t that is actually deterministically produced by these later deliberate states).
- (g) NR (Bush decided immediately before to get down at t).
- (h) NR (It is Bush's act of getting down at t that is actually deterministically produced by his decision, immediately before t, to get down at t).

Thus, by applying Rule B-co, we get

- (i) Bush isn't morally responsible for the fact that he gets down at t. (Ibid., 416)

Now, in order for the compatibilist to deny that (i) is true—that is, that (9) is true—she will have to deny the move from (g) and (h) to (i). But how can she do that? In order to do such a thing, she'll have to grant that not even Bush is morally responsible for the fact that he decides, before t, to get down at t, and that his so deciding is what actually causes him to get down at t. Thus, the compatibilist will have to grant that Bush is responsible for getting down when he does, even though he's not responsible for the fact that he decided to get down, and that it's this decision that causes him to get down.

Of course, if the compatibilist doesn't like this consequence, then, as Shabo points out, the compatibilist can try and reject the move from (e) and (f) to (g). But, if she tries to do that, she'll run into similar problems as with her attempt to reject the move from (g) and (h) to (i). For, if she is to reject the move from (e) and (f) to (g), then she'll have to grant that Bush is morally responsible for the fact that he decides before *t* to get down at *t*, while denying both that Bush is morally responsible for the fact that his deliberative states are exactly as they are, and that these deliberative states actually produce his decision. But, to quote Shabo, "how can [Bush] be morally responsible for making the decision he does if he isn't morally responsible for that decision's being the one that is actually deterministically produced?" (Ibid., my insertion). And similar problems occur when the compatibilist tries to hold the line at (e), and so on. Thus, Shabo concludes that McKenna has failed to show that *causal* versions of the Direct Argument (versions of the argument that rely on Rule B-co, or a principle like it) are dialectically improper.

Even so, Shabo thinks that since Rule B-co includes appeals to deterministic causal sequences, Rule B-co is too tendentious to win the day for the incompatibilist defender of the Direct Argument. This is because it's not obvious, or uncontroversial, that deterministic causal sequences are even *relevant* to the issue of moral responsibility. Thus, Shabo concludes that the Direct Argument, by itself, cannot win the day for the incompatibilist.

But, I think that Shabo's reasoning is mistaken. For one thing, as in the last section, I think that concluding that Rule B is dialectically improper because cases like Bush Deliberates can't establish the principle is the wrong way to think about how philosophical principles are established (or at least, about the structure of the typical Rule B proponent's commitment to, or endorsement of, Rule B). As I've argued, and following Schnall and Widerker (2012; forthcoming), Rule B is *prima facie* established by uncontroversial cases. This is why van

Inwagen, in his *Essay* uses cases like Snakebite and Plato: they are meant to illicit in us intuitions about a plausible principle. Now, when we take this very plausible principle—viz., Rule B—and apply it to cases where an agent’s action is involved, we get the consequence that the agent isn’t directly morally responsible for her action. This is surely a tough pill for the compatibilist to swallow, but that’s why it’s important (if the Direct Argument is to be undermined) for the compatibilist to provide a counterexample to Rule B. She can’t, as McKenna does, presuppose that a causal or logical sequence’s passing through an agent makes a difference. She can’t presuppose anything like this because *that’s exactly what’s at issue!*

So, I conclude that Rule B is not dialectically improper; so, no need to think of the relationship between (L&P) and ‘p’ in the Direct Argument as *merely* causal. And here’s another reason I conclude as I do. Let’s take (a) – (i), from Shabo’s filling in of the causal sequence of 1* – 9, and restate them in terms of a *logical* sequence. I think if we do this, we’ll find that the difficulties incumbent upon the compatibilist to reject certain stages of (a) – (i), given a causal story, will remain given a logical story. Consider:

- (a*) NR (Bush finds himself in his actual deliberative situation—a situation in which his salient choice concerns whether or not to get down at t—shortly before t).
- (b*) NR (It is the initial stages of Bush’s deliberations that materially imply his finding himself in the deliberative situation he does).
- (c*) NR (The initial stages of Bush’s deliberations are exactly as they are).
- (d*) NR (It is the subsequent stages of his deliberations that are materially implied by these initial stages).
- (e*) NR (Bush’s subsequent deliberative states are exactly as they are).
- (f*) NR (It is his decision, immediately before t, to get down at t that is materially implied by these later deliberate states).
- (g*) NR (Bush decided immediately before to get down at t).

(h*) NR (It is Bush's act of getting down at t that is materially implied by his decision, immediately before t, to get down at t).

Thus, by applying Rule B, we get

(i*) Bush isn't morally responsible for the fact that he gets down at t.

Now, as with (a) – (i), if the compatibilist wishes to reject (i*), then she'll have to grant that Bush is neither morally responsible for the fact that he decided immediately before t to get down at t, nor the fact that his getting down at t is implied by his so deciding, all while claiming that Bush is morally responsible for the fact that he gets down at t. But, how can we plausibly conclude that Bush is responsible for getting down when he does, if we're willing to grant that he's neither responsible for the fact that he decides to get down, nor for the fact that his decision implies that *he does* get down? This seems to me to be the same sort of untoward consequence that the compatibilist must accept in the case where the compatibilist denies the move from (g) and (h) to (i). Moreover, the same difficulties remain for the compatibilist if she wishes to hold the line at (g*) and so on.

Of course, what I say about the implausibility of rejecting the move from (g*) and (h*) to (i*) might just rely on the fact that I think Rule B is valid. But this seems to me to be no problem at all, since I've defended its validity from what I take to be the most promising counterexamples on offer. And, moreover, since Rule B is *prima facie* established by non-controversial cases, it remains incumbent upon the compatibilist to give reasons why Rule B doesn't work in controversial cases. I think the only way to do this is to provide a counterexample to Rule B. But, none of the counterexamples I've examined so far do the trick.⁵⁵ Thus, I conclude that Rule B is valid. And if so, then rejecting (i*) while granting (g*) and (h*) (and any other similar

⁵⁵ In chapter six I'll do more than this. I'll argue that counterexamples to Rule B are *impossible*.

move) is implausible. Which conclusion shows us that Bush Deliberates might be able to establish a principle like Rule B, after all.

V. Conclusions

In this chapter I have evaluated, and rejected, several arguments that purport to show that the Direct Argument is dialectically inappropriate (in, at least, two ways), or that its role in the relevant dialectic is one that is merely supplemental. As we saw, Fischer rejects the Direct Argument because he thinks that Rule B is invalid, and any of the various revisions to the Rule are such that they lead to a dialectical stalemate. But I argued that Fischer is right only if Rule B is invalid. And since Rule B *isn't* invalid (anyway, we haven't seen any reason, yet, to think that it is), then Fischer is wrong. Whatever may be the case with respect to the various and sundry revisions to Rule B, these revisions are unneeded. Rule B, itself, is valid.

Moreover, we saw that Michael McKenna questions the dialectical appropriateness of the Direct Argument because Rule B cannot be established by anything other than non-controversial cases. According to McKenna, in order to establish a principle like Rule B, whatever cases are used to do any such thing must be such that they 'pass through' a normally functioning agent. But, as we saw with Schnall and Widerker, this is not the right way to think about how to establish a principle like Rule B. Often in philosophy, what we do is take a very plausible principle, apply it to non-controversial cases, then show that given the principle's *prima facie* establishment, if we apply the principle to controversial cases, we get such and such an outcome.⁵⁶ Thus, there is no reason—apart from an antecedent commitment to compatibilism—

⁵⁶ Take 'ought implies can', for instance, and reconsider the case from footnote 24, where it's alleged that I morally ought to end world hunger. It seems fairly obvious to me that I am under no such obligation (perhaps I'm under the obligation to *help* end world hunger; but I am not obligated to end world hunger). What reason could there be for the seeming obviousness of this claim? One thing a proponent of 'ought implies can' might say is that the reason I am under no such morally obligation is, at least in part, because I *can't* do any such thing. And since

to think that non-controversial cases *don't* establish Rule B. All that's needed in the dialectical context is that such a principle be *prima facie* established. What the compatibilist must do, then, is give reasons for rejecting such a principle that don't presuppose the falsity of the principle. I argued that, in order to do this, the compatibilist will have to provide a counterexample to Rule B. But, since no such counterexample has been offered, we haven't any reason for giving up Rule B.

Lastly, we saw that Seth Shabo is inclined to agree with McKenna about the dialectical appropriateness of the Direct Argument on a particular *reading* of the argument; viz., on a *logical* reading of the argument. But, he thinks, if we read the relationship between the later stages and earlier stages of a deterministic world in a particularly *causal* way, then it's less obvious that the Direct Argument is dialectically inappropriate. But I argued that all of this is just a mistake. For one thing, there's no good reason to reject Rule B in favor of a strictly causal reading like Rule B-co, and for two reasons: 1) none of the alleged counterexamples to Rule B that are on offer is, in fact, successful, and 2) Rule B is *prima facie* established through non-controversial cases. For another, when we examine Shabo's reasons for thinking that Rule B-co isn't dialectically improper, these same reasons seem to apply to Rule B just as well. Such conclusion arose from my suggestion that the reasons for thinking that the compatibilist's denial of (i) while granting (g) and (h) is implausible can be extended to her denial of (i*) while granting (g*) and (h*). And *this* conclusion shows that, in the end, Bush Deliberates *does* establish Rule B. This conclusion is yet another reason to deny both McKenna's conclusion that

it's uncontroversial (I assume) that I am under no moral obligation to end world hunger, if that obvious fact is partly explained in terms of my inability to do so, then 'ought implies can' seems to be *prima facie* valid. Now, if we consider a typical Frankfurt-style case, if it's really true that the agent in question could not have done otherwise than she did, 'ought implies can' implies that the agent is not morally responsible for her action. This, it seems to me, is another well-known example of a very plausible principle that is applied to non-controversial cases in order to infer its *prima facie* establishment. And taken as *prima facie* established, the principle gets applied to controversial cases—like Frankfurt cases, e.g.—and the results are fodder for philosophical discussion.

the Direct Argument ought to be done away with, as well as Shabo's initial reason to reject *logical* versions of the Direct Argument. Thus, I conclude that Rule B is valid, and the Direct Argument remains as strong as ever.

Now, given that Rule B is safe from questions of dialectical impropriety, we might be tempted to conclude that the Direct Argument is, itself, safe. But such a conclusion is too quick. For, while it is true both that Rule B is safe from counterexample, and that it is dialectically appropriate, there remains a need to defend Rule A, the rule that says no one is, or could be, even partly directly morally responsible for a necessary truth. While this Rule might seem to be a trivial truth, there is one recent attempt to reject the principle. In the next chapter, I turn my attention to a defense of Rule A from said objection.

CHAPTER FOUR

An Objection to Rule A

In chapters two and three, I set out to defend the most controversial of the Direct Argument's inference rules, Rule B, from various objections. In particular, I argued that the alleged counterexamples under consideration failed; they aren't counterexamples at all. And, moreover, I argued that Rule B is *not*, contra Fischer, McKenna, and Shabo, dialectically improper. So, summing up the work I've done so far, I concluded the last chapter with the suggestion that Rule B is safe (so far) from concerns of counterexample or impropriety.

Now, one might be tempted to think that, since Rule B hasn't yet been undermined, that the Direct Argument has won the day for incompatibilism. One might be tempted to think this, I'd imagine, because the premises of the Direct Argument are extremely plausible, and the only inference rule left to question is Rule A, which says that nobody can be even partly directly morally responsible for a necessary truth. And since Rule A seems to be an uncontroversial truth, and given the extremely plausible premises of the Direct Argument and Rule B's apparent validity, it looks as if the Direct Argument is both valid and sound. Thus, incompatibilism about causal determinism and moral responsibility is true (or most likely true).

But such a conclusion is too quick. For one thing, and as we'll see in the next chapter, there is an objection that calls into question whether or not all deterministic worlds have a remote past. The upshot of an objection like this, it should be clear, is that it casts doubt on (at least) premises 1, 2, 3, and 4 of the Direct Argument (for, all of these premises include a reference to the remote past). Discussion of this objection, however, will have to wait until the next chapter. What I wish to discuss in the present chapter is the *other* reason for thinking that it's too quick to

conclude, just on the basis of the preceding chapters, that the Direct Argument has won the day for incompatibilism. What I wish to discuss in this chapter, then, is a recent objection by Stephen Kearns (2011) to Rule A.

According to Kearns, Rule A is false because there are counterexamples that show that people are (or could be) partly directly morally responsible for a necessary truth. So, in what follows, I wish to examine three of Kearns's attempts to provide a counterexample to Rule A. I'll argue that these examples fail; and since these three alleged counterexamples exhaust the *types* of counterexamples Kearns wishes to give, if I'm right that these counterexamples fail, then it will follow that Kearns has not successfully shown that Rule A is false. Moreover, if this chapter's arguments are successful, then, given the arguments of the previous two chapters, it will follow that both of the Direct Argument's inferences rules are safe from (what I take to be) their most daunting and recent critiques.

II. Kearns's First Alleged Counterexample to Rule A: Murder!

To begin to see how Kearns's objection to Rule A will go, consider the following scenario:

Murder! Steven murders someone, and it is uncontroversial that he's responsible for his murdering since he does so knowingly and intentionally and could have done otherwise, etc. He is, therefore, responsible for the fact that he *actually* murders someone. However, the fact that he actually murders someone is necessarily true. It is true in every possible world that, in the actual world, Stephen murders someone. (Kearns, 2011, p. 309)⁵⁷

Kearns thinks that, given the set-up of Murder!, it's necessarily true that Stephen actually murders someone. Why think a thing like that? Well, Kearns thinks that *the actual world* is a rigid designator for a particular possible world (Ibid., p. 311). So, call the actual world α . If *the*

⁵⁷ For sake of expediency, I've made minor changes to the wording of Kearns's Murder! case. Nothing crucial rests on these changes.

actual world rigidly designates α , and Stephen murders in α , then, necessarily, Stephen murders in the actual world. What Kearns thinks follows from this is that, necessarily, Stephen actually murders; and, of course, Stephen is morally responsible for actually murdering. And even though it's a necessary truth that Stephen actually murders, he is at least partly morally responsible for this fact because, as Kearns puts it, "in some sense, Stephen *makes it the case* that he actually murders someone" (Ibid., 309, my emphasis). The actual world is the way it is, in part, because of what Stephen does; and since this is a necessary truth, Stephen is morally responsible for a necessary truth. So, Rule A is false.

But I don't think that Murder! succeeds in showing Rule A to be false. To see why not, I think it prudent to remove some of the (what I think is) obfuscatory language that Kearns uses. For example, Kearns asks us to assume that Stephen murders in the actual world, where the locution 'the actual world' is a rigid designator. So, as above, I will continue to call 'the actual world' α , instead. Here is why I think calling *the actual world* ' α ', will help get clear some, at least initially, confusing stuff. Suppose, along with Kearns that, in α , Stephen murders. I think Kearns is right to conclude that if Stephen murders in α , then it's necessarily true *that Stephen murders in α* . But I wish to point out that Kearns thinks that it follows from its being necessarily true *that Stephen murders in α* that he *actually* murders. This is confusing. For, does Kearns mean to argue that it's necessarily true that *in fact* Stephen murders? Or does he mean simply to argue that it's necessarily true that in *the actual world* (i.e. α) Stephen murders. I agree with the latter claim, but think that the former claim is obviously false. If it's possible for Stephen to refrain from murdering (and Kearns grants that it is (Ibid., p. 310)), then it's possible that Stephen refrains from murdering *in fact*. So, Kearns must mean to argue for the claim that it's necessarily true that Stephen murders in α .

Kearns, then, thinks that ‘actually’ denotes the *location* of Stephen’s murder. But, if that’s right, then I think that Kearns is mistaken. For, I think that ‘actually’ acts as an *intensifier* to ‘murders’ (i.e. ‘actually’ acts as a word that adds emotional context to ‘murder’, but nothing of propositional significance). What’s more, I think this reading of ‘actually’ is utterly plausible; in fact, I think it’s the correct reading. Moreover, it seems to me that that using ‘actually’ to denote the location of Stephen’s murder is a bit of a rhetorical trick to pull the reader’s intuition toward the conclusion that, if Stephen is morally responsible for actually murdering, and necessarily *that Stephen actually murders* is true, then Stephen is morally responsible for a necessary truth. To see better why it is that I say using ‘actually’ to denote a locator of Stephen’s murder is a rhetorical trick, allow me to restate the proposition *that Stephen actually murders* (a proposition that is *ex hypothesi* a necessary truth) using my preferred name for *the actual world*, viz., α .

S_α : That Stephen murders in α

S_α is necessarily true. S_α is necessarily true because, if Stephen murders in α , then it’s true in all possible worlds *that Stephen murders in α* . That is, it is true in all possible worlds *that Stephen murders in α* . So, S_α is necessarily true.

Is Stephen morally responsible for S_α ? I doubt it. To see why, suppose another possible world, β , had obtained instead of α . And suppose that in β , Stephen refrains from murdering. Now, it’s true in β *that Stephen murders in α* (since that’s necessarily true); but, is Stephen, in β , morally responsible for the fact *that Stephen murders in α* ? α didn’t obtain; and the world that obtained, β , is a world in which Stephen *doesn’t* in fact murder. Given such a case, I think it’s intuitive to conclude that Stephen is *not* morally responsible for S_α since Stephen’s murdering doesn’t obtain.

Now, Kearns might respond that there it is still true *in some sense* that Stephen *makes it the case* that S_{α} . And, he might continue, if this is true, then there's at least some sense in which Stephen is morally responsible for the fact *that Stephen murders in α* . Thus, Kearns might conclude that Stephen is at least partly morally responsible for the fact *that Stephen murders in α* .

But, in reply, I wish to cast doubt on the claim that Stephen makes it the case *that Stephen murders in α* . For, I think that in order for it to be true that Stephen makes it the case *that Stephen murders in α* , the fact *that Stephen murders in α* has to depend, in the relevant way, on Stephen and what he does. So, let's consider the proposition *that Stephen murders in α* . By hypothesis, this proposition is necessarily true. Now, it might help us figure out whether or not Stephen makes it the case *that Stephen murders in α* by thinking about what this proposition is about. I say it's about Stephen and whether or not he murders in α . Even so, I claim that Stephen is not responsible for the fact that the proposition *that Stephen murders in α* is about him and whether or not he murders in α . I say Stephen is not responsible for this fact because I think that the truth of *that Stephen murders in α* does not depend on Stephen.

Following Merricks (2007), I think that some, but not all, truths require truthmakers. I claim that *that Stephen murders in α* is one of those truths that does not require a truthmaker. Or, even if it does, its truthmaker is *not* Stephen. Why not? Because *that Stephen murders in α* 's truth is independent of whether or not Stephen is, in fact, actual (i.e. whether or not Stephen, in fact, exists), or whether or not Stephen, in fact, murders. And I think that in order for it to be the case that Stephen makes it the case *that Stephen murders in α* , it has to be the case that the truth of the proposition depends on Stephen (in the sense of 'depends on' that truth depends on

the world⁵⁸); that is, it has to be the case that Stephen *makes it true that Stephen murders in α* . But, even though the proposition is about Stephen and what he does, he does not make it true *that Stephen murders in α* ; for, this proposition is true regardless of what Stephen does, or whether or not he exists (and how can a non-existent thing make something the case?). So, I conclude that Stephen doesn't make it the case *that Stephen murders in α* . And since his making it the case *that Stephen murders in α* was supposed to show that he is at least partly directly morally responsible for a necessary truth, I conclude that nothing of the sort has been shown. Thus, I conclude Rule A seems safe from Kearns's argument from Murder!.

But suppose someone were to object as follows:

Your argument seems to be assuming something like the following conditional:

If a proposition, P, will be true no matter what an agent, S, does, then S does not make it the case that P is true.

But this conditional is not obviously true. For, don't Frankfurt-style cases cast some doubt on this particular assumption? Given Black's presence, plans, and powers, Smith is going to be fatally shot no matter what Jones does. But in fact Jones, himself, freely chooses to shoot Smith, subsequently shoots Smith 'on his own' without Black's intervention, etc. By the above assumed conditional, Jones doesn't make it the case that *that Smith is fatally shot* is true (since, again, that was going to be true no matter what). This is not obviously correct, though, given that the contribution that Jones's free choice (etc.) actually made to Smith's being fatally shot.⁵⁹

In reply, I argue that Frankfurt cases, such as the one under consideration, do *not* show that the above conditional is false, and here is why. The truth of *that Smith is fatally shot* does not depend on, in the sense of 'depends on' that truth depends on the world, what Jones does. Rather, the truth of *that Smith is fatally shot* depends on Smith and whether or not he is fatally shot. But for Jones to be morally responsible for the fact *that Smith is fatally shot*, it has to be

⁵⁸ For more about the notion of 'depends on' at play, here, see, again, chapter two, footnote 12. In chapter six I discuss this notion at length.

⁵⁹ Thanks to E.J. Coffman for the objection.

the case that this fact depends on (in the sense of ‘depends on’ that truth depends on the world) what he does. Since it doesn’t, Jones is not morally responsible for the fact *that Smith is fatally shot*.⁶⁰

Before I finish with Murder!, however, I want to offer another reason to doubt that Stephen is morally responsible for S_{α} . I think that Kearns, in his attempt to argue that Stephen makes it the case *that Stephen murders in α* assumes something like principle ER, from chapter two. Recall:

ER: If S is directly morally responsible for p, and $p \supset q$, then S is directly morally responsible for q.

Here is why I think that Kearns’s argument that Stephen makes it the case *that Stephen murders in α* relies on something like ER. Consider another way Kearns might respond to my above argument that Stephen *does not* make it the case *that Stephen murders in α* :

By hypothesis, Stephen murders someone; that is, Stephen *in fact* murders someone. So, Stephen makes it the case that he murders someone. But this fact implies *that Stephen murders someone in α* . So, it follows that Stephen makes it the case *that Stephen murders someone in α* . And since all should agree that Stephen is directly morally responsible for the fact that he, in fact, murders someone, and this fact implies *that Stephen murders someone in α* , we should all agree that he is directly morally responsible for the fact *that Stephen murders someone in α* .

But such a response relies on two invalid inferences. One is ER, from above, and the other is something like it that I’ll call:

MC: If S makes it the case that p, and $p \supset q$, then S makes it the case that q.

But, as I argued in chapter two, while ER is superficially plausible, it is false by counterexample. And so is MC, for the exact same reasons. For, again, suppose that Butch Jones is morally responsible for his being (or that he makes it the case that he is) a great University of Tennessee

⁶⁰ If Jones is morally responsible for anything it’s the truth of *that Jones fatally shoots Smith* since this fact *does* depend on Jones and what he does.

football coach. Butch Jones's being a great University of Tennessee football coach implies that there is such a place as the University of Tennessee. But surely Butch Jones is not morally responsible for (and neither does he make it the case that) there being such a place as the University of Tennessee, even if he is (we may assume) morally responsible for being (or made it the case that he is) a great football coach at the University.

So, if Kearns is relying on ER (or MC), then he is erroneously led to believe that Stephen is responsible for (or makes it the case that he) actually murdering someone because he thinks that Stephen's actually murdering someone is a consequence of his murdering someone (something I am happy to agree that Stephen makes the case). Thus, he might conclude via something like ER, that Stephen is at least partly directly morally responsible for actually murdering. But, since ER (and MC) is false, the envisaged reasoning is unsound. I conclude, then, that Kearns's argument from Murder! fails to show that Rule A is false.

III. Kearns's Second Alleged Counterexample to Rule A: Hey Jude

In this section, I wish to turn our attention to a second counterexample that Kearns gives, a counterexample that is of a different *type* than the first. Consider:

Hey Jude [HJ]: Paul McCartney composes the melody of Hey Jude and does so knowingly and intentionally, etc. He is morally responsible for various facts concerning Hey Jude like the fact that Hey Jude starts the way it does and for the fact that Hey Jude has a beautiful melody since he *created* Hey Jude. However, these facts that McCartney is responsible for are necessary truths. It is necessarily true that Hey Jude starts in the distinct way it does and that it has a beautiful melody.

I say that HJ is a different *type* of counterexample to Rule A because it's designed to undermine Rule A in a way that doesn't rely on a person's being responsible for a world-indexed truth. Murder!, you'll recall, relied on Stephen's being morally responsible for a world-indexed truth, and since world-indexed truths are necessary truths, it was supposed to follow that Stephen is

morally responsible for a necessary truth. But, as Kearns rightly points out, an incompatibilist *could* (though, given my argument from the last section, I don't think that an incompatibilist *should*) concede that Murder! is a counterexample to Rule A, but that the Direct Argument needn't rely on Rule A. All the Direct Argument need rely on, the imagined incompatibilist might claim, is the following:

Rule A*: Necessarily, for any proposition, p, if it is necessary that p (and this necessary truth is not world-indexed), then no one is even in part directly morally responsible for the fact that p. (Ibid., p. 311)

The argument *for* Rule A*'s truth, then, can go as follows:

P1. An agent is in part directly morally responsible for a fact only if the agent makes this fact obtain.

P2. No agent can make a non-world-indexed necessary fact obtain.

C1. Therefore, no one is in part directly morally responsible for any non-world-indexed necessary fact (i.e. Rule A* is true). (Ibid., p. 312)

It is this argument to which HJ is supposed to provide a counterexample. So, since Murder! would not have provided a counterexample to P1 – C1, HJ is a different *type* of counterexample to Rule A because it rules out Rule A*, as well.⁶¹

Now, why think that HJ provides a counterexample to Rule A*? Kearns thinks HJ shows that A* is false because McCartney is morally responsible for some necessary truths. McCartney is responsible for, among other things, the melody and beauty of “Hey Jude”. And, since “Hey Jude” has these properties necessarily—that is, “Hey Jude” couldn't have been any other way—McCartney is morally responsible for some necessary truths since he's the one who created “Hey Jude”. So, Rule A* (and Rule A) is false.

⁶¹ To be clear, HJ (and the alleged counterexample that I consider in the next section) serves to perform two functions. The first is to show that P2 of the argument for Rule A* is unsound. The second is to show that a person can be at least partly directly morally responsible for a necessary truth—and, insofar as this is the case, Rule A* (and Rule A) is false.

But I think that HJ fails to show that Rule A* (and Rule A) is false. To see why, note, first, that Kearns thinks that “it is a necessary truth that Hey Jude starts a certain way, and is beautiful”, and that “if Hey Jude had had a different melody, it wouldn’t have been Hey Jude” (Ibid., p. 313). This is just what HJ says: “Hey Jude” *just is* a certain way, and McCartney is morally responsible for “Hey Jude” because he created it.

But this seems wrong to me. Why should we suppose that, if it’s true that “Hey Jude” is necessarily a certain way, that McCartney created it? It seems to me that supposing “Hey Jude” *just is* a certain way is to suppose that “Hey Jude” has some essential properties. But, if that’s right, then “Hey Jude” just seems to be an instantiation of something like *Hey Judeness* or *Hey Judaity*. I think it’s dubious whether or not there is such a thing as Hey Judaity, but I think that Kearns’s contention that “Hey Jude” *just is* a certain way (that is, “Hey Jude” is necessarily the way it is, whether or not Paul McCartney had ever existed, or if he *had* existed but never actualized it) commits him to such an ontology.

So, suppose that this is what Kearns thinks, and that his thinking this way about “Hey Jude” is right. Again, I think this means that “Hey Jude” is an *instantiation* of *Hey Judaity*. And, if that’s right, then McCartney *did not* create “Hey Jude”; rather, he *actualized* it, i.e., he created an instance of the property *Hey Judeness*.⁶² So, what McCartney is morally responsible for is not his *creating* “Hey Jude”—*Hey Judaity* seems to exist necessarily as an abstract object—he’s morally responsible for *actualizing* it. And since he needn’t have actualized it, then he’s not morally responsible for a necessary truth. He’s not morally responsible for the fact that,

⁶² Recall Haji’s ‘Hal’s Creation’ cases from chapter two. In each case Hal *actualizes* a possible world. Why is it better to say that Hal *actualizes* a possible world rather than *create* a possible world? I assume this is because Haji accepts an ontology of possible worlds such that possible worlds exist necessarily as abstract objects. Thus, if Hal want’s one of these possible worlds to be actual, he can’t *create* it, strictly speaking; it already exists. What he *can* do is make one of them actual; that is, he can *actualize* one of them. The same, I think, must be truth about “Hey Jude” if we’re given that Hey Judeness or Hey Judaity exists necessarily. Paul McCartney can’t, strictly speaking, *create* “Hey Jude”; what he can do is *actualize* it. For more on this, see Alvin Plantinga’s (1974).

necessarily, “Hey Jude” starts a certain way or has a certain melody. He’s morally responsible for the fact that “Hey Jude” is actualized; and that’s not a necessary truth. So, I conclude that HJ fails to show that Rule A* (and Rule A) is false.

To see another reason to think that HJ fails to show that Rule A* (and Rule A) is false, consider the following argument, one that is generated straight from HJ.

1. Paul McCartney is at least partly directly morally responsible for the fact that “Hey Jude” starts the way it does.
2. Necessarily, “Hey Jude” starts the way it does.
3. Therefore, Paul McCartney is at least partly directly morally responsible for a necessary truth (i.e. Rule A* (and Rule A) is false)

Given 1 – 3, I think it’s pretty clear that HJ fails as a counterexample to Rule A* (and to Rule A); for, I think that premise 2 is clearly false. To see that this is so, imagine that you and I have just come from seeing Paul McCartney play a live music concert. And suppose, further, that I am no fan of McCartney’s, I don’t know anything about his music; but you are a fan, and you do know something about his music. Now, suppose that I ask you something like the following:

Hey, what was that third song McCartney played? The one that said something about taking a sad song and making it better. I liked that one; what was it called?

I think, obviously, your answer to my question should be that the song in question is called “Hey Jude”. And I think this is true *even if* Paul McCartney (unbeknownst to me) adlibbed the beginning, or played the verse a step up (that is, in a higher key) from the way he normally plays it.

For, suppose that Paul McCartney did these things, that he adlibbed a bit. Would it be plausible for you to respond to my question in something like the following way?

Well, it’s difficult to say what song McCartney played third. I know to which song you refer, and it *sounds* an awful lot like this other song of his, “Hey Jude”, but it wasn’t *that*

song; for, he played the beginning a different way (or he played the verse a step up from normal). So, I don't know what to tell you, but it wasn't "Hey Jude".

If you answered me in this way, and I went and listened to the recorded version of "Hey Jude", I think I'd have every right to think you'd not been straightforward with me, or that you'd confused my question for some other. For, I think that "Hey Jude" does *not* necessarily start a certain way. There are many things about "Hey Jude" that can change, it seems to me, without it being the case that "Hey Jude" isn't being played. Now, this of course leads to sorities problems with respect to the identity of things like songs, etc.; but, that's another issue entirely. The point is that, unless Kearns is willing to accept a dubious metaphysical claim like *that there is such a thing as Hey Judaity*, then I think he should concede that "Hey Jude" isn't *necessarily* a certain way, or, anyway, that "Hey Jude" doesn't necessarily start in a particular way. And if not, then, given that the intro to "Hey Jude" can change without the song's numerical identity being changed, premise 2 is false. So, I conclude that HJ fails as a counterexample to Rule A* (and Rule A).

But, perhaps Kearns *isn't* committed to the existence of *Hey Judaity*, and that Paul McCartney really did *create* "Hey Jude". If so, then I think Kearns is guilty of basing his HJ argument on a premise like ER (or MC). Kearns erroneously believes that Paul McCartney is morally responsible for the fact that "Hey Jude" has the melody that it does because he reasons that "Hey Jude" has the melody that it does is a consequence of Paul McCartney's having composed it (something that I am happy to agree that McCartney makes the case). Thus, by ER (or MC), Paul McCartney is morally responsible for the fact that "Hey Jude" has the melody that it does. But we saw that ER (and MC) is false. So, I conclude that Kearns's argument from HJ fails to show that Rule A* (and Rule A) is false.

IV. Kearns's Third Alleged Counterexample to Rule A: Torturing Babies

In this section, I'd like to consider another type of alleged counterexample to Rule A* (and Rule A), namely:

Torturing Babies [TB]: Karen decides to make eating cake immoral by making it the case that when someone eats cake, thousands of people are killed. Moreover, Karen decides to make torturing babies for fun immoral. She does so by making it the case that thousands of people are killed when someone tortures a baby for fun. (Ibid., p. 314 – 315)

TB is a different type of counterexample to Rule A* (and Rule A) because it purports to show that a person can be partly directly morally responsible for a necessary truth, even when that necessary truth's being true is *overdetermined*. To be clear, Kearns thinks TB shows that Rule A* (and Rule A) is false because, though torturing babies is necessarily immoral—i.e. it's necessarily true that torturing babies is wrong—Karen is at least *partly* directly morally responsible for the fact that torturing babies is wrong because she, at least in part, *makes it the case that* torturing babies is wrong. Perhaps she overdetermines its being wrong to torture babies, but torturing babies is still at least partly wrong because of what Karen does (viz., insuring that thousands of people will die if someone tortures a baby). So, Rule A* (and Rule A) is false because Karen is partly morally responsible for a necessary truth.

But as with the foregoing alleged counterexamples, I think that TB fails to show that Rule A* (and Rule A) is false, as well. And I think the fact that Karen's actions overdetermine the wrongness of torturing babies helps make it clear why TB fails as a counterexample. Now, before I say why I think that the fact that Karen's actions overdetermine the wrongness of torturing babies helps make it clear why TB fails as a counterexample, it needs to be noted that Kearns thinks such overdetermination is *not* a problem for TB. He thinks overdetermination isn't a problem because he thinks Fischer and Ravizza's Erosion* example shows that

overdetermination doesn't rule out moral responsibility. Recall, from chapters two and three, the following:

Erosion: Imagine that Betty [a soldier charged with destroying an enemy fortress] 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 at T2, starting a natural avalanche that would have crushed the enemy camp at T3.

Erosion*: [The details are the same as Erosion except that, in this case,] the conditions of the glacier *do* actually cause the ice and rocks to break free, triggering an avalanche that arrives at the fortress precisely at the same time as the independent avalanche triggered freely by Betty. Each avalanche is sufficient for the destruction of the enemy fortress. Kearns thinks that, in Erosion*, Betty is clearly morally responsible for the destruction of the enemy fortress.

The conclusion of Erosion*, with which Kearns agrees, is that Betty is morally responsible for the fact that the enemy base is destroyed, and this is true even though that fact is overdetermined. Thus, Kearns thinks overdetermination doesn't rule out moral responsibility.

But, as I argued in the previous chapter, this conclusion isn't at all clear to me. The destruction of the enemy fortress is overdetermined (that much is not up for debate); so, it's being true *that the enemy fortress is destroyed* doesn't depend on Betty; for, the fact *that the enemy base is destroyed* isn't about Betty or what she does. And I think in order for Betty to be morally responsible for some fact or other, it needs to be the case that that fact depends on her in the relevant way (i.e. that the fact in question is *about* Betty and what she does). Thus, it seems to me that Betty *isn't* morally responsible for the fact *that the enemy base is destroyed*. And the same will be true for *anyone* with respect to an overdetermined fact.

Moreover, I think that what Betty plausibly *is* responsible for is the fact *that Betty causes an avalanche that destroys the enemy base*.⁶³ And the same can be said, *mutatis mutandis*, for Karen and her overdetermining the immorality of torturing babies. She's not *at all* responsible for it's being immoral to torture babies; for it's being immoral to torture babies (which is *ex hypothesi* necessarily true) doesn't depend in the relevant way on Karen and what she does. That is, the fact *that it's immoral to torture babies* doesn't depend on Karen because it's not about Karen in any relevant way. Thus, it seems to me that she is not at all directly morally responsible for the fact *that torturing babies is immoral*. What she plausibly *is* responsible for is her *attempt*, or, perhaps, her *intention*, to make torturing babies immoral. But this latter bit is not a necessary truth, while the former is. And since Karen is not at all morally responsible for the former—but is the latter—she is not at all morally responsible for a necessary truth. I conclude that TB fails to show that Rule A* (and Rule A) is false.

But I don't want to move too hastily here, so here I'll provide the same argument I ran against Kearns's argument from Murder!, and his argument from HJ. In his argument from TB, Kearns seems to me to erroneously believe that Karen makes it the case that it's wrong to torture babies because Kearns reasons that its being wrong to torture babies is a consequence of Karen's insuring that thousands of people will die if a baby is tortured. Thus, by MC, Karen makes it the case that it's wrong to torture babies; and if she makes the case that it's wrong to torture babies, then it's plausible that she's at least partly directly morally responsible for having done so. But as we've seen, MC is false. So, I conclude that Kearns's argument from TB fails to show that Rule A* (and Rule A) is false.

⁶³ I'll repeat, here, what I said in footnote 48 of chapter three: I think that it's even *more* likely that what Betty is directly morally responsible for is her *intending to destroy the enemy base*. In any case, she's not directly responsible for the fact *that the enemy base is destroyed*.

V. Conclusions

In this chapter I have considered, and rejected, three alleged counterexamples to Rule A. More precisely, I have considered one alleged counterexample to Rule A *in particular*, namely, Murder!, and two alleged counterexamples to Rule A*, a revised version of Rule A. I argued that all of these counterexamples fail, in the first place, because each case fails to show that the truth of the fact for which the agent is allegedly morally responsible *depends on* the agent in the relevant way. After all, necessary truths are true regardless of whether or not the agent in question exists. And if that's true, then it cannot be the case that the truth of the fact in question depends on the agent. Thus, I conclude that none of the agents in question are even in part directly morally responsible for the truth of the facts under consideration.

Moreover, I argued that it seems as if Kearns's argument, in each case, relies on one or both of the fallacious inference rules, ER and MC. I argued that in each alleged counterexample, Kearns attempts to show that someone is (or could be) directly morally responsible for a necessary truth because the relevant agent in each case *makes it the case* that some necessary truth or other is true by virtue of the fact that the agent makes it the case that some fact that *implies* the necessary fact is true. But, this reasoning doesn't follow. For it's perfectly possible, as my Butch Jones example shows, to make it the case that p, while p implies q, without, thereby, making it the case that q. And if I'm right that Kearns's alleged counterexamples rely on this reasoning, this is all the more reason to conclude that his counterexamples fail. Thus, I conclude that his counterexamples fail.

So far as I know, Kearns is the only philosopher to have (in print) questioned the validity of Rule A. I have argued that his attempts to question the Rule by counterexample fail. Thus, I conclude that Rule A is safe; Rule A is valid. Moreover, given the arguments of the foregoing

chapters, it seems that Rule B is safe, too; for, the most recent and prominent attempts to undermine the Rule fail. If my arguments have been so far successful, it follows that both Rules A and B are on fairly sturdy footing. This is good news for proponents of the Direct Argument since the Argument's inference rules are its most controversial parts.

Even so, this is not the end of the story for the compatibilist's fight against the Direct Argument. There is a recent objection to the Direct Argument that *grants* the truth of the Argument's inference rules. Moreover, this objection *grants* that the Direct Argument is successful in showing that in deterministic worlds that have a remote past, there is no moral responsibility. But *this* truth is consistent with its also being true that there are deterministic worlds that *lack* a remote past but are such that include moral responsibility. Thus, if this objection is right, while it concedes that the Direct Argument works, it shows that the Direct Argument's conclusion doesn't demonstrate that *incompatibilism* is true. This is because incompatibilism is supposed to be necessarily true, if true at all. Thus, the so-called No Past Objection, if successful, undermines the strength of the Direct Argument. So, since I think that the Direct Argument *is* a conclusive argument for the truth of incompatibilism, I take up the task of considering the No Past Objection in the next chapter. It is to this task I now turn.

CHAPTER FIVE

The No Past Objection and a Revised Direct Argument: The Direct Argument*

In the foregoing chapters, I dealt with alleged counterexamples to the inference rules upon which the Direct Argument rests. As we've seen, none of these alleged counterexamples is successful; so, at least for now, it seems to me safe to conclude that both Rules A and B are valid.⁶⁴ And since we've now concluded that both Rules A and B are valid, things look pretty good for the fate of the Direct Argument, and pretty bad for the fate of compatibilism. For, if the Direct Argument is sound, then it's supposed to be that incompatibilism about moral responsibility and causal determinism follows. Unfortunately for the proponent of the Direct Argument, however, there is a recent objection whose conclusion is that the Direct Argument's soundness is *not* sufficient for concluding that incompatibilism is true. It's exactly this problem I intend to deal with in the present chapter.

To begin, Joseph Campbell (2007, 2008, 2010) has recently presented a novel objection to the Consequence Argument—an argument for the incompatibility of *free will* and determinism from van Inwagen's *An Essay on Free Will*—namely, what Andrew Bailey (2012) calls the No Past Objection. Like the Direct Argument, the Consequence Argument begins with a premise that says that 'determinism' is a thesis about the fact that if you conjoin any complete statement about the facts at any one time in the past with the laws of nature, then this conjunction entails a unique future. The upshot of the No Past Objection, however, is that a *remote* past—the sort of past to which the Consequence Argument (and, for that matter, the Direct Argument) refers—isn't a necessary condition of the 'determinism' thesis. And since the premises of the

⁶⁴ In the next chapter I'll make an outright case for this claim with respect to Rule B. I'll argue that counterexamples to Rule B are impossible, and demonstrably so.

Consequence Argument aren't necessary truths, the Consequence Argument fails to support incompatibilism about free will and determinism since incompatibilism, in this sense, if true, is necessarily true. To be clear, the upshot of the No Past Objection is that if it is successful, at best, the Consequence Argument can establish only this: necessarily, no deterministic universe *with a remote past* contains free willed agents. But *that* thesis leave open the following thesis: possibly, *some* deterministic universe that *lacks* a remote past contains free willed agents. And if this last thesis is true, then compatibilism is true. So, even if the Consequence Argument is sound, it does not rule out compatibilism, and so is not a valid argument for incompatibilism about free will and determinism.

Now, I'm not concerned with the Consequence Argument; I'm concerned with the Direct Argument. But a critic could argue that the No Past Objection is *also* an objection to the Direct Argument. Moreover, it's an objection that *grants* the validity of Rules A and B and so it's an objection that avoids the arguments in the foregoing chapters, i.e., the arguments that purport to show that neither Rule A nor Rule B have been defeated by counterexample. With this in mind, a full defense of the Direct Argument would need to address the No Past Objection. In this chapter, then, I'd like to suggest how I think proponents of the Direct Argument should respond to the No Past Objection. Now, it seems to me that the No Past Objection *does* undermine the Direct Argument, at least as it's currently stated.⁶⁵ But even so, it doesn't undermine the Direct Argument*, the novel version of the Direct Argument I develop which avoids the No Past Objection and is consistent with what's been argued thus far.

⁶⁵ For reasons to think that the No Past Objection fails to undermine the *Consequence* Argument (or, rather, a suitably modified version of the Consequence Argument), other than the reasons I'll offer here for the Direct Argument (which can be made to support a defense of the Consequence Argument), see Alicia Finch (2013).

II. The No Past Objection

To see why Campbell thinks that the No Past Objection undermines the Consequence Argument, consider:

Consequence Argument [CA]: If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (van Inwagen, 1983, p. 16)

Campbell thinks the Consequence Argument fails because the No Past Objection is successful.

Here, then, is how Bailey presents the No Past Objection:

The premises of the Consequence Argument are not necessary truths because there needn't be a past. The Consequence Argument thus does not support incompatibilism, a thesis that is necessarily true if true. (Bailey, 2012, p. 352)

Why think the premises of the Consequence Argument are contingent truths (if they're true at all)? Because there needn't be a remote past. Why needn't there be a remote past? Consider the following scenario cooked up by Campbell:

AW: Consider, for instance, the possible world W. Suppose that W is a determined world such that some adult person exists at every instant. Thus, W has no remote past. At its first moment of existence lived Adam, an adult person with all the knowledge, powers, and abilities necessary for moral responsibility. Shortly after Adam comes Eve, and the rest is history. (Campbell, 2007, p. 109)

The upshot of AW is that it's false that Adam isn't free with respect to the remote past; for *there is no* remote past for Adam; Adam exists at W's *very first* instance. Moreover, other cases successfully show that the determinism thesis needn't include any notion of a remote past.

Consider:

OAW: Suppose that there is a deterministic world, W*, where time is circular. In that world exists oscillating Adam. Oscillating Adam has always existed and will always continue to exist. He is in the grips of an everlasting, eternal recurrence. Oscillating Adam spends his time growing 'older' and getting 'younger'. He begins each cycle with powers comparable with the average 25 years old and eventually develops powers comparable with the average 50 years old. Then he slowly regresses back to the state at

which he began, and the cycle starts all over again. (Campbell, 2010, pp. 72 – 73)

And:

IAW: There is a deterministic world, W^{**} , where time has no beginning and no end. In that world exists Infinite Adam, who exists at every time in W^{**} . He neither comes into existence nor goes out of existence. Suppose further that at every time he exists, he has powers comparable with the average 35-year-old. (Finch, 2013, p. 160)

Of course OAW, and IAW, have a similar upshot to the one that AW has: it's false that Adam⁶⁶ isn't free with respect to the remote past; for, there is no remote past; Adam exists at every time, t , in W^* and W^{**} .

So, since CA depends on a transfer principle about the lack of power over the past to the lack of power over the present, CA fails if the No Past Objection is successful. Since Campbell thinks the No Past Objection *is* successful—because, as AW, OAW, and IAW show, ‘determinism’ as a thesis doesn’t require a remote past; it merely requires a set of natural laws that, once instantiated, entail a unique future—he concludes that the Consequence Argument fails.

But we can extend this criticism to the Direct Argument. Consider:

Direct Argument: If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But we are not responsible for what went on before we were born, and neither are we responsible for what the laws of nature are. Therefore, we are not responsible for the consequences of these things (including our present acts).

Since the Direct Argument depends on a transfer principle about non-responsibility for the past to non-responsibility for the present, the Direct Argument fails to show that incompatibilism is true if the No Past Objection is successful. Since Campbell thinks the No Past Objection *is* successful—because, as AW, OAW, and IAW show, ‘determinism’ as a thesis doesn’t require a remote past—he should conclude that the Direct Argument fails.

⁶⁶ Unless otherwise indicated, the name ‘Adam’ will be used to denote any one of the ‘Adams’ from AW, OAW, or IAW (so, Adam, Oscillating Adam, or Infinite Adam).

This is an interesting criticism as applied to the Direct Argument. And, as I explained, it's a criticism that grants the validity of both Rules A and B and, so, it's a criticism that sidesteps my defenses of both Rules. But, even if the No Past Objection successfully undermines the Direct Argument, it doesn't successfully undermine a very similar argument to the Direct Argument, an argument I call the Direct Argument*. To see why not—and before I introduce the Direct Argument*—consider some preliminary points. The upshot of any one of AW, OAW, or IAW is that there are possible worlds without a remote past. It's possible, after all, for Adam to exist at every temporal moment of W's (or W*'s, or W**'s) existence up until his death (or whatever); so, it's possible that Adam lack a temporal remote past and, thus, it's possible for the Direct Argument to fail to apply to him even though he lives in a deterministic world (that is, a world that includes a set of natural laws that, once instantiated, entail a unique future).

I think that cases like AW, OAW, and IAW undermine the Direct Argument as it is currently stated. In the next section, I will introduce a revised version of the Direct Argument—what I call the Direct Argument*—that successfully avoids cases like AW, OAW, and IAW while holding fast to the original Argument's inference rules, as well as what I take to be the metaphysical spirit of the Direct Argument, viz., that if something besides us determines our actions, then we are not directly morally responsible for those actions.

III. The Direct Argument*

Here is the setup for the Direct Argument*. According to AW, OAW, and IAW, it's possible for a contingently existing thing to lack a remote past. In particular, Adam doesn't have a remote past and he's a contingently existing thing. Let's grant this. But even if we grant that it's possible for contingently existing things to lack a remote past, I think we should conclude

that, necessarily, all contingently existing things have a remote past*. Now, what is a remote past*? A remote past* is just the idea that for any thing, *x*, if some other thing, *y*, comes in any sense *before* *x*'s existence (or instantiation, or obtaining) and *x* has no influence as to whether *y*—that is, if *y* is prior to *x*'s existence (or instantiation, or obtaining) in any sense (e.g. temporally prior, causally prior, logically prior, etc.), and *y* is beyond *x*'s influence—then *y* is remotely past* to *x*. I think that, for any contingently existing thing, *x*, *x*'s existence is caused; thus, *x*'s existence has a cause. And if *y* is the cause of *x*'s existence, then, since *y* is at least causally prior to *x*, it follows that any contingently existing thing, *x*, has a remote past*. More to the point, here is my argument that *Adam, qua* contingently existing thing, has a remote past*. Consider:

D1. Whatever contingently exists has a cause. (Premise)

D2. Adam is a contingently existing thing. (Premise)

F1. So, Adam's existence has a cause. (D1 – D2)

D3. If a thing *y* comes, in some sense (e.g. 'causally', 'logically', 'temporally', etc.), *before* another thing *x*'s existence, and *x* has no influence as to whether *y*, then *y* is remotely past* to *x*. (Premise)

D4. Whatever causes Adam to exist comes, in some sense (e.g. 'causally', 'logically', 'temporally', etc.), *before* Adam's existence, and Adam has no influence as to whether something causes his existence. (Premise)

F2. So, whatever causes Adam to exist is remotely past* to Adam. (D3 – D4)

Thus, because Adam is a contingently existing thing at all worlds in which he exists, necessarily, Adam has a remote past*.

But this is true about more than just Adam. It's also true about the laws of nature (L). It's also true about L because, even though L exists necessarily (we may assume⁶⁷), L is only *contingently* true;⁶⁸ that is, there are possible worlds in which L fails to be true. So: L is contingently true. If so, then consider the following argument:

D1*. Whatever is contingently true has a cause. (Premise)

D2*. L is contingently true. (Premise)

F1*. So, L's being true has a cause. (D1* – D2*)

D3*. If a thing y comes, in some sense (e.g. 'causally', 'logically', 'temporally', etc.), *before* another thing x's being true, and x *has no influence as to whether* y, then y is remotely past* to x. (Premise)

D4*. Whatever causes L to be true comes, in some sense (e.g. 'causally', 'logically', 'temporally', etc.), *before* L's being true and L has no influence as to whether something causes it to be true that L. (Premise)

F2*. So, whatever causes L to be true is remotely past* to L. (D3 – D4*)

So, because L is contingently true at all worlds in which L is true, necessarily L has a remote past*. But the thesis, 'determinism', includes L; so, necessarily, 'determinism' includes a remote past*. So, necessarily, if determinism is true, then there's a remote past* and some laws of nature. What's more, any possible world where determinism is true is a world that has a remote past*. So, I conclude that 'determinism' is a thesis about the conjunction of the remote past* and the laws of nature, and the fact that that conjunction entails any future true fact whatever. Thus, I conclude that the Direct Argument *should be*, and, I think, *intends to be*, an argument that

⁶⁷ I think we should assume this because, following van Inwagen (1983, pp. 60-61), I think that the laws of nature are propositions; and I think that propositions exist necessarily. This is not to say, however, that these propositions *obtain* or *are instantiated* necessarily, as I suggest in the text.

⁶⁸ I take it that this is the majority view; however, it's not without its detractors. For example, Brian Ellis (2001), and Ellis with Caroline Lierse (1994) think that certain laws of nature are necessarily true. I ignore this discussion, here, but note that there is a debate about this in the literature.

purports to show that non-responsibility for the remote past* and the laws of nature implies non-responsibility for the present.

Before I formally present the Direct Argument*, I wish to consider a possible objection that might crop up regarding this notion of a ‘remote past*’. Specifically, the imagined objection targets premises D3 and D3* from the foregoing arguments. Here’s how I imagine such an objection might go:

D3 and D3* are non-sensical. ‘Past’ is a temporally loaded word. To be sure, there is such a thing as *priority*; there are, as you suggest, notions such as causal priority, logical priority, and the like. But when you add on the word ‘remote’ to ‘remote past*’ and claim that you’re simply using ‘past*’ as a substitute for ‘prior’, your case begins to crumble. For a ‘remote past’ is clearly temporal. Is this not the case for a ‘remote past*’? How can, for example, something that is logically prior to another thing be at all ‘remote’ with respect to it? And the same sort of question applies, *mutatis mutandis*, to the notion of causal priority. Thus, no sense can be made of a remote past*—no *non-temporal* sense, anyway. So, D3 and D3* are non-sensical.

I concede that I feel the force of an objection like this, but I think that, ultimately, it’s mistaken; and for two reasons. To begin to see the first reason, recall that ‘P’ in the formal version of the Direct Argument denotes a ‘remote past’—the very thing that the No Past Objection calls into question. Now, what is a ‘remote past’, exactly? As the above imagined objection has it, a remote past is clearly temporal, not simply because ‘past’ is clearly temporal, but because ‘remote’ as a qualifier is also temporally loaded. But I’m not convinced that the imagined objector has this right.

‘P’ in van Inwagen’s original version was labeled as P_0 to “denote a proposition that expresses the state of the world at T_0 ” (van Inwagen, 1983, p. 70), that is, before any human agent was alive. Thus, ‘the remote past’ is an idea that captures a proposition about the state of the world at any time before there were human agents. Now, what makes the proposition about the past as denoted by ‘ P_0 ’ remote? Is it the *time* of the proposition in question? Or is it the fact

that the proposition is *unable to be influenced* by any human agents? I think it must be this latter suggestion that makes the remote past ‘remote’; thus, the reason D3 and D3* include clauses about x having influence as to whether y.

If I’m right about this, then this is consistent with there being such a thing as a remote past*; that is, this is consistent with its being the case that a remote past* makes sense. For, whatever is logically, causally, etc., prior to Adam’s existence in AW, OAW, or IAW is such that it is not something that Adam can influence. And *this* is the point of ‘remoteness’ in the idea of a ‘remote past’. So, since whatever causes Adam to exist is not something that Adam can influence (for, Adam doesn’t exist yet), whatever causes Adam to exist is remote, in this sense, to Adam. And the same is true with respect to L and what causes L to be true.

But in case what I’ve just said isn’t all that clear, here’s a second, more straightforward, response to the imagined objector. ‘Remote past*’ is a technical term, a term of art that I developed. And what this term refers to is any thing, y, that comes prior (in any sense) to another thing x, such that x has no way of impacting, or influencing y in any way. And surely this is the case for any thing y that causes another thing, x’s, existence, or its being true (as in the case of L). For, how could x, before its having existed (or its having been true), impact or influence y in any way? It couldn’t; and, this is true about the relationship between x and y regardless of whether or not there is any temporal gap between them. Thus, I conclude that a good deal of sense can be made of the notion ‘remote past*’. The idea of a ‘remote past*’ captures the notion of both temporal and non-temporal senses of the term ‘prior’ while specifically denoting prior conditions that are specifically beyond the influence of any human agents.

It seems to me, then, that this sort of objection really *highlights* the point of the Direct Argument in the first place—and why I have revised the Direct Argument* as I have, below. The point of the Direct Argument (and Direct Argument*), I think, is something like this. Human agents are contingent beings that not only need not be here, they are—to borrow a phrase from Martin Heidegger—*thrown into the world*. We do not (indeed, *cannot*) bring ourselves into the world. Something else brings us into the world, and this something comes, in some sense, before us and is such that we have no influence over it. Now, if we're thrown into a deterministic world, and the something that comes, in some sense, before us is such that it determines our every step, then this is incompatible with our being responsible for those very steps. For, we're not responsible for what comes, in any sense, before us that is beyond our influence; and if we're not responsible for what comes, in any sense, before us that is beyond our influence, and what comes before us, in whatever sense, is such that it is beyond our influence and determines our every step, then we're not responsible for the every step that what comes before us, in that sense, determines. Thus, moral responsibility cannot exist in a deterministic world.

I conclude, then, that the imagined objection fails to undermine premise D3 (and D3*) of the argument that Adam has a remote past* (and that L has a remote past*). If we're willing to grant that D1 (or its variant) is true, and we're willing to grant that whatever causes (or brings about) Adam's existence (or L's being true) is beyond Adam's influence (because Adam isn't around to influence it), then I think we should conclude that whatever goes on prior to Adam's existence (or L's being true) is remotely past* to Adam (or L). Thus, whatever causes (or brings about) Adam's existence (or L's being true) is remotely past* to Adam (or L). Therefore, D3 (and D3*) is safe.

Here, then, is the Direct Argument*⁶⁹:

If determinism is true, then our acts are the consequences of the laws of nature and the remote past*. But we are not responsible for what is remotely past* to us, and neither are we responsible for what the laws of nature are. Therefore, we are not responsible for the consequences of these things (including our present acts).

More formally (where 'P*' stands for 'the remote past*'):

- | | | |
|-----|---------------------------------------|--------------------------------------|
| (1) | $\square (P^* \ \& \ L \supset p)$ | By definition of 'determinism' |
| (2) | $\square (P^* \supset (L \supset p))$ | 1, exportation |
| (3) | NR $(P^* \supset (L \supset p))$ | 2, Rule A |
| (4) | NR P^* | NR for what is remotely past* to one |
| (5) | NR $(L \supset p)$ | 3, 4, Rule B |
| (6) | NR L | NR for the laws of nature |
| (7) | NRp | 5, 6, Rule B. |

The No Past Objection fails to undermine the Direct Argument*. The No Past Objection fails to undermine the Direct Argument* because the No Past Objection only addresses temporal senses of the word 'prior', the sense being invoked in the term 'remote past'. Moreover, an important upshot of the Direct Argument* is that—as far as I can see—there are no possible No Past* Objections that can be raised against it. This is because necessarily, if determinism is true, then there is a remote past* and some laws of nature. But we aren't responsible for what is remotely past* to us, and neither are we responsible for what the laws of nature are. So, we aren't responsible for the consequences of these things either. And this is just what the Direct Argument* says.

⁶⁹ For what it's worth, *mutatis mutandis*, there is a Consequence Argument* that, I believe, also gets around the No Past Objection.

I think that the Direct Argument* is just what the Direct Argument means to argue in the first place. But even if it isn't, I think it's a successful version of the same sort of argument. So, I conclude that the Direct Argument* successfully shows that determinism and moral responsibility are incompatible.

IV. Objections, Round 1: A Potential Counterexample to (1) of the Direct Argument*

Premise 1 of the Direct Argument* goes like this:

(1) $\square (P^* \& L \supset p)$.

Here, then, is an objection to (1):

In the original Direct Argument, the thesis 'determinism' says that given a complete description of the state of the universe at any given time plus the laws of nature logically entails a complete description of the state of the universe at any *other* time. But with your substitution of 'remote past' with 'remote past*' you seem to be assuming that any deterministic universe must be such that its "ultimate causes" or "first causes" plus the laws of nature entail every truth about (for example) our behavior. But consider the following. Suppose that God is the ultimate first cause of the universe, and that he's free to create a wide range of "first moments" (there are numerous ways the Big Bang could happen, and God is free to actualize any of them). Suppose, further, that God creates a universe with a deterministic set of natural laws, L. In this deterministic world, (1) of the Direct Argument* is false: the remote past*, viz. God's existence, plus the deterministic laws, L, doesn't entail every other truth. God was, after all, free with respect to what the "first moment" was like; had he freely actualized a different Big Bang, some actually true proposition might have been false. Thus, at most what the Direct Argument* gets you is the conclusion that in deterministic universes that are such that the "first causes" plus the laws jointly entail every other truth, no one is responsible for anything they do.⁷⁰

This is an interesting objection to (1), but I think that it ultimately fails to provide a successful counterexample, and for two reasons. Here is the first: the objection seems to assume that there are deterministic universes without "first causes". This seems to me to be an objection to both

D1. Whatever contingently exists has a cause, and

D1*. Whatever is contingently true has a cause.

⁷⁰ I owe this objection to E. J. Coffman.

But, I think that D1 and D1* are fairly unassailable.⁷¹ That is, I think that there are no possible deterministic universes that lack a “first cause”, that is, something that is remotely past* to it that brings about its existence. So, e.g., if L is true, then it’s true because something remotely past* to it caused it to be true. Moreover, if Adam exists, then he exists because something remotely past* to him caused him to exist. Thus, if this objection is assuming that there are possible deterministic universes without “first causes”, then the counterexample fails. It fails because F2* (whatever causes L to be true is remotely past* to L) is true, and D1* (and D1) is true.

Here is the second reason this counterexample fails. There are no universes where L is true, but are such that there’s nothing remotely past* to L that, when conjoined with L, doesn’t entail every other truth. To be sure, God’s mere existence is remotely past* to L. And, if God exists and created the universe, it was possible—assuming, as we are, that God was free to choose to do otherwise—that God could have actualized a different world than the one he in fact did. But, again, L necessarily has a particular remote past*, viz., whatever it is that caused L to be true. (Suppose that God caused L to be true. Is it God’s *mere existence* that caused L to be true? No. Or, anyway, it better not be the case that God’s mere existence caused L to be true since we’re assuming that God was free to create (or refrain from creating).) Moreover, God’s mere existence conjoined with the laws, L, doesn’t suffice for the existence of a deterministic universe; rather, it must be that God’s existence, plus L, plus *God’s willing a particular way*, suffices for the existence of a deterministic universe.

Think about it like this. God is free to create either universe A or universe B, both having deterministic laws, but each having a different Big Bang. Whether or not A or B is the universe that exists depends on more than God’s mere existence and his causing L to be true; it depends on God’s decision to create A or B. But God’s decision to create A (or B) plus L entails every

⁷¹ Sort of. See the next section for more on this.

further true fact about A or B. Thus, there are no possible deterministic universes that include a remote past* and L, but don't entail every further true fact. Thus, (1) is true, and the imagined objection fails to provide a successful counterexample to the Direct Argument*.

V. Objections, Round 2: *Ex Nihilo Nihil Fit*?

My revised version of the Direct Argument, the Direct Argument*, depends on its being true that there is such a thing as a remote past*, or that it makes sense to talk of a remote past*. The way I tried to show that there is such a thing—and that such a thing makes sense—was by giving the argument D1 – F2 (and D1* – F2*). But I can imagine objections to two of the argument's premises, viz., D1 and D3 (and D1* and D3*). D2 and D4 (and D2* and D4*) I take to be uncontroversial, and F1 and F2 (and F1* and F2*) follow deductively from the truth of the premises. So if there's a flaw in the ointment (to borrow a phrase from Alvin Plantinga⁷²), it'll be with respect to either D1 or D3 (or D1* or D3*). Thus, in this section and the next, I wish to defend D1 and D3 (and D1* and D3*) from some potentially troubling objections. If I'm successful in so doing, I think that will give us good reason to think that there is such a thing as a remote past* (and that Adam and L have it). I'll start with an objection to D1 (and D1*).

Recall that the first premise of the argument that Adam has a remote past* says:

D1. Whatever contingently exists has a cause.

The undergirding thought behind D1 is what I take to be a truism, the idea that *ex nihilo nihil fit*.⁷³ That is, from nothing, nothing comes. It seems to me that if this is true—and I don't see how it could be false—whatever contingently exists comes from some thing that is not itself.

⁷² From, e.g., his (1974b, p. 106).

⁷³ What I say about D1 applies, *mutatis mutandis*, to D1* as well.

Call whatever this something is, ‘x’. And whatever this something, x, is, it can be said to be the cause of the thing that only contingently exists.

But I can imagine non-causalists about agency, such as Carl Ginet, arguing that the undergirding thought behind D1—that *ex nihilo nihil fit*—is false; so D1 (and its variants) is false. For, they think that agents bring about events (that is, they instantiate events, or make events obtain) without *causing* them when they perform actions freely. And since agent-involving events are only contingently instantiated, they are contingently instantiated things that are uncaused. They are things that come from—are caused by—nothing. Thus, D1 (and its variants) is false.

To be clear about how this line of reasoning works, consider one way in which Ginet describes a free action’s being up to an agent at a time (though the action is uncaused):

What is it for it to have been up to me...whether that event would occur? More specifically, what is it for it to have been up to me *at a certain time* whether that event would occur?...It was up to me at time T whether that even would occur only if I made it the case that it occurred and it was open to me at T to keep it from occurring; and it cannot have been up to me at T to keep it from occurring if whether it occurred depended entirely on facts in place by T; that is, it was up to me at T whether it would occur only if whether it would occur did *not* depend entirely on facts already in being at T. (Ginet, 2007, p. 245. Italics in the original)

So, if asked “what caused the event of the agent’s doing such and so?”, the response from the non-causalists would be something like “*nothing* caused the event of the agent’s doing such and so, if the agent’s doing such and so is freely done,” (or, maybe, “nothing *caused* the event of the agent’s doing such and so...”.) even though an agent *brought it about* that the event obtained, or *made the event happen*. If the non-causalists are right about this, then there’s at least a *prima facie* reason to reject D1. For, the event of the agent’s doing such and so is a contingently instantiated thing; yet, it comes from—is caused by—nothing. Thus, if the non-causalists are right, then D1 is false.

What is ironic, and not a little troubling, about an objection like this, is that Ginet is the father of original Consequence Argument, the argument that says that determinism is incompatible with free will!⁷⁴ What's important about this is that, if Ginet cannot endorse D1 – F2, then he cannot endorse the Direct Argument*. But since D1 – F2 is an argument purporting to show that Adam has a remote past*, it's an argument that, if successful, saves not only the Direct Argument* from the No Past Objection, it also saves the Consequence Argument* from the No Past Objection.⁷⁵ Thus, if Ginet cannot endorse D1 – F2, he cannot endorse the Consequence Argument*, a revision of the Consequence Argument that replaces 'remote past' in the definition of 'determinism' with 'remote past*'.

Be that as it may, there are two ways in which I'd like to respond to this sort of objection to D1. The first is as follows. We should reject non-causalism. For, if non-causalism is true, then D1 is false; but, D1 looks extremely plausible; so, non-causalism seems implausible. And if we have a choice between accepting something plausible versus accepting something implausible, I think we should (generally speaking) accept the thing that seems plausible instead of the implausible thing where those two things are in conflict. Thus, we should reject non-causalism.

The second way is this. Note that non-causal views do *not* necessarily suggest that the truism, *ex nihilo nihil fit*, is false. While the non-causalist is happy to grant that contingent things can be brought about without having been caused (and so, nothing caused them to be

⁷⁴ See, for example, his (1966).

⁷⁵ What is the Consequence Argument*? It's the revised version of the Consequence Argument that doesn't succumb to the No Past Objection. Here it is, informally:

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past*. But it is not up to us what is remotely past* to us, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.

brought about), they do not have to commit themselves to the idea that something can come from nothing. This is because the Ginet-type non-causalist will think that uncaused actions *did* come from something, namely, an agent.⁷⁶ Thus, the non-causalist can agree with the undergirding intuition behind a premise like D1. Because this is so, I think that D1 can be modified in such a way that it's consistent with non-causal views of agency while saving the conclusion that for any contingently instantiated thing, that thing has a remote past*.

To see how properly to modify D1 in light of the envisaged objection, consider another way in which Ginet describes an agent's bringing it about that she acts a certain way.

For it seems evident to me that, given that an action was uncaused, all its agent had to do to make it the case that she performed that action was to perform it. If my deciding to vote for the motion, for example, was uncaused, then it follows that nothing other than me made it the case that I decided to vote for the motion, and it also follows that I made it the case that I so decided: I did so simply by so deciding. If my raising my hand was uncaused (that is, *nothing other than me* determined or made it the case that I raised my hand), then I made it the case that I raised my hand simply by raising my hand. (Ibid., p. 247. My italics.)

So, the idea is that there are some contingently occurring things (namely, actions), that can be 'brought about' or 'made the case' by an agent, without having, thereby, been caused by the agent. So, if Ginet (or any other non-causalist) is right, then I think we can modify D1 as follows (even if it reads a bit clunky):

D1.** Whatever contingently exists is either caused, or brought about (or made the case).

Such a revision would make F1 read as follows:

F1.** So, Adam's existence is either caused, or brought about (or made the case),

and F1* read as follows:

F1*.** So, L's being true is either caused, or brought about (or made the case).

⁷⁶ Moreover, Ginet maintains that uncaused actions can be explained in terms of an agent's having reasons for performing said actions. See, e.g., the discussion on pages 251 – 252 of his (2007).

This would, of course, change D4 thusly:

D4.** Whatever either causes, or brings about (or makes the case) Adam's existence comes, in some sense (e.g. 'causally', 'logically', 'temporally', etc.), *before* Adam's existence, and Adam has no influence as to whether something either causes, or brings about (or makes the case) his existence.

And D4** thusly:

D4*.** Whatever either causes, or brings it about (or makes it the case) that L is true comes, in some sense (e.g. 'causally', 'logically', 'temporally', etc.), *before* L's being true and L has no influence as to whether something either causes, or brings it about (or makes it the case) that L is true.

F2.** So, whatever either causes, or brings it about (or makes it the case) that Adam exists is remotely past* to Adam,

and F2*:

F2*.** So, whatever either causes, or brings it about (or makes it the case) that L is true is remotely past* to L.

Even so, I think, as we can see in F2** and F2***, such revisions are consistent with the conclusion that for any thing, x, if x contingently exists (or, if x is contingently true), then x has a remote past*. And if that's right, then non-causalists can endorse the Direct Argument* simply by making the necessary modifications to the argument D1 – F2 (and D1* – F2*). Thus, the non-causal worries about D1 do not at all undermine the Direct Argument*.

VI. Conclusions

Having defended the Direct Argument's inference rules—Rule A and Rule B—from alleged counterexamples, the remaining task for a full defense of the Direct Argument was to deal with the No Past Objection. Thus, the point of this chapter was to defend the Direct Argument from said objection. As I argued, however, the No Past Objection seems successful against the Direct Argument as it is originally stated; for, the Direct Argument relies on a transfer principle about non-responsibility for the remote past to non-responsibility for the

present. So, my defense of the Direct Argument in this chapter resulted in my having given a revised version of the Direct Argument, the Direct Argument*.

As I've shown, the Direct Argument* is the same sort of argument as the Direct Argument; the one crucial difference is that instead of relying on a transfer principle about non-responsibility for the remote past to non-responsibility for the present, the Direct Argument* relies on a principle about non-responsibility for the remote past* to non-responsibility for the present. One important advantage of the Direct Argument*'s relying on such a principle, I claimed, is that there are no possible No Past* Objections. I claimed this because, as I've shown, necessarily, for any contingently existing thing, *x*, *x* has a remote past*. And there are no possible deterministic worlds where this is not true.

This claim, you'll recall rested (essentially) on two controversial premises.

D1. Whatever contingently exists has a cause, and

D3. If a thing *y* comes, in some sense (e.g. 'causally', 'logically', 'temporally', etc.), *before* another thing *x*'s existence, and *x* has no influence as to whether *y*, then *y* is remotely past* to *x*.

The envisaged objection to D1 was from a non-causalist perspective about human agency. That is, said objection has in mind the ideas of philosophers like Carl Ginet who think that agents perform free actions without having caused them. So, since free actions are contingently instantiated things, if it's true that agents perform free actions without having caused them—and nothing *else* causes their instantiation—then it follows that free actions are uncaused. And if so, then free actions are a counterexample to D1.

I argued, however, that this need not worry the proponent of the Direct Argument*. For, there are two plausible ways to respond to such an objection. First, since D1 is extremely plausible, its denial is implausible. Thus, since non-causalism leads to the denial of D1—

something that is implausible—we have good reason to reject non-causalism. Second, there isn't, after all, a *need* to reject non-causalism⁷⁷ simply to save D1. D1, I argued, can be modified to account for non-causal theories of agency. All D1's defender would need to do (and as I've done in D1**) is make D1 an either/or claim. That is, she'll need to adjust D1 to read that whatever is contingently instantiated is *either* caused *or* brought about (or made the case). This is perfectly consistent with the non-causal theory of agency.⁷⁸

The envisaged objection to D3 had to do with the concept of a 'remote past*' and whether or not such a concept makes any sense. The objection, you'll recall, argued that the concept 'remote past*' is temporally loaded, and thus of no use to the Direct Argument*'s defender. But I argued that this objection is mistaken. For one thing, the 'remoteness' to the idea of a 'remote past' in the Direct Argument hasn't anything to do with temporality at all; rather, it has to do with the fact that whatever is *remote* to a person is beyond that person's ability to impact it, or influence it. This same notion of 'remoteness' is captured in the concept 'remote past*'. Moreover, it's false that 'past*' is temporally loaded. For, 'past*' captures the idea of *priority*, non-temporal senses of priority (especially) included. And since whatever it is that causes it to be the case that (or brings it about that, or makes it the case that) some thing, x, is instantiated is, itself, *prior*, or *before* (in some sense) x, then that thing is past* to x—indeed, remotely so. Thus, I concluded that the concept of a 'remote past*' *does* make sense, and so D3 is unscathed by the imagined objection.

Finally, there was the objection to (1) of the Direct Argument*, the objection claiming that there are possible deterministic universes that are such that the remote past* plus the laws of

⁷⁷ At any rate, there isn't a need to reject non-causalism in order to save D1. There might, however, be other reasons to reject non-causalism.

⁷⁸ And, of course, as I've done in §V, the remaining premises of the argument would need to reflect said changes to D1. But these changes don't do any violence to the argument, itself.

nature do *not* entail every further true fact about that universe. But I argued that this objection fails because the counterexample rests on two mistakes. First, there are no deterministic universes where L is true but lack “first causes”. And, second, there are no deterministic universes where L is true but whose remote past*, when conjoined with L, fails to entail every further true fact about that universe.

This concludes my defense of the Direct Argument. Or, more strictly, this concludes my defense of the Direct Argument*. As I’ve argued, the Direct Argument*’s inference rules are safe (at least so far) from counterexample. And, moreover, the Direct Argument* is impervious to the No Past Objection. And since there seem to be no other reasons to reject the Direct Argument*, I conclude that the Direct Argument* is sound. But since the Direct Argument*’s conclusion entails incompatibilism about moral responsibility and determinism, I conclude that since the Direct Argument* is sound, incompatibilism is true. Thus, incompatibilism about moral responsibility and determinism is true.

Such conclusion is important. For one thing, it solves a centuries old philosophical puzzle. For: if I’m right, we can know that incompatibilism is true. But I think there are additional important upshots to such conclusion in the offing. In the next chapter I’ll take to a discussion of what I think are some important take-aways from the success of the Direct Argument*. Moreover, I’ll strengthen my argument for the soundness of the Direct Argument* by showing that counterexamples to Rule B—the Direct Argument*’s most controversial inference rule—are *impossible*. Importantly, this will show that not only are the alleged counterexamples considered in chapter two unsuccessful, they are necessarily unsuccessful since no counterexamples to Rule B are possible.

CHAPTER SIX

Looking Forward: Ramifications of a Successful Direct Argument*

I have argued that the Direct Argument* is successful. That is, I have argued that the Direct Argument* successfully shows that causal determinism and moral responsibility are incompatible. If I'm right about all of this, then I think there are at least two sufficiently interesting, and salient ramifications. So, in this chapter, I intend to examine them.

Here, I think, are the two most important upshots of the Direct Argument*'s success: 1) plausibly, we can know that indeterminism is true; and 2) we can know that free will is compatible with indeterminism. This last upshot might seem fairly innocuous, but I think it is both important and surprising. It is surprising because it's an upshot that comes by way of an argument that makes no use whatever of the concept 'free will' and what such a concept may or may not require. It's important because, even though I have nowhere argued about free will, one take-away from all I've argued so far—or, so I'll go on to show—is that the Direct Argument*'s conclusion will help guide the *free will* debate going forward.

I'll take to a discussion of these two take-aways in the order that I've presented them just now. But, first, I wish to redeem a promissory note that I made back in chapter two. That is, first, I want to discuss, at length, Truth Dependence_{MORAL} (TDM), a principle I introduced in chapter two. This will serve two functions. The first is that it will strengthen my argument in chapter two, the chapter in which I defend Rule B from various recent attempts to provide a counterexample. My discussion of TDM will strengthen chapter two's argument because if TDM is true, then counterexamples to Rule B are impossible. And, this, of course, is highly important to the relevant discussion. For, I will argue that by reflecting on an obvious fact about

truth, we can see that there's good reason to think that TDM is true. Thus, it falls right out of my overall conclusion—viz., that an obvious fact about truth implies that there can't be counterexamples to Rule B—that compatibilism commits its adherents to denying an obvious fact about truth.

The second function of my discussing TDM is that such a discussion will allow me to actively *defend* TDM, to show that TDM *is* true. So, I'll do this first. Then, in the sections following, I'll discuss the two important upshots of a successful Direct Argument*.

II. Counterexamples to Rule B Are Impossible

To begin my defense of the claim that counterexamples to Rule B are impossible, I follow Trenton Merricks (2007, 2009, 2011a, 2011b) by noting that truth depends (in a very trivial way) on the world. It's true, for example, *that dogs bark* because dogs bark; it's true *that Turner exists* because I exist; *that I write this paper at t* is true because I write this paper at t, and so on. No one should disagree with this “truism about truth” (Merricks, 2009, p. 31). That is, no one should disagree that truth depends on the world in this trivial way.⁷⁹

And I do not think anyone *would* disagree with this truism about truth. But I think that reflection on this truism will reveal that any alleged counterexamples to Rule B must fail. They must fail because it's impossible to give a *bona fide* counterexample to Rule B given this truism about truth.

⁷⁹ This isn't to say that there's nothing more substantive to say about the way truth depends on the world. One might wonder, for example, how true counterfactuals can 'depend on the world'. But I leave aside discussions of such analyses of truth—e.g. Truthmaker theory, Correspondence theory, Truth Supervenes on Being (TSB) theory, etc.—because a discussion of such analyses not only takes us too far afield, it's also irrelevant to the present discussion. What's crucial to my project is that all agree that what's true depends on how things are; how things are doesn't depend on what's true. So, e.g., if it's true *that if Roger is in C, then he would A*, this is true (minimally) because I would A if I was in C. Perhaps there's more that *can* be said about how this (assumed) truth 'depends on' the world, but it's not important for the present discussion. Even so, for a thorough treatment of the competing theories about truth, see Merricks (2007).

To begin to see why I say this, note that Merricks (2009, 2011a, 2011b) argues that the following is a corollary to the truism about truth (what I'll call):

Truth Dependence_{CHOICE} [TDC]: For all agents, S, and all propositions, p, if S has a choice about what p's truth depends on (in the sense of 'depends on' in which truth depends on the world), then S has a choice about p's truth.

To see that TDC is a corollary of the truism about truth, imagine that the proposition *that Jones kills Smith* is true. Given the truism about truth, *that Jones kills Smith* is true because Jones kills Smith; that is, *that Jones kills Smith* depends on what Jones does. And if Jones has a choice about whether or not he kills Smith, it seems easily to follow that he has a choice about whether or not it's true *that Jones kills Smith*. For, suppose that we thought Jones *doesn't* have a choice about the truth of *that Jones kills Smith*. We'd think this because we'd think Jones doesn't have a choice about whether or not he kills Smith. So, not only do we see that there's a close connection between 'having a choice about' and the nature of truth, we can see that the truism about truth reveals a principle about 'having a choice about' the truth of a proposition, namely, TDC.

I think that something similar is true about the relationship between moral responsibility and the nature of truth. To see this, suppose it's true *that Jones kills Smith*. It's true *that Jones kills Smith* because Jones kills Smith; that is, the truth of *that Jones kills Smith* depends on Jones and what he does. So, given the truism about truth, it follows that *that Jones kills Smith* would not have been true had Jones not acted as he did.

Now, if all of that is right, then I think the truism about truth has the following corollary in addition to TDC:

Truth Dependence_{MORAL} [TDM]: For all agents, S, and all propositions, p, if S is directly morally responsible for that which p's truth depends on (in the sense of 'depends on' in which truth depends on the world), then S is at least partly directly morally responsible for p's truth.

And if TDM *is* a corollary to the truism about truth, then, as I'll go on to argue, since the truism about truth is necessarily true, counterexamples to Rule B are impossible. To see why I say that TDM is a corollary to the truism about truth, notice that given the truism about truth and Jones's moral responsibility for killing Smith, it follows that *that Jones kills Smith* would not have been true had Jones not acted as he did. Now, suppose that we thought that Jones isn't directly morally responsible for the fact *that Jones kills Smith*. I say we'd think this because we'd think that Jones isn't directly morally responsible for the thing upon which the truth of *that Jones kills Smith* depends, viz., Jones's killing of Smith. And this generalizes. Thus, for all S and all p, if S is directly morally responsible for the thing upon which the truth of p depends, then S is at least partly directly morally responsible for p's truth.

If TDM is true, so I'll go on to show, then it is impossible to give a counterexample to Rule B. So, what I intend to do in this section is the following. First, I will show how TDM's (assumed) truth reveals a way in which two recent alleged counterexamples to Rule B fail. But these reasons will generalize; so, TDM's (assumed truth) will reveal how *all* alleged counterexamples fail. Second, I will defend TDM from two objections, one that purports to attack TDM indirectly by attacking a principle that TDM allegedly entails, and another that attacks TDM directly. I'll conclude that these objections fail to undermine TDM; so if TDM is true, its truth undermines all attempts to provide a counterexample to Rule B.

Such conclusion is important. For, I think that TDM is intuitively plausible. Moreover, I think that if the truism about truth is true, then TDM is true. That is, I think that the truism about truth implies TDM. Thus, to deny TDM (which any compatibilist must, assuming (i) that Rule B implies incompatibilism, and (ii) that there can be counterexamples to Rule B if it's invalid) will come at great cost; for, the objector to TDM will have to deny the truism about truth.

I now wish to consider two recent alleged counterexamples to Rule B. My aim is to show that counterexamples to Rule B are in fact impossible; so, I use these two recent purported counterexamples in order to illustrate this point. I begin by revisiting two alleged counterexamples from chapter two, viz., John Martin Fischer and Mark Ravizza's 'Erosion' case, and Ishtiyaque's Hal's Creation(2). As we've seen, each case is a different *type* of alleged counterexample. Fischer and Ravizza's case is a case of overdetermination, while Haji's case is a case of libertarianly free action—that is, a free action done in a non-deterministic universe—where there is *no* overdetermination involved. Importantly, I'll show that all of these cases fail to provide a counterexample for the same reasons. Thus, such reasoning will generalize and *all* alleged counterexamples to Rule B will fail.

I begin by revisiting an alleged counterexample to Rule B, John Martin Fischer and Mark Ravizza's Erosion case. Recall:

Erosion: Imagine that Betty [a soldier charged with destroying an enemy fortress] 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 at T2, starting a natural avalanche that would have crushed the enemy camp at T3. (Fischer and Ravizza, 1998, p. 157)

Erosion, you'll remember, is supposed to generate the following counterexample to Rule B:

1. NR The glacier is eroding
2. NR (The glacier is eroding $\supset$ there is an avalanche that crushes the enemy base at T3)
3. NR There is an avalanche that crushes the enemy base at T3.

But, even apart from my arguments in chapter two, I don't think that Erosion successfully shows that Rule B is invalid; and reflection on the truism about truth will help us to see why. Recall that it's true *that Turner exists* because I exist, and it's true *that I write this paper at t* because I

write this paper at t . The truism about truth asserts just the notion that truth depends on the world in this very trivial way. And, as I argued above, the truism about truth has TDM as a corollary.

Now, there are objections to TDM (or something like it) in the offing, and I'll consider those in the next section. For now, however, let's assume that TDM is true. Given TDM's truth, does Erosion supply a successful counterexample to Rule B? I think that it does not.

To begin to see why not, let's assume, with Fischer and Ravizza, that Betty is morally responsible for the fact that an avalanche crushes the enemy base when it does. Now, let's write the equivalent of the conditional in 2 as follows:

2.** $\sim (\text{The glacier is eroding}) \vee (\text{There is an avalanche that crushes the enemy base at } T3).$

Since the first disjunct of 2** is false, we can see that 2** is true just in virtue of the fact that the second disjunct is true. But notice that 2**'s truth depends on the fact *that there is an avalanche that crushes the enemy base at T3*. Since Betty is, *ex hypothesi*, morally responsible for the fact *that there is an avalanche that crushes the enemy base at T3* because she's morally responsible for the avalanche that crushes the enemy base when it does, we can conclude on the basis of TDM that 2* is false. That is, Betty *is* morally responsible for the fact expressed in 2** because 2**'s truth depends on (in the sense of 'depends on' in which truth depends on the world) the fact *that there is an avalanche that crushes the enemy base at T3*, something for which Betty is *ex hypothesi* morally responsible. Therefore, Betty, *contra* Fischer and Ravizza's claim, is morally responsible for the fact that lies within the scope of the 'NR' operator in 2; and so, 2 itself is false. Thus, Erosion fails as a counterexample to Rule B.

Now, Fischer and Ravizza might respond as follows.

You've failed to show that Betty is morally responsible for the conditional contained in 2 because you've failed to show that the relevant portion of 2's truth depends on anything that Betty has done. By hypothesis, *that there is an avalanche that crushes the enemy base at T3* would have been true no matter what Betty did; there was an 'ensuring condition'—namely, the erosion of the glacier—that would have led to the crushing of the enemy base even if Betty had never planted the explosive device. Thus, you've failed to show that Betty's actions are what the truth of 2 depends on; so, you've failed to show that the truth of 2 depends on Betty. Erosion hasn't yet been undermined.

But in reply I ask the following question: Is Betty morally responsible for the fact *that there is an avalanche that crushes the enemy base at T3*, or not? If she *is* responsible for the fact *that there is an avalanche that crushes the enemy base at T3*, and the truth of the conditional that lies within the scope of the 'NR' operator in 2 depends on this fact, then, given TDM, Betty is morally responsible for the truth of the conditional that lies within the scope of the 'NR' operator in 2.

Now, if Betty isn't directly morally responsible for the fact *that there is an avalanche that crushes the enemy base at T3*, then, of course, there is no problem for Rule B. In such a case, Rule B is confirmed rather than refuted. In any case, equipped with TDM, we can see that Erosion fails as a counterexample to Rule B.

Next, let's revisit Haji's Hal's Creation(2) case. Recall:

Hal's Creation(2): Hal-2 [an essentially omniscient, sempiternal, amoral—that is, lacks knowledge of moral right, wrong, or obligatory—being] has the ability to create (or actualize) any one of an infinite number of possible worlds... Suppose Hal-2 creates a world, W1, in which, after due reflection, Yasmin [in a libertarianly free way] donates a large sum of money (at some time, ts) to a credible charity, UNICEF. Yasmin really cares about the plight of the needy children; she donates because she wishes to help the kids and not, for instance, because she wants a big tax break. We safely suppose that she is morally praiseworthy for her bountiful donation. Under appropriate circumstances normal agents would be deserving of praise for such an act.

Further, recall that Hal's Creation(2) generates the following substitution instance of Rule B:

Hal's Argument(2):

(4H): NR (Hal-2 creates W1).

(5H): NR (Hal-2 creates W1 $\supset$ Yasmin donates (at ts in W1)).

(6H): Therefore, NR (Yasmin donates (at ts in W1)).

But, as we can now see, there's a further reason to think that Hal's Creation(2) has failed to provide a counterexample to Rule B. For, equipped with TDM, we can see yet another reason to think that 5H is false.

For, as we did with the Erosion case, we can see that 5H's conditional can be written this way:

(5H*): $\sim$ (Hal-2 creates W1) $\vee$ (Yasmin donates (at ts in W1)).

As with 2** from the Erosion case, we can see that since the first disjunct of 5H* is, *ex hypothesi*, false, 5H* is true just in virtue of the truth of the second disjunct. That is, 5H* is true *because* Yasmin donates to UNICEF when she, in fact, does. But Yasmin is responsible for this fact. Thus, given TDM, Yasmin is at least partly directly morally responsible for the truth of 5H*. And if Yasmin is at least partly directly morally responsible for the truth of 5H*, then it follows that she's at least partly directly morally responsible for the truth of the conditional contained in 5H of Hal's Argument(2). So, 5H itself is false, and Hal's Creation(2) is no counterexample to Rule B.

I've shown that Erosion, and Hal's Creation(2) fail to provide successful counterexamples to Rule B if TDM is true. But it should be clear that I've shown more than that. For, since every alleged counterexample to Rule B will have to be of the same form, every alleged counterexample to Rule B suffers the same affliction. Namely, in each case, someone is (or was, or will be) *ex hypothesi* directly morally responsible for the thing upon which the truth of the consequent contained within the second premise of each substitution instance of Rule B depends. And *this* means that someone is (or was, or will be) at least partly directly morally

responsible for the thing upon which the truth of the conditional contained within the second premise of each substitution instance of Rule B depends. Thus, if TDM is true, counterexamples to Rule B are impossible.

III. Is Truth Dependence_{MORAL} True?: Objections to TDM

I will now consider two potential objections to Truth Dependence_{MORAL}. The first objection I'll consider is from John Martin Fischer and Eleonore Stump (2000) that casts doubt on a principle that TDM seems to entail, and, thus, seems indirectly to attack TDM. The second objection I'll consider attacks TDM directly.

In recent work, Fischer and Stump attempt to undermine an argument by Ted Warfield (1996). To begin to see how Fischer and Stump's potential criticism of TDM will go, consider:

Erosion*: [The details are the same as Erosion except that, in this case,] the conditions of the glacier *do* actually cause the ice and rocks to break free, triggering an avalanche that arrives at the fortress precisely at the same time as the independent avalanche triggered freely by Betty. Each avalanche is sufficient for the destruction of the enemy fortress.⁸⁰

Fischer and Stump think that Erosion* is a counterexample to Rule B. Ted Warfield (1996) seems to agree; so, Warfield constructs the following revision of Rule B, call it:

Rule Beta $\Box$: $NRp, \Box (p \supset q) \vdash NRq$

The upshot of Rule Beta $\Box$ is that it's supposed to make it much more difficult to construct counterexamples like Erosion* because the connection between the eroding glacier at T1 and the destruction of the enemy fortress at T3 has to be one of broadly logical necessity. But, of course there is no such connection.

Fischer and Stump, however, think that there *are* counterexamples to Rule Beta $\Box$ that can be constructed, and that don't presuppose causal determinism. Consider the following:

⁸⁰ Initially from Mark Ravizza (1994), but also found in Fischer and Ravizza (1998, p. 160ff).

Counterexample B: [Given the set up of a case like Erosion*] let ‘r’ be a conjunction of these propositions:

(r1) the actual laws of nature obtain, and

(r3) there is an avalanche, which destroys the enemy camp at T3.

Now, without doubt, there is a logically necessary connection between r and q (since q is identical to r3), but the question of whether causal determinism of any sort obtains is irrelevant. Here we have:

(1) NRr, and

(2) $\Box (r \supset q)$, but it isn’t the case that

(3) NRq. (Fischer and Stump, 2000, p. 50 – 51)

Now, as Fischer and Stump rightly point out, Warfield anticipates this sort of attempt at a counterexample. In so doing, he presents a thesis that Fischer and Stump call

(W1): If no one is even partly morally responsible for a conjunction, then no one is even partly morally responsible for either conjunct of the conjunction. (Ibid., p. 51)⁸¹

The idea, here, is that if W1 is true, then (1), from Counterexample B, is false since it’s not the case that no one is even partly morally responsible for r3, a conjunct of r. So, (1) is false.

But, Fischer and Stump think that such a conclusion is mistaken, and this because of the relation between conjunctions and conditionals. To begin to see the worry, notice that, according to the details of Erosion* and Counterexample B, it’s not the case that if the actual laws of nature obtain, then there won’t be an avalanche that destroys the enemy base at the time it does. That is:

(4) $\sim (L \supset \sim q)$

is true. Moreover, (4) is trivially equivalent to:

(5) $(L \ \& \ q)$.

⁸¹ But originally from Warfield (1996, p. 218).

Now, Fischer and Stump think that it's counterintuitive to think that anyone is morally responsible for the fact expressed in (5) because it's strange to think that anyone could be morally responsible for the fact expressed in (4). Thus, Fischer and Stump think there's a problem with accepting a principle like W1.

And more problems abound, they claim. For, if W1 is true, then the following is true, too:

(W2): Given a true antecedent of a conditional, a person is partly morally responsible for the conditional's being false if he is partly responsible for the falsity of the consequent of the conditional. (Ibid., p. 52)

Moreover, if W2 is true, then Fischer and Stump think (and I agree) that the following ought to be true as well:

(W3): Given a true antecedent of a conditional, a person is partly responsible for the conditional's being true if he is partly responsible for the truth of the consequent of the conditional. (Ibid.)

Thus, if W1 is true, then W3 is true. This, Fischer and Stump conclude, is an untoward consequence of accepting a principle like W1.

If W1 entails W3, then one way to show that W1 is false is by showing that W3 is false. This is exactly what Stump and Fischer attempt to do. To see why Fischer and Stump think that W3 is false, recall:

(2): $\Box (r \supset q)$.

Now, by Rule A of the Direct Argument—which says that no one can be even partly directly morally responsible for a necessary truth—we get the fact that nobody is even partly directly morally responsible for the fact expressed in (2). (2), after all, is a necessary truth; for it's just trivial that, necessarily, the conjunction of the actual laws of nature and q implies q . But, if W3 is right, then it follows that someone *is* morally responsible for the fact expressed in (2); *Betty* is

responsible for the fact expressed in (2) because she's responsible for q , and (2)'s truth depends on q 's being true. So, Warfield is left with a dilemma: either Rule A—the Direct Argument's inference rule that says no one is (or could be) even partly directly morally responsible for a necessary truth—is false, or W3 is false. If W3 is false, then W1 is false, and Warfield's defense Rule Beta $\square$ fails. On either horn, the Direct Argument (or, more importantly, the Direct Argument*) gets skewered.

Moreover, Truth Dependence_{MORAL} appears to entail W3. For, W3, so the objection goes, merely expresses one way in which a person can be responsible for the truth of a proposition: if a conditional has a true antecedent, and a person is directly morally responsible for the truth of the consequent, then the person is at least partly directly morally responsible for the truth of the conditional. If this is what TDM entails, and if Fischer and Stump's objections to W3 go through, then TDM is defeated.

But I think that Fisher and Stump's argument against W3 does *not* defeat TDM; for, I think that W3 is *not* entailed by TDM. In what remains of this section, I will defend TDM from the claim that it entails W3.

To begin to see how my defense will go, recall:

(2): $\square (r \supset q)$.

(2) expresses the following fact:

DEB: Necessarily, if the actual laws of nature obtain and the enemy base is crushed by an avalanche at T3, then the enemy base is crushed by an avalanche at T3.

Fischer and Stump think that, if W3 is true, then it follows that someone—namely, Betty—is responsible for DEB. And, so the objection continues, since TDM entails W3, it follows by TDM that Betty is responsible for DEB, a necessary truth. Thus, if TDM is true, Rule A is false.

But, I think that this claim is too hasty; for, DEB's truth doesn't depend on Betty. For, DEB is true even if it isn't the case that the enemy base is crushed when it, in fact, is. Indeed it's difficult to say just what DEB's truth *does* depend on; it's a trivial truth. This is important: TDM speaks in terms of being responsible for a proposition whose truth *depends on* what an agent does. But DEB's truth is not like this; it does *not* depend on Betty or what she does.

Why do I say that DEB's truth (and, so, (2)'s truth) doesn't depend on Betty? Well, for one thing, DEB is true even if it isn't the case that the enemy base is crushed when it in fact is. Moreover, DEB's truth doesn't depend on Betty in the relevant way—the way relevant for moral responsibility—because it's true regardless of how things are. But, if DEB's truth is to *depend on* Betty in the relevant way, this can't be the case; it has to be that DEB's truth hinges on whether or not things *are* a certain way; viz., that things are such that Betty acts in a particular way. Or another way to put the point is this: some truths require a truthmaker; I think that DEB is not one of those, but even if it is, Betty is not its truthmaker. So, DEB's truth *doesn't* depend on Betty. But, since TDM requires that DEB's truth depend on Betty (or someone, anyway) in order for her (or anyone) to be directly morally responsible for its truth, even if W3 is false, TDM is unharmed.

Now, I agree that TDM entails a principle *similar* to W3, but I argue that any such principle must have an appeal to the relevant notion of dependence. For, suppose that it's true *that if hobbits don't exist, then I write this paper*. Given that it's true *that hobbits don't exist*, the foregoing conditional is true if and only if I write this paper. Thus, I'm responsible for the conditional's being true if and only if I write this paper. Now, suppose that I'm not responsible for its being true *that if hobbits don't exist, then I write this paper*. If I'm not responsible for the truth of this conditional, then this is because I'm not responsible for writing this paper; that is,

I'm not responsible for the thing upon which the truth of the conditional depends. Thus, I conclude that any W3-like principle that TDM entails must include the relevant notion of dependence.

I agree with Charles Hermes (forthcoming) that "any plausible account of moral responsibility must make the following disjunction principle true:"

DPR: If a disjunction has only one true disjunct, and an agent is responsible for the truth of that disjunct, then the agent is responsible for the truth of the disjunction. (Ibid)

Here is Hermes's justification for thinking that DPR is a required inference for any plausible account of moral responsibility:

[S]uppose that Mary knows that one of her two children broke her favorite vase and wants to know who is responsible. Mary would discover what she wanted to know by learning that Johnny is the only person responsible for breaking the vase. After all, if Johnny is the only person responsible for breaking the vase, then Johnny is also responsible for the fact that one of her children broke the vase. It is easy to see how Mary can derive this conclusion, if DPR is valid. If DPR is invalid, Mary's inference appears to be equally problematic. (Ibid.)

Or think of it this way. Suppose that I murder Jones. By murdering Jones I am obviously morally responsible for the fact *that Roger murders Jones*; however, I am also morally responsible for the fact *that one of the people at the University of Tennessee murders Jones*. So, I agree that any plausible account of moral responsibility will make DPR true, but this is because I think that DPR contains within it an implicit appeal to the relevant notion of dependence to which TDM refers. For suppose it's true that $\sim p \vee q$ (and p is true). This disjunction is true *just because* q is true; that is, this disjunction *depends on* q in the sense of 'depends on' in which truth depends on the world. So, DPR contains an implicit appeal to the relevant notion of dependence to which TDM refers. Moreover, if DPR is true, then Rule B is *established* and with it the Direct Argument (or, perhaps, the Direct Argument*).

But W3 is not like DPR in its appeal to the relevant notion of dependence. Thus, I conclude that TDM does *not* entail W3; it entails a restricted W3, one that appeals to the relevant notion of ‘dependence’ to which TDM refers. Thus, Fischer and Stump’s argument against W3 fails to show that TDM is false.

To see what I take to be a more worrisome objection to TDM, let’s revisit the Zombie Case from chapter two. Recall:

Sara does cutting edge scientific research. She knows her craft well. She fulfills the requisite requirements to be morally responsible for her research, and reports her research to her boss, Ted. Unbeknownst to Sara, however, Ted is a mastermind controlling a large conglomerate of labs. Ted uses Sara’s work, along with the work of many other scientists (whose work we can safely assume Sara would not grasp), to create a virus that, when released, turns half of the world into flesh-eating zombies. Sara bears moral responsibility for a part of the way things are: namely, that her research took place. But Sara doesn’t bear part of the moral responsibility for the way things are. It would be inappropriate to blame Sara for the zombie outbreak. Plausibly, Sara is non-culpably ignorant, and so gets off the hook.

The Zombie Case is, I think, supposed to show that someone could be morally responsible for some fact but lack the epistemic requirements for responsibility with respect the ‘dependent’ fact. So, for example, Sara bears moral responsibility for the fact *that her research took place*, but it’s alleged that she bears no responsibility for the fact *that her research took place and there’s a zombie outbreak*, and this because she doesn’t have any good reason to believe that the truth of this second fact—this conjunctive fact—depends on her.

But, I think that this objection can be met with the brick-maker story from chapter two.

Recall:

Suppose that I am a mason, tasked with the making, and laying, of a single brick that’s part of the foundation of a beautiful mansion. And suppose that my creating and laying this brick is, other things equal, a morally praiseworthy action. Now, suppose, also, that all I know how to do, at least when it comes to building things, is make and lay bricks; I don’t know anything about engineering, blue-prints, or anything else relevant to the task of building a house. Moreover, I’m not so much as *capable* of knowing such things (for whatever reason).

According to TDM, if I'm directly morally responsible for the fact *that this particular brick is made* and the truth of *that this particular brick is made and the beautiful mansion is built* depends on whether or not I make this particular brick, then I'm at least partly directly morally responsible for this conjunctive fact; I'm partly directly morally responsible for the fact that things are such that *that this particular brick is made and the beautiful mansion is built*.

Back in chapter two, I asked the reader to further suppose that the homeowner thanked me (as a type of moral praise) for the fact *that her house is built*, and then asked whether or not we should think the homeowner is out of line in her so thanking. I said, then, that it wasn't obvious to me that the homeowner *is* out of line in her thanking me. But let's assume that she is. Does this show that TDM is false? I think that it does not. For, TDM doesn't imply that I am responsible for the fact *that her house is built*; rather, TDM implies that I am responsible for the fact *that this particular brick is made and her house is built*.

So, it seems to me that meeting the epistemic conditions for responsibility for the making of the brick suffices for meeting the epistemic conditions for responsibility for the whole of which the brick plays a part. Moreover, I think that this objection confuses what's at issue. TDM doesn't imply that a person is responsible for all of the conjuncts in a conjunction, for example. Nor does it imply that a person is responsible for the antecedent in a conditional. All it implies is that a person is responsible for the truth of the conjunction, or the truth of the conditional (or, etc.) given that the truth of those things depends on what they do. For, suppose it's true *that Sara's research took place*. Now, suppose that Sara isn't responsible for this fact. I think the only thing to conclude here is that she's not responsible for this fact because she's not responsible for the thing upon which the truth of the proposition depends, viz., doing her research. But this is just to say that if she's responsible for the thing upon which the truth of the

proposition depends, then she's responsible for the truth of the proposition. And given the Zombie Case, the truth of the proposition *that her research took place and there's a zombie outbreak* depends on what Sara does. Thus, though Sara is responsible for the fact *that her research took place and there's a zombie outbreak*, it doesn't follow that she's at all responsible for the fact *that there's a zombie outbreak*.⁸² And it seems to me that this is what the Zombie Case is supposed to elicit.

Thus, I conclude that TDM is safe from these objections. Moreover, I can't think of any better objections to TDM than what I've here considered. So, I think that TDM is safe from objection, full stop. And if TDM *is* safe, then counterexamples to Rule B are impossible.

IV. Incompatibilism is True, and What This Might Say About Our World

So far I've been at pains to defend the Direct Argument* from the most pressing objections that have been brought to bear against it. I've concluded that the Direct Argument* is safe from these objections. It's impossible to give a counterexample to Rule B, for example; so, Rule B is true (and necessarily so). Rule A is true since no one can be even partly morally responsible for a necessary truth. And there are no possible No Past* Objections that can be brought against the Direct Argument*. So far as I can see, there are no other places in the Direct Argument* on which to press. Thus, the Direct Argument* is successful.

⁸² Objection: Doesn't the truth of *that there is a zombie outbreak* depend on Sara? If so, then, by TDM, isn't Sara at least partly directly morally responsible for the fact *that there is a zombie outbreak*? If your answer is yes to either of these, then the Zombie Case is, contrary to what you say, a counterexample to TDM.

Reply: No, the truth of *that there is a zombie outbreak* does *not* depend on what Sara does. At least, it does not depend on what Sara does *in the relevant way*. Recall that the notion of 'dependence' at issue in TDM is the very trivial sort of dependence in which truth depends on the way the world is. So: is it true *that there is a zombie outbreak* if and only if Sara does her research? According to the details of the case, the answer must be no (e.g., Sara's boss's nefarious plan plays a part, etc.). There are, presumably, very many ways in which it could be true *that there is a zombie outbreak*. Thus, I conclude that the truth of *that there is a zombie outbreak* does *not* depend on Sara—not, anyway, in the sense of 'dependence' at issue in TDM. Thus, the Zombie Case is no counterexample to TDM.

But if the Direct Argument* is successful, this means that incompatibilism about moral responsibility and causal determinism is true. Thus, we should conclude that incompatibilism about moral responsibility and causal determinism is true. In this section, then, I wish to gesture at what I think are some potentially important take-aways from the conclusion that incompatibilism is true. This conclusion is important; for, as I'll now go on to suggest, it's very plausible that it tells us some important things about our world. Moreover, it helps to illuminate—and sharpens the focus of—the areas of the free will/moral responsibility debate that are in need of further exploration. What I say in this section deserves a dissertation all its own, and what I say, herein, isn't meant to exhaust everything that might or could be said about the issues I'll raise. Even so, I'd like to take the time to think about where, and on what, philosophers might focus their future philosophical efforts given the success of the Direct Argument*.

To begin to see where, and on what, philosophers thinking about free will and moral responsibility should focus their efforts, I wish to discuss the two upshots of the successful Direct Argument* that I mention above (§I). First, there's the idea that a successful Direct Argument* plausibly reveals to us that indeterminism is true. Second, there's the idea that free will is compatible with indeterminism.

With respect to the first, this is a striking revelation; for, 'indeterminism' (as well as 'determinism') is, as van Inwagen puts it, "a thesis about the motion of particles of matter in the void" (van Inwagen, 1983, p. 210). That is, indeterminism is a thesis about the physical world, and how it operates. Presumably, theses about the physical world are theses best left investigated by the physical sciences (e.g. biology, physics, chemistry). Yet, I am now claiming that we can know the answer to this scientific question by reflection from the armchair. I claim

that we can know that the universe is indeterministic because we can know that we're morally responsible for (at least some of) our actions.

But such reflection is consistent with what it is that quantum physics seems to suggest about our universe; namely, that at the microscopic level, the movement of particles in the void is indeterministic; how particles behave is a matter of statistical probability as opposed to certainty.⁸³ So, the Direct Argument*'s success is consistent with what empirical scientific evidence purports to show. This is good reason, it seems to me, to conclude that we can know, on the basis of the Direct Argument*'s success coupled with our being morally responsible for (at least some of) our actions, that indeterminism is true.

Perhaps unsurprisingly, some philosophers take a distinctly contradictory approach. Derk Pereboom (2001), and Bruce Waller (1990), for example, think that we don't know whether or not determinism (at the macro-level) is true since our scientific evidence isn't conclusive; rather, what we have good reason to believe is that we lack moral responsibility. That is, they agree with me that incompatibilism is true, but they think our best scientific evidence suggests either that determinism is true, or, even if it isn't, that given our scientific understanding of the universe, there are no plausible accounts of human agency that are compatible with indeterminism and moral responsibility. So, what these "hard incompatibilists"⁸⁴ favor is concluding on the basis of scientific evidence—plus some philosophical argumentation, e.g., the Direct Argument*—a particular metaphysical thesis; namely, that we aren't morally responsible for anything that we do.

⁸³ Most philosophers in the free will/moral responsibility debate realize this. See, e.g., van Inwagen's *An Essay on Free will*, and Pereboom's *Living Without Free Will*.

⁸⁴ This term is from Pereboom (2001, p. xix).

In response, I offer what I take to be van Inwagen's argument from chapter six of his *An Essay on Free Will*. As I understand it, the argument goes something like this:

P1. We're morally responsible for our actions only if determinism is false.

P2. We're morally responsible for our actions.

C1. Thus, determinism is false.

Given the success of the Direct Argument*, all sides to this disagreement should agree that P1 is true; so, the bugbear is at P2. But, to quote van Inwagen, "surely we cannot doubt the reality of moral responsibility?" (van Inwagen, 1983, p. 206). Yet this is exactly what hard incompatibilists do.

I think that P2 is just obviously true. But recently, Stephen Kearns (forthcoming) has argued that philosophers, like me, who think that P2 is just obviously true, should consider the following. The above assumption that P2 is obviously true, relies on the assumption that we know the following conjunction is false:

~DET: Determinism is true and incompatibilism is true. (Kearns, forthcoming, p. 3)⁸⁵

But Kearns claims that we cannot know that ~DET is false because

[i]f we can easily work out the falsity of [~DET] based on our commonsense knowledge, then were we to find out the truth of one of the conjuncts, we could easily work out the falsity of the other conjunct. But this does not seem to be true. Finding out that incompatibilism is true would not make it easy to work out that [determinism] is false. Finding out that [determinism] is true would not make it easy to work out that incompatibilism is false. Thus we do not know that [~DET] is false based on our commonsense knowledge. (Ibid., p. 7, my insertions)

So, Kearns's idea is that, even if I'm right about the success of the Direct Argument*, there's no good way to get from *that* conclusion to the conclusion that determinism is false. And this is

⁸⁵ In the original, Kearns title's ~DET as '(A)', and labels 'determinism' as 'action determinism'. Nothing about what follows depends on my having made these minor changes. Also, since the article is forthcoming, any page number that I give as a reference is given as if the first page of the article is p. 1.

true even if our belief that we are morally responsible for stuff is part of our commonsense knowledge.

We might be tempted to wonder why we can't 'easily work out' that determinism is false even if we find out that incompatibilism is true and couple that with our commonsense knowledge that we are morally responsible for some of our actions. For his part, Kearns thinks we cannot easily work out that determinism is false from these considerations because we need conclusive scientific evidence—not mere reasoning from the armchair—to figure out the answer to questions about how the physical world works. We cannot, as Kearns puts it, “figure out that some of our actions (or their immediate causal antecedents) are undetermined from the armchair, even reasoning from knowledge we currently possess” (Ibid., p. 9).

But this seems wrong to me. I think that, if I know that $(p \ \& \ q)$ entails r , and I know that $(p \ \& \ q)$ is true, then by straight-forward *modus ponens* I know that r is true. So, suppose that I know that incompatibilism is true, and that I know that I'm morally responsible for some of my actions. Suppose, further, that I know that this conjunction of facts entails the falsity of $\sim\text{DET}$ (which it does). If I know all of these things, then I know that $\sim\text{DET}$ is false. This seems to me to be a case of easily working out the falsity of $\sim\text{DET}$.

Clearly, I can know that the truth of incompatibilism coupled with the truth of our being morally responsible for some of our actions entails the falsity of $\sim\text{DET}$. But *can* I know that the antecedent is true? That is, can I know that incompatibilism is true and that we are morally responsible for some of our actions? According to John Maier (forthcoming), I cannot.

Maier begins his argument for the claim that we cannot know that P2 is true (and so, we cannot know the truth of the conjunction of P2 and the claim that incompatibilism is true) by

introducing the notion of ‘effective ignorance’. He defines the term this way (where P1, P2, and C1 are as above):

Effective Ignorance: An agent is effectively ignorant of C1 just in case her argument would be exactly as epistemically effective as it actually is if she had no evidence for C1 other than her evidence for P1 and P2. (Maier, p. 5)⁸⁶

The idea is that the proponent of P1 – C1 who is effectively ignorant of C1 is committed to an argument that doesn’t at all benefit from scientific evidence. For, even if she had evidence besides P1 and P2 (e.g. evidence for quantum-level indeterminacy), her argument would be just as effective as she hadn’t had that sort of evidence. Why think that the proponent of P1 – C1 is effectively ignorant? Because the argument’s origin is prior to the discoveries of twentieth-century physics. Thus, if the proponent of P1 – C1 knows that C1 is true *now*, then the proponents of P1 – C1 that antedate her knew C1 *then*.

Further, Maier claims that “someone who is effectively ignorant of C1 cannot know both P1 and P2” (Ibid.). To see why Maier claims that someone who is effectively ignorant of C1 cannot know both P1 and P2, he makes, and argues for, three claims. They go as follows:

Particularity: S can know P2 only if she can know some instance of P2.

Authority: S can know some instance of P2 only if she can know some instance of P2 on partly non-testimonial grounds.

Humility: If S knows P1, then she cannot know any instance of P2 on non-testimonial grounds. (Ibid.)

These three theses, conjoined, entail that anyone who is effectively ignorant of C1, cannot know both P1 and P2. Thus, we cannot know, on the basis of knowing P1 and P2 that C1 is true.

While there might be available objections to either Particularity or Authority, the place to press, I think, is Humility. Before we get to my objection to Humility, however, let us consider

⁸⁶ As with Kearns’s forthcoming paper, for any page number I cite for Maier’s paper, the page number will be given as if the first page is p. 1.

the way in which Maier attempts to motivate the thesis. First, Maier provides the following epistemic principle that underlies Humility (where ‘an excusing condition’ is any condition K such that, if it obtains, then an agent is not morally responsible for her action (e.g. infancy, insanity, etc.)):

Epistemic Priority of Excuse [EPE]: If K is an excusing condition for some particular action A of some particular agent R, then one cannot know on partly non-testimonial grounds that R is morally responsible for A unless one independently knows that K does not obtain. (Ibid., p. 10)

To see EPE in action, Maier gives us the following scenario:

[Jones is a witness in a court case and] is testifying, of Smith, that he is responsible for what he has done, and is not purporting to base her claim entirely on the testimony of others. There is something legally inapt about an appeal to determinism in such a case, but let us imagine that Jones holds that Smith is responsibly only if it is false that he has a certain brain lesion. It is pointed out, by Smith’s lawyer, that Jones is in no position to know whether Smith has that brain lesion. We can imagine versions of the case where Jones response to this point by acknowledging that, on this point at least, she needs to defer to someone else. But let us say that she does not do this. She rather claims to know that agents who have certain brain lesions are not responsible and also, though she is ignorant of whether Smith has this brain lesion, that Smith is responsible. Here we may reasonably protest that Jones does not know that Smith is responsible after all. (Ibid.)

The idea is that the Jones case gives us reason to think that EPE is true; for, we’re supposed to think that Jones cannot know that Smith is morally responsible for what he has done if Jones doesn’t know whether or not Smith has an excusing condition (i.e. whether or not he has a brain lesion). The upshot is that determinism is such an excusing condition if P1 is true. So, if S knows P1, and EPE is true, then S can’t know that P2 is true unless S independently knows that determinism is false. So, if anyone is effectively ignorant as to whether determinism is false, and EPE is true, then no one can know that both P1 and P2 are true.

My first complaint about Maier’s argument for Humility is that its undergirding epistemic principle, EPE, isn’t very well motivated. Indeed, the Jones case—which is meant to motivate the intuitions for EPE—it seems to me, is setup in such a way as to presuppose the truth of EPE.

For, why should we think that Jones *is* ignorant as to whether or not Smith has the relevant brain lesion. If Jones *knows* that Smith is morally responsible for his actions, and his being morally responsible for his actions entails the absence of the relevant brain lesion (and Jones knows this), then it follows that Jones knows that Smith doesn't have the relevant brain lesion.

Of course we might wonder *if* Jones knows that Smith is morally responsible for what he has done. That is, we might wonder whether or not Jones's belief that Smith is morally responsible is *true*. But, it seems infelicitous to set up the case in such a way that we're meant to presuppose that she *couldn't* know such a thing without first knowing that Smith lacks the relevant brain lesion. We might, just as easily, set up the case in such a way that Jones is *not* ignorant as to whether or not Smith has the relevant brain lesion because she's not ignorant as to whether or not he's morally responsible for what he did.

But suppose that the Jones case isn't unhelpful the way that I claim it is. Even if so, I think that the case fails to show that EPE is true. What it *does* show, if anything, is that it is difficult to know whether or not someone is morally responsible for something without independent knowledge of whether or not there is an excusing condition that obtains. But notice that this says nothing about whether or not we can know that *we* (i.e. *ourselves*) are morally responsible for our actions. And P2 is a claim about whether or not *we* (i.e. *ourselves*) are morally responsible for our actions. I think that I, for example, can know that I am morally responsible for some of my actions, and this is true whether or not I have independent evidence that there are no excusing conditions (e.g. determinism) present. Let me explain.

My belief that I am morally responsible for some of my actions seems true to me in the same way that it seems true to me that the world didn't just pop into existence five minutes ago. It also seems true to me in just the same way that it seems true to me that there exist other people

besides myself. Of course there are arguments entailing the falsity of each of these apparent truths; one argument concludes that we should be skeptics about knowledge; the other concludes that we should be solipsists. But, I think that we should not conclude either of these things; and I do not think this on the basis of any argument. Instead, I believe it's true that the world has been around for longer than five minutes, and that there are other people, in a very basic way. That is, I don't believe these things are true on the basis of any argument or form of reasoning.⁸⁷ The same is true, I think, for my belief that I am morally responsible for some of my actions.⁸⁸

Moreover, I think I *know* that the world didn't just pop into existence five minutes ago. And I think I *know* that there are other people besides myself. Thus, if I'm right that my belief that I'm morally responsible for some of my actions is like these others, then I think I *know* that I'm morally responsible for some of my actions. So, in the parlance of EPE, if I know that I'm morally responsible for some of my actions, then I know this on partly non-testimonial grounds without independently knowing that K does not obtain. Therefore, if I know that I am morally responsible for some of my actions, then EPE is false. And if EPE is false, then Humility is false, and Maier's argument is defeated.

Of course one might object that I *don't* know that I am morally responsible for some of my actions; that is, one might claim that I don't know that P2 is true. But, if my belief that I am morally responsible for some of my actions (i.e. that P2 is true) is basic in the way that my belief in an actual past (and my belief that there are other people) is basic, then any argument against my knowing such a thing must mirror the arguments for skepticism and solipsism. But, I think this means that any opponent of the claim to know that P2 is true ought to be a skeptic and a

⁸⁷ For a thorough discussion on how these sorts of beliefs are 'basic', as I suggest, see Alvin Plantinga's (1969, 1993a, 1993b, and 2000).

⁸⁸ Similar points are made by van Inwagen (1983). See, especially, pp. 204 – 214.

solipsist since the argument against knowing P2 would have to mirror the arguments against knowing that there is an external world (and knowing that there exist other people besides oneself). But I think that we should be neither skeptics nor solipsists; so, I think that we should not say that we cannot know P2 is true.⁸⁹

We might, however, claim that P2 is false. Pereboom, for example, thinks that P2 is false because “there are good empirical reasons to believe that we do not in fact have [the sort of free will required for moral responsibility]” (Pereboom, 2001, p. 132).⁹⁰ So, what do we do? I claim that my belief that P2 is true is a basic sort of belief, one that I do not believe on the basis of evidence or argument; Pereboom claims that we have evidence that P2 is false. This seems like an impasse.

Even so, I think it’s worth pointing out that most people—philosophers and laymen—think that we are morally responsible for (at least some of) our actions. To be sure, there are differing accounts of what it means to be ‘morally responsible’ for something. But the position that we are *not* morally responsible for anything, is a minority position.⁹¹ And I think the reason for this is that such a position cuts severely against the human experience. And our experience

⁸⁹ For an engaging discussion of the similarities between the moral responsibility skeptic arguments (and the free will skeptic arguments) and the external world skeptic arguments, see Jean-Baptiste Guillon (forthcoming).

⁹⁰ Pereboom thinks that *libertarian* free will is what’s required for moral responsibility. Moreover, he favors so-called ‘agent-causal’ views of (libertarian) free will over so-called ‘event-causal’ views. He thinks that event-causal views of (libertarian) free will are undermined by the fact that all events governed by the physical laws are found on a continuum of “alien-deterministic events—events such that there are causal factors beyond our control by virtue of which they are causally determined or truly random events—those that are not produced by anything at all, or partially random events—those for which facts beyond the agent’s control contribute to their production but do not determine them, while there is nothing that supplements the contribution of these factors to produce the events” (2001, p. 127). Agent-causal views place our actions and choices beyond this ‘continuum’; however, Pereboom thinks that “empirical objections provide good (albeit not conclusive) reasons to believe that we are not agent-causes,” and that, therefore, we have “good reason to believe that our actions and choices are not agent-caused events, but are rather either alien-deterministic, or partially random, or truly random events” (Ibid., p. 128).

⁹¹ This point is conceded by Pereboom (2001, p. xviii). This is not to say, however, that there are not a number of outstanding philosophers who hold (or held) this view. I’ve mentioned Pereboom and Waller, but there’s also Galen Strawson, C. D. Broad, Spinoza, and B. F. Skinner to name a few.

makes such a view seem quite implausible. Now, unless we have sufficient reason to doubt our shared human experience, I think we should conclude that such a position *is* implausible. So, even if Pereboom is right that we have good empirical reason to think we lack the sort of free will required for moral responsibility—and so, good empirical reason to think that we lack moral responsibility—we should deny his conclusion. We should, instead, say that we have overriding evidence to the contrary; namely, our human experience as of being morally responsible for our actions.

Moreover, if our belief that we are morally responsible for some of our actions really is a basic belief, then we simply must decide which of the following are the more trustworthy: science, or our basic beliefs. I take it as a given that our cognitive faculties are reliable (i.e. that our cognitive faculties are such that they, when functioning properly, lead to, and are designed to (by God, or evolution, or both), form a preponderance of true beliefs). If our cognitive faculties are such that they reliably form true beliefs, then it seems plausible that our basic beliefs—those beliefs that we’re just hard-wired to believe—reliably track the truth. So, since I take it as a given that our cognitive faculties are reliable, I conclude that our basic beliefs reliably track the truth. So, I conclude that I have good reason to think that my basic belief (if it is one) that I am morally responsible for some of my actions is true.

Science, however, has this funny—but completely appropriate—habit of changing its mind. If I had to guess (which I’ll do now), I’d guess that the vast *majority* of scientific theses have been disconfirmed by evidence. As I understand it, this is one way in which science works: come up with a plausible hypothesis H for some remarkable fact F, then try and gather evidence E that confirms H. If, however, there happens to be some evidence E* that disconfirms H, then H is thrown out. Scientific progress is chock-full of this sort of happening, and so much the

better for science. The point, though, is this: if I had to guess which is the more reliable, I'd guess that my basic belief is more reliable than our scientific theories (e.g. that determinism is true).

If I'm right that P2 is true, then this leads to our second upshot of the Direct Argument*'s success: that free will is compatible with indeterminism. Though I haven't discussed free will at all in this dissertation, the success of the Direct Argument* is relevant to the free will discussion. Here is why. The Direct Argument* establishes that incompatibilism is true. We have good reason to think that we are morally responsible for (at least some of) our actions. Moreover, there's good reason to think that moral responsibility requires free will.⁹² Thus, we have good reason to think that we have free will. Thus, we have good reason to think that free will is compatible with indeterminism.

This conclusion does not establish whether or not free will is compatible with determinism. Nevertheless, the conclusion that free will is compatible with indeterminism is important to the free will debate. Galen Strawson (2003), for example, argues that moral responsibility is impossible. That is, it doesn't matter whether or not a world is deterministic, moral responsibility is incompatible with both, and this I think, is because free will is incompatible with both. Strawson explains his position thusly:

(1) It is undeniable that one is the way one is, initially, as a result of heredity and early experience, and it is undeniable that these are things for which one cannot be held to be in any way responsible (morally or otherwise). (2) One cannot at any later stage of life hope to accede to true moral responsibility for the way one is by trying to change the way one already is as a result of heredity and previous experience. For (3) both the particular way in which one is moved to try to change oneself, and the degree of one's success in one's attempt at change, will be determined by how one already is as a result of heredity

⁹² Notably, Fischer (1995, 1998, 2007, 2010) denies this position on the basis of Frankfurt cases. Frankfurt cases, he thinks, successfully show that freedom to do otherwise (i.e. the sort of 'free will' I mention in the text) is not required for moral responsibility. His 'semi-compatibilist' position is one in which he believes we *are* morally responsible for (at least some of) our actions, but that we needn't have free will (in the sense of 'two-way' or 'regulative' control—i.e. freedom to do otherwise) in order to be so responsible.

and previous experience. And (4) any further changes that one can bring about only after one has brought about certain initial changes will in turn be determined, via initial changes, by heredity and previous experience. (5) This may not be the whole story, for it may be that some changes in the way one is are traceable not to heredity and experience but to the influence of indeterministic or random factors. But it is absurd to suppose that indeterministic or random factors, for which one is *ex hypothesi* in no way responsible, can in themselves contribute in any way to one's being truly morally responsible for how one is. (Strawson, 2003, p. 215)

The idea, I think, is that we do what we do because of the way we are; we (allegedly) have no regress stopping account of how we can be said to have *control* over the way we are; and if our regress stopping account is that some part of the way we are happens because of random factors, then we don't have control over this either. Thus, we're not morally responsible for anything, regardless of whether or not determinism is true.

I just explained Strawson's lengthy quote in terms of 'control'. For I think that this must be what he's attempting to show: if we don't have any control over the way we are, then we're not responsible for the way we are (and, thus, not responsible for what we do). And this is true regardless of whether or not 'the way we are' is determined or happens at random, i.e., is undetermined. So, even though Strawson is speaking about moral responsibility, I think his conclusion is such that it renders 'control'—that is, the sort of free will required for moral responsibility—impossible.

But the Direct Argument*'s conclusion, coupled with our knowledge that P2 is true, shows that Strawson (or anyone else who thinks that free will is not compatible with indeterminism) is wrong. Thus, since we're morally responsible for (at least some of) our actions, and moral responsibility plausibly requires free will, we can conclude that free will is compatible with indeterminism. So, the sort of free will required for moral responsibility is compatible with indeterminism.

Such conclusion, I think, reveals some important ways in which to advance the discussion on free will and moral responsibility. We are, hereby, given two things: moral responsibility is incompatible with determinism, and indeterminism is true. Now, what needs to be answered is this: what account of free will is sufficient to explain the kind of free will required for moral responsibility? Which, in turn, spawns the following questions. Does free will require indeterminism? If not, then is it possible to have free will but not be morally responsible for anything? If free will *does* require indeterminism, then what is it about *the agent* that must be undetermined in order for her will to be free (e.g. is it events in her brain? Is it the agent *herself*? Etc.)? Whatever the answers to these question are, I think attempting to answer *these* questions will help philosophers figure out what free will *is*, and, moreover, it will help them advance the relevant debate.

V. Conclusions

The goal of this dissertation has been to defend the Direct Argument for the incompatibility of causal determinism and moral responsibility. In particular, the goal has been to defend the Argument's key inference rules as well as its metaphysical assumptions, which undergird the Argument. In chapters two, three, and six (the present chapter), I defended Rule B, the most controversial of the Direct Argument's inference rules. In chapter four, I defended Rule A from (what I take to be) the only criticism of it on offer in the literature. In chapter five, I defended one of the Direct Arguments most salient metaphysical assumptions, namely, that for any contingently existing being in a deterministic universe, it will be such that there's something remote to it that determines its steps. That is, for any possible deterministic universe, there will be a remote past*, even if not a remote past.

To defend Rule B from its interlocutors, I devoted chapter two to a discussion of some recent attempts to give a counterexample to the Rule. Fischer and Ravizza's well-known Erosion case, for example, is considered by very many philosophers to have successfully shown that Rule B is invalid. But I argued that Erosion does *not* successfully show that Rule B is invalid; rather, it reveals a confusion in the setup of the case, namely, that Erosion fails to show that Betty is *directly* morally responsible for the fact that the enemy base is crushed by a glacier. It's important that Erosion doesn't successfully show that Rule B is invalid because much of the literature devoted to the Direct Argument's inference rules deals with *revisions* of Rule B, e.g., revisions of Rule B that are devised to capture only 'one path' cases (McKenna), or Rule B as a 'causal rule' of inference rather than a logical one (Shabo). And if Rule B hasn't been undermined, then the Rule B defender needn't take time to consider *revisions* of Rule B at all; for Rule B is valid, so it doesn't really matter whether or not there are counterexamples to certain revisions of the Rule.

But even recent attempts to provide a counterexample to Rule B, itself, fail. For, as we saw with various cases from Haji, Widerker, and Schnall with Widerker, these attempts are insufficient for a variety of reasons. Not least of which, though, was defended in the present chapter. For, as we've now seen, the principle, Truth Dependence_{MORAL}, is likely true, and any plausible theory of moral responsibility will have to say that it is true. And if so, then counterexamples to Rule B are impossible. For, as any purported counterexample will claim, the agent in the case will be responsible for the fact expressed in the conclusion to which the 'NR' operator is affixed, which means that she is thus responsible for the truth of the consequent of the second premise. But the truth of the conditional in the second premise *depends on* the truth of the consequent which. According to TDM, this means that the agent is responsible for the truth

of the conditional in the second premise. Thus, there are no possible substitution instances of Rule B where the agent is responsible for the conclusion, but isn't responsible for (at least) the truth of the conditional in the second premise. Thus, counterexamples that purport to undermine Rule B, itself (like Haji's, Widerkers, etc.), are doomed to fail.

But even if there are no possible counterexamples to Rule B, it doesn't follow that such a Rule is one that we should give much attention and credence to. For, as we saw in chapter three, some philosophers have argued that the Rule is dialectically inappropriate. Fischer, for example, thinks that Rule B leads to a "dialectical stalemate", and McKenna, for his part, thinks that Rule B should have never been accepted in the first place since it nearly begs the question against the compatibilist. But, as I argued, neither of these worries successfully undercuts the Rule. For example, Fischer is right only if Rule B is false, and we are, instead, thinking of some *modified version* of Rule B. Some modified version of Rule B might lead to a dialectical stalemate, this much I'm happy to agree with. But it's not so clear that Rule B *itself* leads to a dialectical stalemate. And given Rule B's validity, Fischer will have to show that Rule B—and not one of its revisions—leads to a dialectical stalemate. Moreover, McKenna is right only if there's some non question-begging reason to think that Rule B can be established only by examples that pass through a 'normally functioning' agent. But, this seems to be the very question at issue; so, there isn't much reason to think that McKenna is right about Rule B. Thus, from my discussions of Rule B in chapters two, three, and six (the present chapter), I concluded that Rule B is valid.

Rule A says that no one is, or could be, even partly directly morally responsible for a necessary truth. Such a rule seems trivially true, and most philosophers agree. Stephen Kearns, however, has tried to give several counterexamples to Rule A (or some variation thereof). In chapter four, I argued that Kearns's attempts to provide a successful counterexample to Rule A

fail. I gave several reasons for thinking that each of these counterexamples fails; but, I think the chief reason for their failing is that each case relies on two invalid rules of inference, what I called

ER: If S is directly morally responsible for p, and $p \supset q$, then S is directly morally responsible for q, and

MC: If S makes it the case that p, and $p \supset q$, then S makes it the case that q.

But ER and MC are clearly invalid. For example, it's true that I make it the case (or am directly morally responsible for the fact) that I write this dissertation; and it's true that my writing dissertation implies that I was born; but of course it doesn't follow that I made it the case (or am directly morally responsible for the fact) that I was born. So, Rule A is safe from objection.

In chapter five I considered Joseph Campbell's 'No Past Objection', and conceded that *this* objection has some real bite to it. For, I think that the objection is, strictly speaking, successful. The way the Direct Argument is worded, the fact that there are possible worlds that are deterministic but lack a remote past undermines at least premises 1, 2, 3, and 4 as currently written. But I argued that this objection misses the point of the Direct Argument's key metaphysical assumption that for any contingently existing being in a deterministic universe, he'll be such that there's something remote to him—that is, beyond his influence—that determines his steps. And the same is true of the natural laws in any deterministic universe; they'll be such that something over which they have no influence causes them to be true. Thus, I argued that there are no possible deterministic universes that lack a remote *past*•. If I'm right about all of this, then we can simply replace the notion of 'remote past' in the Direct Argument with 'remote past*'—a notion that I think properly gets at the metaphysical spirit of the Argument. And such revision gives us the Direct Argument*. This argument, I claimed, is successful. Thus, incompatibilism about moral responsibility and causal determinism is true.

In the present chapter I've done two things. First, I argued that a corollary to the truism about truth—that is, the idea that truth depends on the world in a very trivial way—is the principle, Truth Dependence_{MORAL} (TDM). And since no one should deny the truism about truth, no one should deny TDM. But if TDM is true, then counterexamples to Rule B are impossible; thus, counterexamples to Rule B are impossible. Second, I gave what I think are two important upshots of a successful Direct Argument*. The first of these was that we can know that incompatibilism about moral responsibility and causal determinism is true; and the second is that we have good reason to think that free will—something not discussed in the dissertation—is compatible with determinism's being false. This last upshot might seem fairly modest; however, I think its importance to the free will debate cannot be overlooked. For, since we have good reason to believe that free will is compatible with indeterminism, and free will is required for moral responsibility, we have good reason to think that there's a cogent explanation of indeterministic free will in the offing. What such a thesis will look like I don't know. But realizing that there likely *is* such a thesis helps to advance not just the free will discussion, but also metaphysics more generally, and the broader philosophy of mind.

REFERENCES

- Bailey, Andrew M. 2012, "Incompatibilism and the Past," *Philosophy and Phenomenological Research* 85: 351 – 376.
- Campbell, Joseph Keim 2007, "Free Will and the Necessity of the Past," *Analysis* 67: 105 – 111.
- _____ 2008, "Reply to Brueckner," *Analysis* 68: 264 – 269.
- _____ 2010, "Incompatibilism and Fatalism: Reply to Loss," *Analysis* 70: 71 – 76.
- Conee, Earl and Richard Feldman (2004). *Evidentialism: New Essays in Epistemology* (New York: Oxford University Press).
- Crisp, Thomas and Ted A. Warfield 2000, "The Irrelevance of Indeterministic Counterexamples to Principle Beta," in *Philosophy and Phenomenological Research* 61: 173 – 184.
- Ellis, B. and Lierse, Caroline., 1994, "Dispositional Essentialism", *Australasian Journal of Philosophy*, 72: 27–45.
- Ellis, Brian., 2001, *Scientific Essentialism*, Cambridge: Cambridge University Press.
- Feldman, Fred 1986, *Doing the Best We Can: An Essay in Informal Deontic Logic*, Boston: D. Reidel Publishing Company.
- Finch, Alicia 2013, "On Behalf of the Consequence Argument: Time, Modality, and the Nature of Free Action," *Philosophical Studies* 163: 151 – 170.
- Fischer, John Martin and Mark Ravizza 1998, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.
- Fischer, John Martin and Eleonore Stump 2000, "Transfer Principles and Moral Responsibility," *Philosophical Perspectives* 14: 47 – 55.
- Fischer, John Martin, Robert Kane, Derk Pereboom, and Manuel Vargas, 2007, *Four Views on Free Will* Malden, MA: Blackwell Publishing.
- Fischer, John Martin (1994). *The Metaphysics of Free Will*. Malden, Massachusetts: Blackwell Publishers.
- _____ 2006, *My Way: Essays on Moral Responsibility*, New York: Oxford University Press.
- _____ 2010, "The Frankfurt Cases: The Moral of the Stories." *Philosophical Review* 119: 315 – 336.
- Frankfurt, Harry 1969, "Alternative Possibilities and Moral Responsibility," in *The Journal of Philosophy* 66: 829 – 839.

- Ginet, Carl 1966, "Might we Have no Choice?," in *Freedom and Determinism*, Keith Lehrer ed., New York: Random House Publishing.
- _____. 2007, "An Action can be Both Uncaused and Up to the Agent," in *Intentionality, Deliberation and Autonomy* eds. Christoph Lumer and Sandro Nannini, Burlington, VT: Ashgate Publishing Company: 243 – 255.
- Guillon, Jean-Baptiste (forthcoming), "Van Inwagen on Introspected Freedom," *Philosophical Studies*.
- Haji, Ishtiyaque 2008, "Reflections on the Incompatibilist's Direct Argument," *Erkenntnis* 68: 1 – 19.
- _____. 2009, *Incompatibilism's Allure: Principal Arguments for Incompatibilism*, Toronto: Broadview Press.
- _____. 2010, "On the Direct Argument for the Incompatibility of Determinism and Moral Responsibility," in *Grazer Philosophische Studien* 80: 111 – 130.
- _____. 2012, *Reason's Debt to Freedom*, New York: Oxford University Press.
- Hermes, Charles (forthcoming), "Truthmakers and the Direct Argument," *Philosophical Studies*.
- Kearns, Stephen 2011, "Responsibility for Necessities," *Philosophical Studies* 155: 307 – 324.
- _____. (forthcoming), "Free Will Agnosticism," *Nous* 00: 1 – 18.
- Maier, John (forthcoming), "The Argument From Moral Responsibility," *Australasian Journal of Philosophy*
- McKenna, Michael 2001, "Source Incompatibilism, Ultimacy, and the Transfer of Non-Responsibility," *American Philosophical Quarterly* 38: 37 – 51.
- _____. 2008, "Saying Good-bye to the Direct Argument the Right Way," *Philosophical Review* 117: 349 – 383.
- Merricks, Trenton 2007, *Truth and Ontology*, New York: Oxford University Press.
- _____. 2009, "Truth and Freedom" *Philosophical Review* 118: 29 – 57.
- _____. 2011a, "Replies to Hasker and Zimmerman," in *Molinism: The Contemporary Debate* ed. Ken Perszyk, New York: Oxford University Press, 90 – 95.
- _____. 2011b, "Truth and Molinism," in *Molinism: The Contemporary Debate* ed. Ken Perszyk, New York: Oxford University Press, 50 – 72.

- O'Connor, Timothy 1993, "On the Transfer of Necessity," *Nous* 27: 204 – 218.
- Pereboom, Derk 2001, *Living Without Free Will*, New York: Cambridge University Press.
- Plantinga, Alvin 1969, *God and Other Minds*, Ithaca: Cornell University Press.
- _____ 1974a, *The Nature of Necessity*, Oxford: Oxford University Press.
- _____ 1974b, *God, Freedom, and Evil*, Grand Rapids, MI: Eerdmans Publishing Company.
- _____ 1993a, *Warrant: The Current Debate*, New York: Oxford University Press.
- _____ 1993b, *Warrant and Proper Function*, New York: Oxford University Press.
- _____ 2000, *Warranted Christian Belief*, New York: Oxford University Press.
- Ravizza, Mark 1994, "Semi-Compatibilism and the Transfer of Nonresponsibility," *Philosophical Studies* 75: 61 – 93.
- Shabo, Seth 2010a, "Against Logical Versions of the Direct Argument: A New Counterexample," *American Philosophical Quarterly* 47: 239 – 252.
- _____ 2010b, "The Fate of the Direct Argument and the Case for Incompatibilism," *Philosophical Studies* 150: 405 – 424.
- Schnall, Ira M. and David Widerker 2012, "The Direct Argument and the Burden of Proof," *Analysis* 72: 25 – 36.
- _____ (forthcoming), "The Direct Argument for Incompatibilism," in *Libertarian Free Will* ed. David Palmer, New York: Oxford University Press.
- Strawson, Galen 2003, "The Impossibility of Moral Responsibility," in *Free Will* 2nd Edition ed. Gary Watson, New York: Oxford University Press.
- van Inwagen, Peter 1983, *An Essay on Free Will*, Oxford: Oxford University Press.
- Waller, Bruce 1990, *Freedom Without Responsibility*, Philadelphia: Temple University Press.
- Warfield, Ted A. 1996, "Determinism and Moral Responsibility are Incompatible," *Philosophical Topics* 24: 215 – 226.
- Widerker, David 2002, "Farewell to the Direct Argument," *The Journal of Philosophy* 99: 316 – 324.
- Zimmerman, Michael J. 1996, *The Concept of Moral Obligation*, Cambridge: Cambridge University Press.

_____ 2008, *Living With Uncertainty: The Moral Significance of Uncertainty*, Cambridge: Cambridge University Press.

VITA

P. Roger Turner obtained his B.S. in Government from Liberty University in 2003. After working for Alion Science and Technology, in Washington D.C., he attended Liberty University, again, to earn his M.A. in Religious Studies. He completed the M.A. in 2010, and attended the University of Tennessee to pursue his Ph.D. in philosophy completing the degree in 2014.