

TRUTH AND MORAL RESPONSIBILITY

Forthcoming in *New Advances in Causation, Agency, and Moral Responsibility*, eds. Fabio Bacchini, Massimo Dell’Utri, and Stefano Caputo (Cambridge Scholars Publishing)

P. Roger Turner

Most philosophers who study moral responsibility have done so in isolation of the concept of truth. Here, I show that thinking about the nature of truth has profound consequences for discussions of moral responsibility. In particular, by focusing on a very obvious fact about truth—that truth depends on the world and not the other way around—we can see that widely accepted alleged counterexamples to a key inference rule in an important incompatibilist argument can be shown not only to be unsuccessful, but also *impossible*.

As it is usually understood in philosophical debate, to be morally responsible for something is to be either praiseworthy or blameworthy for it. And this notion of moral responsibility is arguably a key feature of personhood and deeply important to our self-understanding. So, that we *are*, or that we *can be*, either praiseworthy or blameworthy for the things that we do is important to the way in which we view the type of creatures that we are. We are, we think, *moral* creatures, after all.

But now an important question. Is moral responsibility compatible with the truth of causal determinism (where ‘causal determinism’ is “the thesis that there is at any instant exactly one physically possible future,” (van Inwagen, 1983, p.3))? That is, could there be morally responsible agents in a causally determined universe? Or could only *nondeterministic* universes feature morally responsible agency? 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 is morally responsible.

Here are the details of van Inwagen’s 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’ |
| (2) $\Box (P \supset (L \supset C))$ | 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. It is the first challenge against Rule B—i.e. alleged counterexamples to the Rule—that I wish to focus on in this paper. Specifically, I want to argue that if we reflect on the nature of *truth*, we will see that there's good reason to think that counterexamples to Rule B are impossible.

If I am right that the nature of truth shows us that counterexamples to Rule B are impossible, such conclusion is important. For, if Rule B is valid, then incompatibilism is true. Moreover, if Rule B is invalid, then there are possible cases that illustrate its falsity; that is, there

are counterexamples to the Rule. So, 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. That compatibilists must deny an obvious fact about truth is very surprising (and important) indeed!

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.¹

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.

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.

¹ 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 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 the paper 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.

II. Rule B and Some Alleged Counterexamples

Recall that the Direct Argument depends on the following rule of inference:

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

This inference rule is the most controversial aspect of the Direct Argument.² Where ‘NR’ is an operator that means ‘nobody is now, or ever has been, even partly directly morally responsible for the fact that’, what follows is that, if Rule B is valid, then if you’re not (now or ever) even partly directly morally responsible for some fact, p, and you’re not (now or ever) even partly directly morally responsible for the fact that p implies some further fact, q, then you’re not (now or ever) even partly directly morally responsible for the fact that q.

Van Inwagen attempts to motivate Rule B with various cases like the following:

Plato: NR Plato died in antiquity

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

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

And surely the conclusion of Plato—that no one is now (or ever has been, or ever will be) even partly directly morally responsible for the fact that Plato never met Hume—follows given the truth of the first two steps. And since Plato is just a substitution instance of Rule B, Rule B appears to be valid.

But, as I say, Rule B is contentious, and lots of ink has been spilt in discussion of this controversial inference rule. So, now I wish to consider two recent alleged counterexamples to

² Rule B is the most controversial aspect of the Direct Argument, but there are other objections to the argument in the literature. Stephen Kearns (2011) questions the first inference rule upon which the Direct Argument rests; namely Rule A. Rule A says that no one is now (or ever was, or will be) even partly directly morally responsible for a necessary truth. Kearns attempts to call this Rule into question.

Joseph Keim Campbell (2007, 2008, 2010) calls into question the assumption that the ‘determinism’ thesis *necessarily* includes the idea of a remote, or distant, past. The upshot of Campbell’s objection is that, if he’s right, though the Direct Argument still goes through, it fails to establish incompatibilism, since incompatibilism, if true, is necessarily true. But, if there could be deterministic worlds that don’t include a remote, or distant, past, then all the Direct Argument establishes (if it establishes anything) is that moral responsibility and causal determinism are incompatible in worlds that include a remote, or distant, past. But this is consistent with the possibility that there are deterministic worlds that *don’t* include a remote, or distant, past but include morally responsible agents. In my dissertation, I argue that Kearns’s alleged counterexamples to Rule A fail, and that, while Campbell’s objection goes through, it can be met by a revised version of the Direct Argument that continues to rely on both Rule A and Rule B. So, Rule B’s validity continues to be important to a properly revised version of the Direct Argument for incompatibilism about moral responsibility and causal determinism.

Rule B. My aim is to show that counterexamples to Rule B are in fact impossible; so, I use these three recent purported counterexamples in order to illustrate this point. I begin by considering a well-known case called ‘Erosion’, introduced by John Martin Fischer and Mark Ravizza (1998). Many philosophers think that Erosion decisively tells against Rule B; so, if it fails as a counterexample to Rule B, this is important. Next, I’ll consider another alleged counterexample by Ishtiyaque Haji (2010).³ As we’ll see, 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’ll start with Fischer and Ravizza’s attempt to provide a counterexample. 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, 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.)

³ There are, of course, many other attempts to provide a counterexample to Rule B. See, e.g., David Widerker (2002), Michael McKenna (2008), Seth Shabo (2010a; 2010b), Ishtiyaque Haji (2008; 2009).

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, 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 does not follow that

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

6 doesn't follow because, since the outcome (the camp's being destroyed by the glacier) 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 don't think that Erosion successfully shows that Rule B is invalid; and reflection on the truism about truth (§I) 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 see why I say this, consider:

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.

$1^* - 3^*$ is just $1 - 3$ written in the form of Rule B. And the idea is that, even though 1^* and 2^* are true, 3^* is false because Betty is directly morally responsible for the fact that there is an avalanche that crushes the enemy base at T3.

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 (The glacier is eroding) \vee (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.⁴ 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

⁴ Widerker and Schnall (forthcoming) use this exact move against Haji's (2009) attempt to provide a counterexample to Rule B.

morally responsible for the truth of the conditional that lies within the scope of the ‘NR’ operator in 2.⁵

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.

Ishtiyaque Haji (2010) offers the following attempt at a counterexample to Rule B:

Hal’s Creation: 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).

Hal’s Creation generates the following substitution instance of Rule B:

Hal’s Argument:

(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.).

⁵ Suppose that TDM isn’t true. There is still reason to doubt that Erosion provides a counterexample to Rule B. For, there is a problem with 3*. The problem with 3* is that it misses the point of Rule B, that is, the point of *direct* moral responsibility. Following David Widerker (2002), we can think of direct moral responsibility in this way: “S is directly responsible for p just in case S is responsible for p, but *not* in virtue of being responsible for some other fact” (Ibid., 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, Betty is directly morally responsible for the fact *that there is an avalanche that crushes the enemy base at T3* only if she’s not responsible for this fact because of her responsibility for the truth of some *other* fact. And since Betty is, *ex hypothesi*, responsible for the fact *that there is an avalanche that crushes the enemy base at T3* only in virtue of the fact that she *caused* the avalanche, she is not directly responsible for the fact *that there is an avalanche that crushes the enemy base at T3*. Thus, even if TDM isn’t true, Erosion is defeated.

The upshot of Hal’s Argument 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 are 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 the conditional that lies within the scope of the ‘NR’ operator in 5H. But, if so, then Yasmin would have to be morally responsible for

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.).

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 I think that, given TDM, Hal’s Creation does *not* generate a counterexample to Rule B; for, I think that TDM’s truth reveals that 5H is *false*. To see that this is so, let’s—as we did with the Erosion case—rewrite 5H’s conditional 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. So, 5H itself is false, and Hal's Creation is no counterexample to Rule B.

I've shown that Erosion, and Hal's Creation 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.⁶

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 \square : $NRp, \square (p \supset q) \vdash NRq$

The upshot of Rule Beta \square 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 \square that can be constructed, and that don't presuppose causal determinism. Consider the following:

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

⁶ Initially from Mark Ravizza (1994), but also found in Fischer and Ravizza (1998, p. 160ff).

(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)⁷

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

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:

⁷ But originally from Warfield (1996, p. 218).

(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 \Box fails. On either horn, 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): $\Box (r \supset q)$.

(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.

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, consider:

Zombie Case: 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.⁸

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 to 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 is 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 following story. 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, blueprints, 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*.

⁸ I owe this case to Kyle Fritz.

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 thanked me in such a way as to be a type of moral praise given to me for the fact that her home is 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 in this way 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.

But suppose that the homeowner *is* out of line. 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*.⁹ 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. Conclusions

This is a truism about truth: *that dogs bark* is true because dogs bark; *that I write this paper at t* is true because I write this paper at t, and so on. I follow Trenton Merricks (2007, 2009, 2011a, 2011b) when I say that *that truth depends on the world and not the other way*

⁹ 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.

around denotes just this truism, the truism about truth. But this truism, I claimed, has the following corollary:

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

If I'm right about this, then accepting the truism about truth requires accepting TDM; and denying TDM requires rejecting the truism about truth. But no one should reject the truism about truth; so, no one should reject TDM.

But, if TDM is right, then counterexamples to Rule B—the Direct Argument's most contentious transfer principle—are impossible. I argued for this claim by, first, showing that two recent attempts to provide a counterexample (Erosion, and Hal's Creation) fail to so provide given the truth of TDM. Moreover, I argued that TDM's truth shows more than the failure of Erosion and Hal's Creation. It shows that any alleged counterexample to Rule B will fail. For, any purported counterexample to Rule B will have the same form.

The form of any alleged counterexample to Rule B will go as follows: by hypothesis, no one is (or was, or ever will be) even partly directly morally responsible for the propositions governed by the 'NR' operator in the first two premises of the substitution instance of Rule B. But someone *is*, by hypothesis, at least partly directly morally responsible for the proposition governed by the 'NR' operator in the conclusion of the substitution instance of Rule B. But, if TDM is true, then counterexamples of this form (which is all of them!) will fail. For, the truth of the conditional in the second premise depends on what the agent in question does. Thus, in all attempts to give a counterexample to Rule B, someone is at least partly directly morally responsible for the truth of the conditional in the second premise. Thus, if TDM is true, then it's impossible to give a substitution instance of Rule B where the conclusion is false but the

premises are true. Which is just to say that if TDM is true, then it's impossible to give a counterexample to Rule B.¹⁰

References

- Campbell, Joseph Keim 2007, "Free Will and the Necessity of the Past," *Analysis* 67: 105 – 111.
- _____ 2010, "Incompatibilism and Fatalism: Reply to Loss," *Analysis* 70: 71 – 76.
- _____ 2008, "Reply to Brueckner," *Analysis* 68: 264 – 269.
- Fischer, John Martin and Mark Ravizza 1998, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.
- Fischer, John Martin 2006, *My Way: Essays on Moral Responsibility*, New York: Oxford University Press.
- _____ and Eleonore Stump 2000, "Transfer Principles and Moral Responsibility," *Philosophical Perspectives* 14: 47 – 55.
- Haji, Ishtiyaque 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.
- _____ 2008, "Reflections on the Incompatibilist's Direct Argument," *Erkenntnis* 68: 1 – 19.
- Kearns, Stephen 2011, "Responsibility for Necessities," *Philosophical Studies* 155: 307 – 324.
- McKenna, Michael 2008, "Saying Good-bye to the Direct Argument the Right Way," *Philosophical Review* 117: 349 – 383.
- Merricks, Trenton 2011a, "Replies to Hasker and Zimmerman," in *Molinism: The Contemporary Debate* ed. Ken Perszyk, New York: Oxford University Press, 90 – 95.
- _____ 2009, "Truth and Freedom" *Philosophical Review* 118: 29 – 57.

¹⁰ I'd like to thank Joe Campbell, E. J. Coffman, John Martin Fischer, Kyle Fritz, Trenton Merricks, David Palmer, Derk Pereboom, Daniel Speaks, the participants at the Florida State University Graduate Conference on Free Will and Moral Responsibility, and the participants at the 2013 meeting of the Pacific Division of the American Philosophical Association for helpful and instructive comments on earlier versions of this (or parts of this) paper.

- _____ 2011b, "Truth and Molinism," in *Molinism: The Contemporary Debate* ed. Ken Perszyk, New York: Oxford University Press, 50 – 72.
- _____ 2007, *Truth and Ontology*, 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 (forthcoming), "The Direct Argument for Incompatibilism," in *Libertarian Free Will* ed. David Palmer, New York: Oxford University Press.
- van Inwagen, Peter 1983, *An Essay on Free Will*, Oxford: Oxford 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.