

SHABO ON LOGICAL VERSIONS OF THE DIRECT ARGUMENT

Philosophical Studies (forthcoming)

This is the penultimate version of the paper. Please do not cite this draft; rather, see <http://www.springer.com/-/3/AVDkQ4S-2brxj7RSc08Z> for the online first version of the paper.

You can cite *it* using the following DOI: 10.1007/s11098-015-0600-4

I. Introduction

In a recent paper, Seth Shabo (2010) sets out to show that *logical* renderings of the Direct Argument for incompatibilism about moral responsibility and causal determinism, an influential incompatibilist argument for this conclusion, fail. In particular, Shabo argues that the Direct Argument—cashed out in logical terms—fails because it rests on an invalid rule of inference, Rule B. Shabo argues that Rule B, rendered *logically*, is subject to a counterexample that he constructs. If he's right about this, it follows that logical versions of the Direct Argument fail.

But, I think that Shabo is mistaken. So, in what follows, I intend to argue along these lines. In particular, in the next section I will show that Shabo's alleged counterexample to logical renderings of Rule B fails. Since it fails, I will conclude that Shabo has failed successfully to undermine logical versions of the Direct Argument.

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

The Direct Argument goes as follows. Assume the following two principles

Rule A: $\Box p \vdash NRp$, and

Rule B: $NR(p \supset q), NRp \vdash NRq$.¹

¹ 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 to express the operator in this argument as 'NR', as above.

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, what follows is that, if determinism is true, then

- | | |
|---|--------------------------------|
| 1. $\Box (P \ \& \ L \ \supset \ p)$ | By definition of ‘determinism’ |
| 2. $\Box (P \ \supset \ (L \ \supset \ p))$ | 1, exportation |
| 3. NR (P \supset (L \supset p)) | 2, Rule A |
| 4. NR P | Premise |
| 5. NR (L \supset p) | 3, 4, Rule B |
| 6. NR L | Premise |
| 7. NRp | 5, 6, Rule B. |

Let ‘ \Box ’ stand for ‘broadly logical necessity,’ ‘P’ stand for ‘the remote past,’ ‘L’ for ‘the laws of nature,’ ‘p’ for ‘any true proposition whatever,’ and ‘NR’ is the operator 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.

In a recent paper, however, Seth Shabo (2010) 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 logic*, Shabo thinks that he has devised a counterexample that shows noncausal views of Rule B fail. In this section I’ll analyze and, ultimately, undermine Shabo’s alleged counterexample to noncausal renderings 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, 2010, 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:

1*. NR p

2*. NR (p \supset q)

3*. NR q

1* – 3* is supposed to be a substitution instance of Rule B, one that is generated by BA. If this is a successful substitution instance, then it will follow that Rule B is invalid, and BA is a successful counterexample to noncausal, or logical, forms of Rule B. Shabo believes that 1* –

3* 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 of 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 (2014), I argue that it’s plausible to think that 1* is false. For, given that van Inwagen understands the notion of someone’s being ‘partly responsible’ for something to include being responsible for a *logical part* of something,² there’s good reason to think that Jed is partly responsible for the fact under the ‘NR’ operator in 1*. To see why, recall that 1* 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. 105, 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)

² 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 1^* 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 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 r , and so $(L \ \& \ r)$, then the notion of responsibility being invoked “encompasses too much” (Shabo, 2010, 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. 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.

Suppose that p implies q, and call this proposition ‘z’. I think that in cases where p is true, q is obviously a relevant logical consequence of z. Why? Because z’s truth *depends* on q’s being true. Moreover, I think the following principle from is true. Consider:

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. (Turner, 2014, p. 216)

I claim that TDM is a corollary to a “truism about truth”, viz., that “truth depends (in a very trivial way) on the world” (Ibid., p. 215).³ The idea, here, is very simple. For example, it’s true *that dogs bark* because dogs bark; it’s true *that the earth revolves around the sun* because the

³ For more on this (besides what follows, below), see my (2014). My thoughts there, as well as here, are heavily influenced by Trenton Merricks (2007; 2009).

earth revolves around the sun; and so on. To see that TDM is a corollary to this truism about truth, [redacted for blind review]:

[S]uppose 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... 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. (Ibid., p. 215)

The idea, here, is that when a person is morally responsible for the truth of some proposition—*that Jones kills Smith*, say—she's responsible for the truth of the proposition *just because* the truth of the proposition depends on what she does. Conversely, if a person *isn't* morally responsible for the truth of some proposition, it's because she's not responsible for the thing the truth of the proposition depends on (e.g. Jones isn't responsible for the truth of *that Jones kills Smith* because she didn't kill Smith).

If TDM is true, then, while Shabo might have successfully blocked Schnall and Widerker's objection to 1*, he has a problem at 2*. The problem is that, given TDM, 2* is false. I think that 2* 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 $p \supset q$ since he is responsible for the thing upon which the truth of the conditional proposition depends. Thus, I conclude that 2* 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 from the truth of

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

it follows trivially that

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

(Here is why [**] follows trivially from [*]. Since r's truth is stipulated by the BA case, any conditional with r as a consequent is true; thus, it's true that $s \supset r$. And given that $r \supset s$ from [*], this makes s and r material equivalents. So, since $L \ \& \ s \supset q$, it follows that $L \ \& \ r \supset q$. Thus, from [*] it obviously follows that $s \supset r, L \ \& \ r \supset q \vdash L \ \& \ s \supset q$. In other words, [**] follows trivially from [*.])

Now, let 'u' stand for 'L & s' (where 'L & s' is a conjunction of a complete statement of the laws of nature and the proposition 'that Kenny *is* in the boulder's path at t_2 ') and consider:

$$1^{**}. \text{ NR } u$$

$$2^{**}. \text{ NR } (u \supset q)$$

$$3^{**}. \text{ NR } q$$

1** – 3** 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 1** – 3** is a successful counterexample to Rule B. To see why, let's first zoom in on 1**. Is 1** 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 $t2$ '. And it's at least plausibly true that Jed is morally responsible for the fact that Kenny is in the boulder's path at $t2$. That is, it's plausible that Jed is morally responsible for the fact that there's a boulder-and-Kenny-containing path at all. Thus, if Jed *is* morally responsible for this fact, then, given TDM, it follows that Jed is at least partly morally responsible for the conjunction of L and s. That is, it follows that Jed is at least partly morally responsible for u . So, plausibly, 1** is false.

I conclude that 1** is plausibly false because it's plausible that Jed is at least partly 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 2**, 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 $t3$ '. But, given that Jed *intended* for it to be the case that there being a boulder-and-Kenny-containing path at $t2$ materially implies that Kenny is hit by the boulder at $t3$, it's plausible that Jed is at least partly morally responsible for the fact that $u \supset q$. Or, if that isn't convincing, note that the truth of $u \supset q$ depends on q 's being true, something for which Kenny is morally responsible. Thus, by TDM, 2** is false. So, plausibly, both 1** and 2** are false. Thus, 1** – 3** fails to provide a counterexample to Rule B.

If I'm right that 1** – 3** fails to provide a counterexample to Rule B, we have further reason to think that 1* – 3* fails to provide a counterexample to Rule B. To see why, notice that since we get

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

and

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

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 morally responsible for the conjunction of L and s, then it follows that he's also at least partly morally responsible for the conjunction of L and r. That is, if Jed is at least partly morally responsible for u, then he's at least partly morally responsible for p. Thus, if 1** is false, then 1* is false (and similar reasoning applies with respect to 2** and 2*), and Schnall and Widerker were right all along. I think that 1** is false, so I conclude that 1* is false. Thus, Shabo's Bad Angle case fails to provide a successful counterexample to logical renderings of Rule B.

III. Conclusions

I have argued that Shabo's attack against logical renderings of Rule B fails. It fails because Shabo's alleged counterexample to Rule B is merely that: alleged. Since the Bad Angle case is no counterexample to logical renderings of Rule B, I conclude that logical versions of

Rule B—and logical versions of the Direct Argument more generally—remain unscathed from Shabo’s objection.⁴

References

Merricks, Trenton 2007, *Truth and Ontology*, New York: Oxford University Press.

_____ 2009, “Truth and Freedom” *Philosophical Review* 118: 29 – 57.

Shabo, Seth 2010, “Against Logical Versions of the Direct Argument: A New Counterexample,” *American Philosophical Quarterly* 47: 239 – 252.

Schnall, Ira M. and David Widerker 2014, “The Direct Argument for Incompatibilism,” in *Libertarian Free Will* ed. David Palmer, New York: Oxford University Press, 88 – 106.

Turner, P. Roger 2014, “Truth and Moral Responsibility,” in *New Advances in Causation, Agency, and Moral Responsibility* ed. Fabio Bacchini, Stefano Caputo, and Massimo Dell’Utri, New Castle Upon Tyne: Cambridge Scholars Publishing, 214 – 230.

van Inwagen, Peter 1983, *An Essay on Free Will*, Oxford: Oxford University Press.

⁴ I’d like to thank E. J. Coffman, John Martin Fischer, David Palmer, and an anonymous referee for *Philosophical Studies* for helpful comments on earlier drafts of this paper.