

## Rule A

Forthcoming in *Pacific Philosophical Quarterly*; please cite published version

**Abstract:** Rule A: if it's metaphysically necessary that p, we may validly infer that no one is even partly morally responsible for the fact that p. Our principal aim in this paper is to highlight the importance of this rule and to respond to two recent challenges to it. We argue that rule A is more important to contemporary theories of moral responsibility than has previously been recognized. We then consider two recent challenges to the rule and argue that neither challenge successfully undermines the rule's initial appeal.

Let " $\Box p$ " abbreviate "it's metaphysically necessary that p," and let "NRp" abbreviate "p, and no one is even partly morally responsible for the fact that p." Now consider this inference rule:

A.  $\Box p \vdash \text{NRp}$ .

Rule A tells us that if it's metaphysically necessary that p, then we may validly infer that no one is even partly morally responsible for the fact that p.<sup>1</sup> Our principal aim in this paper is to highlight the importance of rule A and to respond to two recent challenges to it.

Here's our plan. We start by explaining the philosophical significance of rule A and the basic idea it captures that, necessarily, no one is deserving of praise or blame for necessary truths. We argue, in particular, that rule A is more important to contemporary theories of moral responsibility than has previously been recognized. We then consider two recent challenges to

---

<sup>1</sup> By "no one" we understand proponents of rule A to mean "no human agent." Some believe that God can be morally responsible for necessary truths. We disagree, but even if we are wrong about that, it won't pose a difficulty for rule A as typically understood by its proponents.

the rule, both of which rely on the idea that our actions are what make certain necessary truths true. We argue that neither challenge successfully undermines the initial appeal of rule A.

### 1. The Significance of Rule A

It's common to distinguish between indirect and direct arguments for the incompatibility of determinism and moral responsibility.<sup>2</sup> Indirect arguments proceed by first trying to establish the intermediate conclusion that determinism precludes the freedom to do otherwise, which is then used to argue for the further conclusion that determinism precludes moral responsibility. Direct arguments, by contrast, try to establish the incompatibility of determinism and moral responsibility directly (i.e., without first trying to show that determinism precludes the freedom to do otherwise), by appealing instead to various transfer of non-responsibility principles.

Rule A plays an important role in direct arguments for the incompatibility of determinism and moral responsibility, as we'll see. However, there is more at stake when it comes to rule A than just the direct arguments. The principle is also crucial to some indirect arguments, and some prominent compatibilist accounts of moral responsibility are committed to its validity as well.

Rule A is most often associated with a direct argument of Peter van Inwagen's (1983: 182-188). Two inference rules are at the heart of that argument. The first is rule A. The second is

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

Rule B is a transfer of non-responsibility principle. It tells us that if no one is even partly morally responsible for the fact that p implies q, and if no one is even partly morally responsible for the fact that p, we may validly infer that no one is even partly morally responsible for the fact that q.

---

<sup>2</sup> Cf. Fischer and Ravizza 1998:17.

Using rules A and B, van Inwagen argues as follows, where “ $P_0$ ” is a proposition describing the intrinsic state of the universe at some time in the distant past prior to the existence of human beings, “ $L$ ” is a conjunction of all the laws of nature, and “ $P$ ” is any true sentence:

- |    |                                    |                                  |
|----|------------------------------------|----------------------------------|
| 0. | Determinism is true                | assumption for conditional proof |
| 1. | $\Box ((P_0 \ \& \ L) \supset P)$  | formal consequence of 0          |
| 2. | $\Box (P_0 \supset (L \supset P))$ | from 1 by exportation            |
| 3. | NR $(P_0 \supset (L \supset P))$   | from 2 by rule A                 |
| 4. | NR $P_0$                           | premise                          |
| 5. | NR $(L \supset P)$                 | from 3 and 4 by rule B           |
| 6. | NR $L$                             | premise                          |
| 7. | NR $P$                             | from 5 and 6 by rule B           |

The conclusion of this argument is that no one is even partly morally responsible for the fact that  $P$ , if determinism is true. But since  $P$  can be any truth you like, this conclusion generalizes; no one is even partly morally responsible for any fact, if determinism is true.

Rule A clearly plays an important role in van Inwagen’s direct argument. Its role in other direct arguments, while less obvious, is no less significant. Consider, for instance, the direct argument advanced by Ted Warfield (1996). This argument replaces rules A and B with

Beta  $\Box$ . NR $p$ ,  $\Box (p \supset q) \vdash$  NR $q$ .

Beta  $\square$  is another transfer of non-responsibility principle. It tells us that if no one is even partly responsible for the fact that p, and if q is a logical consequence of p, we may validly infer that no one is even partly responsible for the fact that q. Using this principle, Warfield argues as follows:

- |                                     |                                  |
|-------------------------------------|----------------------------------|
| 0. Determinism is true              | assumption for conditional proof |
| 1. $\square ((P_o \& L) \supset P)$ | formal consequence of 0          |
| 2. NR ( $P_o \& L$ )                | premise                          |
| 3. NR P                             | from 1 and 2 by Beta $\square$   |

This argument too purports to show that no one is even partly morally responsible for any fact, if determinism is true. Notice, though, that, unlike van Inwagen's direct argument, Warfield's argument doesn't appeal to rule A. One might be led to conclude from this that the argument doesn't depend on that rule. But that would be a mistake. Though Warfield may not appeal to rule A explicitly, we contend that his direct argument is valid only if rule A is valid.

Given Beta  $\square$  and the uncontroversial assumption that there is at least one fact, x, for which no one is even partly morally responsible, rule A is provably valid. Here's the proof:

- |                            |                                |
|----------------------------|--------------------------------|
| 1. $\square p$             | assumption                     |
| 2. $Nx$                    | assumption                     |
| 3. $\square (x \supset p)$ | from 1                         |
| 4. $Np$                    | from 2 and 3 by Beta $\square$ |

Now, if rule A should turn out to be invalid, there would have to be something wrong with this proof of its validity, and obviously the culprit would be Beta  $\square$ . It seems, then, that Beta  $\square$  is valid only if rule A is, which means that direct arguments like Warfield's that employ Beta  $\square$  require the validity of rule A too, even if they don't appeal to that rule explicitly.

Other direct arguments may also depend on rule A in a way, regardless of whether they make explicit use of the principle. Many of these arguments rely, at least implicitly, on the assumption that no one is even partly morally responsible for the laws of nature. (See, e.g., step 6 of van Inwagen's direct argument and step 2 of Warfield's.) But if rule A were invalid, this might undercut at least some of the plausibility of that assumption. For if people could be morally responsible for *metaphysically* necessary truths, it would be difficult to see why they couldn't also be responsible for *physically* necessary truths like the laws of nature.<sup>3</sup>

Rule A is most often associated with direct arguments for the incompatibility of determinism and moral responsibility, but it also has an important role to play in at least some indirect arguments for that conclusion. To see this, consider the following indirect argument:

1. For any true proposition p, no one could have prevented it from being the case that p, if determinism is true.
2. A person is morally responsible for the fact that p only if he could have prevented it from being the case that p.
3. So, for any true proposition p, no one is morally responsible for the fact that p, if determinism is true (from 1 and 2).<sup>4</sup>

---

<sup>3</sup> We owe this point to Kearns 2011: 322.

<sup>4</sup> Cf. van Inwagen 1999: 347.

Rule A is clearly not among the premises of this argument either, nor is it required to validate the inference from the argument's premises to its conclusion. Nevertheless, we contend that the rule plays an indispensable role in this and any other argument that presupposes the likes of 2.

Suppose that someone could be morally responsible for the fact that p, even though it's metaphysically necessary that p. This would invalidate rule A. But it would also show that 2 is false, for no one could have prevented it from being the case that p, if it is metaphysically necessary that p. Any counterexample to rule A would thus be a counterexample to 2 as well. So if rule A is invalid, then 2 is false, in which case any argument that relies on 2 is unsound.

If what we have said thus far is correct, rule A plays a crucial role in a number of prominent arguments for the incompatibility of determinism and moral responsibility. But it isn't just incompatibilists who have a stake in the validity of rule A. At least one prominent compatibilist theory of moral responsibility is also committed to the rule. According to John Fischer and Mark Ravizza's (1998) influential compatibilist theory of responsibility, a person is morally responsible for a fact or state of affairs p only if the person has what Fischer and Ravizza call guidance control over p. For an agent to have guidance control of p, there has to be a causal process leading from the agent's behavior to p that is "sensitive to action." What makes a causal process sensitive to action is a fairly complex matter, but the basic idea, which is all we need for present purposes, is this: the process leading from action to p is suitably sensitive to action only if, holding fixed the non-occurrence of other potential but non-actual triggering events (i.e., events the occurrence of which would initiate a causal sequence leading to p), it wouldn't have been the case that p, had the agent done something other than what he actually did.<sup>5</sup>

---

<sup>5</sup> See Fischer and Ravizza 1998: 111-112. We bypass here both Fischer and Ravizza's formal presentation of their account of guidance control for outcomes, as well as certain complexities having to do with preemptive and simultaneous overdetermination, since these details in no way impact the general points made in the main text.

No one can have guidance control of necessary truths. This is because necessary truths aren't sensitive to action. To see this, suppose that it's metaphysically necessary that  $p$ , and consider any action  $A$  of any agent  $S$ . Holding fixed the non-occurrence of other potential but non-actual triggering events, it would still have been the case that  $p$ , if  $S$  had done something else instead of  $A$ . So whatever process there might have been leading from  $A$  to its being the case that  $p$ , that process wasn't sensitive to the agent's action. But since  $p$  can be any necessary truth and  $A$  any action of any agent  $S$ , this conclusion generalizes. Fischer and Ravizza's theory thus entails that it's impossible for anyone to be responsible for necessary truths, and we suspect that the same will be true for many other compatibilist theories of responsibility, though it's hard to say for sure, since most compatibilist theories only give accounts of responsibility for (contingent) actions and omissions, while remaining silent on responsibility for other facts.

Our focus in this section has been on the significance of rule A. We have defended three main claims. We first argued that rule A is central to many direct arguments for the incompatibility of determinism and moral responsibility, even when it doesn't appear as a premise. We then argued that rule A is essential to indirect arguments that employ the following sort of principle: a person is morally responsible for the fact that  $p$  only if he could have prevented it from being the case that  $p$ . Finally, we argued that prominent compatibilist theories of moral responsibility are committed to the rule. We turn now to the question of its validity.

## **2. Kearns on Rule A**

Rule A is plausible, a fact that hasn't gone unnoticed. The principle seems to van Inwagen "beyond dispute." After all, he says, "No one is responsible for the fact that  $49 \times 18 = 882$ , for the fact that arithmetic is essentially incomplete, or, if Kripke is right about necessary truth, for

the fact that the atomic number of gold is 79” (1983: 184). Similar sentiments are expressed by Warfield, who says that the rule “is (nearly) as trivial and inconsequential as a rule of inference could be” (1996: 218-219). And to Eleonore Stump and John Fischer the principle “seems entirely uncontroversial” (2000: 53). Although we are inclined to agree with these authors about the appeal of rule A, it could be argued that the initial plausibility of the principle can be accounted for by the fact that we aren’t focusing on the right sorts of cases. Notice, in particular, that none of the necessary truths cited by van Inwagen involve or make reference to human agency. Critics of rule A have suggested that once we turn to necessary truths that somehow involve human agency, we find some for which a person could be at least partly responsible.

One such critic is Stephen Kearns (2011), who offers a number of alleged counterexamples to rule A, three of which we’ll consider here. Although these three examples are quite different from one another in many respects, they have the following feature in common: in each of the cases, the agent is said to “make it the case” that p, where p is the metaphysically necessary truth for which the agent is allegedly morally responsible.<sup>6</sup>

### *Murder!*

Suppose that Stephen murders someone and that he is morally responsible for the fact that he murders someone. “This being so,” Kearns says, “it is also clear that [Stephen] is responsible for the fact that he *actually* murders someone.” But it is a necessary truth that Stephen actually murders someone, for, as Kearns goes on to point out, “it is true in every possible world that, in

---

<sup>6</sup> Kearns (2011: 321-322) also suggests cases in which the agent doesn’t make it the case that p, where p is the necessary truth for which the agent is allegedly responsible, but in which the agent is said to be at least partly morally responsible for the fact that p nonetheless. We won’t consider those cases here, in part because they seem to rely for their probative force on the assumption that the examples we will consider are counterexamples to rule A.

the actual world, Stephen murders someone” (2011: 309). Hence, Kearns concludes that Stephen is morally responsible for a metaphysically necessary truth and thus that rule A is invalid.

Let “ $\alpha$ ” rigidly designate the actual world in which Stephen murders someone. It is a necessary truth that Stephen murders someone in  $\alpha$ , for, as Kearns points out, it is true in every possible world that Stephen murders someone in  $\alpha$ . But is Stephen even partly morally responsible for that truth? Kearns claims that he is. But why think that? Kearns suggests that it is because there is some sense in which Stephen makes it the case that he murders someone in  $\alpha$ ;  $\alpha$  “is the way it is,” Kearns says, “in part because of what Stephen does” (2011: 309-310).

We deny that Stephen makes it the case that he murders someone in  $\alpha$ , for the following reason. Possible worlds are maximally consistent states of affairs.  $\alpha$  is thus a maximally consistent state of affairs a defining feature of which is that it has the state of affairs “Stephen murders someone” as a component. We contend that what makes world-indexed truths of the form “X happens in world W” true is the fact that the state of affairs “X happens” is compossible with the conjunction of all the other states of affairs that make up W. What makes “Stephen murders someone in  $\alpha$ ” true on this account is the fact that the state of affairs “Stephen murders someone” is compossible with the conjunction of the other states of affairs of which the maximally consistent state of affairs  $\alpha$  is composed. Notice, though, that Stephen doesn’t make it the case that “Stephen murders someone” is compossible with the other states of affairs of which  $\alpha$  is composed. Which states of affairs are consistent with which is not determined, even in part, by human behavior. Hence, Stephen doesn’t make it the case that he murders someone in  $\alpha$ .

While we don’t find it plausible to think that Stephen is even partly morally responsible for the fact that he murders someone in  $\alpha$ , there is a closely related contingent fact for which he arguably is morally responsible. Stephen had no control over the world-indexed fact that he

murders someone in  $\alpha$ . However, he did have at least some control over whether  $\alpha$  is the actual world, for he presumably had some control over whether he murders someone, and if he hadn't murdered anyone, a possible world other than  $\alpha$ , one in which he isn't a murderer, would have been actual. This observation suggests that while Stephen isn't even partly morally responsible for the necessary truth that he murders someone in  $\alpha$ , he may be least partly morally responsible for the contingent fact that  $\alpha$ , a world in which he murders someone, is the actual world.

### *Hey Jude*

According to Kearns, "It is necessarily true that Hey Jude starts in the distinctive way it does." He goes on to argue that we cannot infer from this, as rule A says we can, that no one is even partly morally responsible for the fact that Hey Jude starts the way it does, the reason being that Paul McCartney is at least partly responsible for that fact. After all, McCartney composed the song. Surely, then, he bears some responsibility for its features, including the fact that it starts in the distinctive way it does. Hence, Kearns concludes that rule A is invalid (2011: 312).

The first thing to say about this example is that it isn't at all obvious that "Hey Jude starts in the distinctive way it does" is a metaphysically necessary truth. As Roger Turner (2015: 211-212) points out, it seems possible for Hey Jude to be played in a different key or for McCartney or some other musician to adlib the beginning during a concert. Also, consider the fact that another musician might decide to 'cover' Hey Jude. When a musician covers a song, she often changes the music in various ways. However, when she does so, we typically don't regard her as playing a different song. It thus seems that there are many things about Hey Jude that could be changed, including its distinctive beginning, without it ceasing to be the particular song it is.

Suppose, though, that this seeming is illusory. Suppose that it is a necessary truth that Hey Jude begins in the distinctive way it does. This suggests that Hey Jude necessarily has a certain structure, a metaphysical essence that might or might not be instantiated. Indeed, Kearns seems to acknowledge as much when he refers to “abstract objects like Hey Jude” (2011: 313). But if that’s right, if Hey Jude is indeed an abstract object, a melodic essence as it were, we think it’s more plausible to understand McCartney’s writing the song not as an instance of *creating* it, for it exists necessarily, but rather as an act of *actualizing* or *instantiating* it. And, of course, it isn’t a necessary truth that McCartney or anyone else actualizes (or instantiates) that melody. This, in turn, suggests that McCartney isn’t even partly morally responsible for the necessary truth that Hey Jude starts in the distinctive way it does, since he didn’t, strictly speaking, create the song, but that he is at least partly morally responsible for the contingent fact that that abstract melody, with its distinctive beginning, has been instantiated in the actual world.

### *Torturing Babies*

“Karen likes to make certain actions wrong to perform. For instance, she decides to make eating cakes wrong by creating a machine that kills thousands of people if anyone eats a cake. She then informs everyone that she has done this” (Kearns 2011: 313-314). As Kearns points out, given the details of this story, “It is now wrong to eat a cake, as it would lead to the death of many people” (2011: 314). Moreover, because Karen made it the case that it is now wrong to eat cake, it’s also plausible to suppose that she is morally responsible for the fact that eating cakes is now wrong. Kearns contends that the following, parallel story yields a counterexample to rule A.

“Karen decides to build a similar machine that kills thousands of people if anyone tortures a baby for their own enjoyment. She then informs everyone that she has done this.”

According to Kearns, just as Karen is morally responsible for the fact that it is now wrong to eat cake, so too she is at least partly morally responsible for the fact that it's wrong to torture babies for fun, "as she has ensured that doing so will lead to the death of thousands of people." However, as Kearns goes on to argue, "it is necessarily true that torturing a baby for one's own enjoyment is wrong" (2011: 314). Hence, he concludes yet again that rule A is invalid.

It is necessarily the case that torturing babies for fun is immoral. However, we deny that Karen is even partly morally responsible for the fact that torturing babies for fun is immoral. What she is morally responsible for is the fact that torturing babies for fun is even more immoral than it otherwise would have been, or perhaps for the fact that we now have additional reasons not to torture babies for fun, reasons we wouldn't have had *sans* Karen's deadly machine. But these are contingent facts. Karen's responsibility for them poses no problem for rule A.

Kearns's defense of the claim that Karen is partly morally responsible for the fact that torturing babies for one's own enjoyment is immoral rests on alleged parallels between the first case in which Karen makes it wrong to eat cake and the second in which she makes torturing babies even more immoral than it otherwise would be. However, there are important disanalogies between the two cases. While in the first story Karen did make it the case that eating cake is wrong, she didn't make it the case in the second story that torturing babies for fun is wrong, for torturing babies was already wrong independently of anything Karen did or refrained from doing.

There is typically nothing untoward about eating cake. So, prior to Karen making her first deadly machine, it presumably wasn't wrong to eat cake. But by making the first deadly machine, Karen changed all this; she made it the case that eating cake is now wrong, and she is arguably morally responsible for the fact that eating cake is now wrong. But the second case is importantly different. In that story, she didn't make it the case that torturing babies is immoral,

for torturing babies for fun was already immoral and would have been immoral no matter what Karen had done or failed to do. It's plausible that what Karen did do, though, is to make it even more immoral to torture babies for fun than it otherwise would have been, by creating additional moral reasons, reasons that don't necessarily apply, not to torture babies for fun. And, it is plausible that she is morally responsible for the fact that we now have these additional reasons not to torture babies for fun. But the fact that we now have these additional reasons isn't a necessary truth, and thus Karen's being morally responsible for it is consistent with rule A.

Kearns's discussion of rule A is ingenious and subtle, and we haven't attempted a point-by-point rejoinder to it here. We have, rather, provided a template for responding to his examples. In each case there are reasons to doubt that the agent made it the case that p, where p is the relevant necessary truth. Moreover, whatever residual appearance there might be that the agents is morally responsible for the relevant necessary truth can be dispelled by noting that, in each case, there is a closely related contingent fact for which the agent arguably is morally responsible. We believe that this template can be applied to Kearns's other examples as well. We conclude that Kearns's challenge, while ingenious, doesn't undermine the otherwise plausible claim that, necessarily, no one is even partly morally responsible for metaphysically necessary truths.

### **3. Hermes on Rule A**

Another recent critic of rule A is Charles Hermes's (2014b). Suppose that Mary murders Matt and that she is morally responsible for doing so. Now consider the following disjunction:

P. Either Mary murders Matt, or she doesn't.

P has only one true disjunct (viz., that Mary murders Matt), and we are assuming that Mary is morally responsible for the truth of that disjunct. According to Hermes, “When a disjunction has only one true disjunct, then whatever is responsible for the disjunct is also responsible for the disjunction” (Hermes 2014b: 388). If he is right about that, it follows that Mary is responsible for the fact that P. But P is a necessary truth, for it’s true in every possible world that either Mary murders Matt, or she doesn’t. So Hermes concludes that rule A is invalid.

A key step in this argument is the following disjunction principle for responsibility (DPR): if a disjunction has only one true disjunct, then whoever is responsible for the disjunct is also responsible for the disjunction. Hermes insists that, “Any plausible account of moral responsibility must make [this principle] true” (2014a: 411). We will examine his reasons for thinking this in due course. But first we explain why we think the principle is false.

Sometimes an agent bears no moral responsibility for the fact that one or the other of two bad outcomes must obtain, as this fact may be entirely beyond the agent’s control. Nevertheless, the agent may have some control over, and may be at least partly morally responsible for, which of those two outcomes obtain. Scenarios of this sort are counterexamples to DPR.

Consider, for instance, the following case. Lindsay launches a missile in the general direction of two cities, A and B, thereby ensuring that one of them will be hit. Which of the two cities will be hit is to be determined by an indeterministic process in the missile’s guidance system. Lindsay’s friend, Jacob, bears no responsibility for the launching of the missile or for its general trajectory, and neither he nor anyone else has any way of preventing the missile from hitting one or the other of the two cities. It is, however, within his power to override the indeterministic process in the missile’s guidance system and to determine which of the two cities

the missile hits. In an effort to make sure that Lindsay's efforts result in maximum civilian casualties, he diverts the missile to city A, which he knows has a larger population than city B.

Now consider the following disjunction:

Q: Either the missile hits city A, or it hits city B.

Q has only one true disjunct (the missile hits city A), and Jacob is partly to blame for the truth of that disjunct. That the missile hits city A is partly his fault. Contrary to what DPR implies, however, Jacob isn't even partly to blame for the fact that Q. Given that he isn't to blame for the missile being launched or for its general trajectory, it isn't his fault that the missile hits one or the other of the two cities. If anyone is to blame for that fact, it's Lindsay. So DPR is false.<sup>7</sup>

It might be objected that, appearances to the contrary, Jacob is at least partly to blame for the fact that the missile hits one or the other of the two cities. After all, he is partly to blame for the fact that the missile hits city A, and the fact that the missile hits city A entails Q. Isn't that enough to make him partly to blame for the fact that the missile hits either city A or city B?

No. To see why, suppose that Kant is responsible for being a great Prussian philosopher. Kant's being a great Prussian philosopher entails that there is such a place as Prussia. It also entails that  $5 + 7 = 12$ . However, Kant isn't even partly responsible for the fact that there is such a place as Prussia, nor is he even partly responsible for the fact that  $5 + 7 = 12$ . Just because p entails q, and S is responsible for p, we can't validly infer that S is responsible for q.

Consider also Bernard Williams's well-known case of Jim and the natives.<sup>8</sup> Recall that, through no fault of his own, Jim finds himself faced with a terrible dilemma: twenty innocent

---

<sup>7</sup> Robinson (n.d.) has independently advanced a similar counterexample to DPR.

<sup>8</sup> Williams 1973: 98-99.

natives have been taken prisoner and will be executed unless Jim kills one of the twenty. Suppose Jim kills one of the twenty. Now consider this disjunction: either Jim kills one of the prisoners, or all twenty are executed. This disjunction has only one true disjunct, and Jim is at least partly responsible for the truth of that disjunct, as he is partly responsible for the fact that he kills one of the twenty prisoners. Contrary to what DPR implies, however, Jim isn't even partly responsible for the truth of the disjunction. It isn't his fault, nor is it to his credit, that either he kills one of the prisoners or all twenty are executed. If anyone is responsible for that disjunction, it's Pedro, the army captain responsible for forcing Jim to make such a terrible choice.

Given the implausible implications of DPR, why might someone think the principle is true? Hermes offers two defenses of it, the first of which is contained in the following passage:

To see why [DPR] is required, suppose 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. (2014a: 411)

The argument here seems to be this. Given that only one of Mary's two children broke the vase, she can validly infer from the fact that Johnny broke the vase and is morally responsible for doing so that he is also morally responsible for the fact that one of her children broke the vase. But unless DPR is true, it is difficult to see how her inference is valid. So we should accept DPR.

This defense of DPR relies on the claim that Mary can *validly* infer from the fact that Johnny broke the vase and is responsible for doing so that he is also responsible for the fact that one of her children broke the vase. But why think that unless you are *already* committed to DPR? It won't help to point out that the fact that Johnny broke the vase, in conjunction with the background assumption that he is one of Mary's only two children, entails that one of her two children broke the vase, for, as we noted a moment ago, we can't validly infer from the fact that S is morally responsible for p and p entails q that S is also morally responsible for q. So unless you are already presupposing the truth of DPR, it isn't obvious that the inference from "Johnny is the only one of my two children responsible for breaking the vase" to "Johnny is responsible for the fact that one of my two children broke the vase" is necessarily valid.

Hermes second defense of DPR is more challenging. He writes:

The exact same features of the world that make it true that "Mary murdered Matt" also make it true that "Mary murdered Matt, or she did not." Suppose that an account of moral responsibility entailed that Mary is responsible for the disjunct but not responsible for the disjunction. Since Mary is responsible for the disjunct, she is responsible for those features of the world that ground the disjunct. Since Mary is not responsible for the disjunction, she is not responsible for those features of the world that ground the disjunction. The features of the world that ground the disjunct, however, just are the features of the world that ground the disjunction. Disjunctions are true because at least one of their disjuncts is true. So, no adequate account of moral responsibility can entail that Mary is responsible for the disjunct but not the disjunction. Otherwise, the account

would require that Mary is and is not responsible for the state of affairs that makes both propositions true. (2014b: 388)

The idea here is that rejecting DPR by assuming that an agent can be morally responsible for the only true disjunct of some disjunction without being morally responsible for the disjunction itself leads to the contradiction that the agent both is and isn't morally responsible for a certain event or state of affairs. So unless we are prepared to endorse contradictions, we ought to accept DPR.

To show that, on pain of contradiction, we must accept DPR, Hermes needs two auxiliary premises. One he states explicitly. The other is implied. The explicit premise is that the same feature of the world that makes it true that "Mary murdered Matt" (viz., Mary's act of murdering Matt) also makes it true that "either Mary murdered Matt, or she didn't." This is so, Hermes suggests, because disjunctions are true in virtue of the fact that at least one of their disjuncts is true. So, since P has only one true disjunct, he concludes that whatever makes that disjunct true must also make P true. The implied premise is that an agent is morally responsible for the fact that p if and only if he is morally responsible for the features of the world that make it the case that p. It is clear that Hermes needs this second premise, for without it he couldn't validly infer steps 5 and 7 in the *reductio* below, both of which are crucial to his argument. It is also suggested by his claim that "Since Mary is responsible for the disjunct, she is responsible for those features of the world that ground the disjunct. Since Mary is not responsible for the disjunction, she is not responsible for those features of the world that ground the disjunction."

With these two auxiliary premises in hand, Hermes's second defense of DPR can be reconstructed a bit more precisely as follows:

1. Suppose, for *reductio*, that DPR is false and that Mary is responsible for the disjunct “Mary murdered Matt” but not the disjunction “Mary murdered Matt, or she didn’t.”
2. The same feature of the world that makes it true that “Mary murdered Matt,” viz., her act of murdering Matt, also makes it true that “Mary murdered Matt, or she didn’t.”
3. S is responsible for the fact that p if and only if S is responsible for the feature of the world that makes it true that p.
4. Mary is responsible for the fact that “Mary murdered Matt” (from 1).
5. Mary is responsible for the feature of the world that makes it true that “Mary murdered Matt,” viz., her act of murdering Matt (from 2, 3 and 4).
6. Mary isn’t responsible for the fact that “Mary murdered Matt, or she didn’t” (from 1).
7. Mary isn’t responsible for the feature of the world that makes it true that Mary murdered Matt or she didn’t, viz., her act of murdering Matt (from 2, 3 and 6).
8. Mary is and isn’t responsible for murdering Matt (from 5 and 7).

To avoid contradictions like 8, Hermes concludes that we should reject our initial assumption and accept DPR. And, as we have seen, if we accept DPR, it is a fairly short step from there to the conclusion that agents could be morally responsible for necessary truths like P.

Another way of escaping contradiction, though, is to keep 1 and reject one of the two auxiliary premises instead. This escape route is more promising than the one Hermes advocates, for several reasons. For starters, 2 isn’t obviously true, but if it is true, then 3 is false.

Consider 2, first. This premise seems to imply that Mary’s act of murdering Matt stands in the *makes true* relation to the disjunct that “Mary murdered Matt,” as well as to the disjunction, P, that “either Mary murdered Matt, or she didn’t.” This isn’t an innocuous

implication. It is controversial whether every truth must have a truthmaker, some entity in the world that makes it true. And while some truths clearly do have truthmakers, it isn't obvious that P is one of them.<sup>9</sup> Some philosophers insist that only contingent truths have truthmakers. If they are right, then P, a necessary truth, doesn't have a truthmaker. And, if P lacks a truthmaker, then Mary's act of murdering Matt isn't what makes P true. But her act of murdering Matt is what makes "Mary murders Matt" true. So, if necessary truths lack truthmakers, 2 is false; the feature of the world that makes it true that "Mary murders Matt" isn't what makes P true.

Perhaps, though, we are reading too much into the locution "makes it true that" in 2. Even if there are truths that lack truthmakers, there is still a trivial, and thus uncontroversial, sense in which truth depends on features of the world. "Dog's bark" is true because dogs bark. "Gretzky exists" is true because Gretzky exists. "There aren't any ghosts" is true because ghosts don't exist. And so on. This is consistent with denying that every truth has a truthmaker, an entity in the world that makes the truth true. So perhaps the phrase "makes it true that" in 2 isn't to be understood as expressing a substantive truthmaker relation. Perhaps it is to be understood instead as expressing the trivial claim that a truth p depends for its truth on the way the world is.

But when "makes it true that" is understood in this way, 2 is false. If dogs didn't bark, it wouldn't be true that "dogs bark." If Gretzky didn't exist, it wouldn't be true that "Gretzky exists." If ghosts existed, it wouldn't be true that "there aren't any ghosts." If Mary hadn't murdered Matt, it wouldn't be true that "Mary murdered Matt." Evidently, though, a similar relation of dependence doesn't hold between Mary's act of murdering Matt and the disjunction, P, that either Mary murders Matt or she doesn't, for if Mary hadn't murdered Matt, it would still have been true that P. The truth of P thus doesn't depend (in the trivial way that truth depends on the world) on whether Mary murdered Matt. The upshot of all this is that if we treat the phrase

---

<sup>9</sup> For some objections to the idea that all truths have truthmakers, see Merricks 2007 and Griffith (forthcoming),

“makes it true that” in 2 as expressing the uncontroversial claim that truth (trivially) depends on the world, then 2 turns out to be false, for while the truth of “Mary murdered Matt” trivially depends on whether Mary murdered Matt, the truth of P doesn’t. P, a necessary truth, would have been true no matter what Mary had done and, indeed, even if Mary had never existed.

But let’s bypass all this complicated business about truth and truthmakers, and let’s suppose that 2 is true for roughly the reasons Hermes cites. Given this assumption, we think a strong case can be made against 3, which, you will recall, says that S is responsible for the fact that p if and only if S is responsible for the feature of the world that makes it true that p.

Consider disjunction Q again:

Q. Either the missile hits city A, or the missile hits city B.

What makes Q true? Q has only one true disjunct (a missile hits city A), and we are now assuming that the feature of the world that makes that disjunct true must also be what makes the disjunction true. The missile’s hitting city A is the feature of the world that makes the disjunct true. So, given our assumption, that must also be the feature of the world that makes Q true. Recall, finally, that Jacob is at least partly to blame for the fact that the missile hits city A, even though he isn’t even partly to blame for the disjunctive fact, Q, that the missile either hits city A or city B. Together these points provide the materials for a counterexample to 3.

Jacob is partly to blame for the missile’s hitting city A. The missile’s hitting city A is, we are supposing, the feature of the world that makes Q true. So, Jacob is partly to blame for the feature of the world that makes Q true. So, 3 entails that Jacob is at least partly to blame for the fact that Q. But, as we have seen, Jacob isn’t even partly to blame for Q. It isn’t his fault that the

missile hits one or the other of the two cities. Hence, 3 is false; an agent can be partly responsible for the feature of the world that makes p true without thereby being responsible for p.

Hermes's objection to rule A hinges on the truth of DPR. We have argued that both of his defenses of that principle are unsuccessful and that there are, moreover, good reasons to think that the principle is false. We conclude that Hermes's objection to rule A is unsuccessful.

#### 4. Conclusion

Rule A is intuitively plausible and is a corollary of various accounts of the control required for moral responsibility, including at least one prominent compatibilist account. Many philosophers—compatibilist and incompatibilist alike—thus have reason to accept rule A. Moreover, we have yet to see a compelling reason to doubt the rule. So, while we would be hesitant to say that rule A is unassailable or beyond dispute, the case for it seems strong indeed.<sup>10</sup>

#### References

- Fischer, J. M. and Ravizza, M. 1998. *Responsibility and Control: A Theory of Moral Responsibility*.
- Griffith, A. n.d. "Towards a Pluralist Theory of Truthmaking," forthcoming in *Erkenntnis*.
- Hermes, C. 2014a. "Truthmakers and the Direct Argument," *Philosophical Studies* 167: 401-418.
- Hermes, C. 2014b. "A Counterexample to A," *Philosophia* 42: 387-389.
- Kearns, S. 2011. "Responsibility for Necessities," *Philosophical Studies* 155: 307-324.
- Merricks, T. 2007. *Truth and Ontology*. New York: Oxford University Press.
- Robinson, M. n.d. "Truthmakers, Moral Responsibility, and an Alleged Counterexample to Rule A," forthcoming in *Erkenntnis*

---

<sup>10</sup> Acknowledgements omitted for blind review.

- Stump, E. and Fischer, J.M. 2000. "Transfer Principles and Moral Responsibility," *Philosophical Perspectives* 14: 47-55.
- Turner, P. R. 2015. "Kearns on Rule A," *Philosophia* 43: 205-215.
- van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- Van Inwagen, P. 1999. "Moral Responsibility, Determinism, and the Ability to do Otherwise," *Journal of Ethics* 3: 341-350.
- Warfield, T. A. 1996. "Determinism and Moral Responsibility Are Incompatible," *Philosophical Topics* 24: 215-226.
- Williams, B. 1973. "A Critique of Utilitarianism," in *Utilitarianism: For and Against*, edited by J.J.C. Smart and B. Williams. Cambridge: Cambridge University Press.