CrossMark

Incompatibilism and the transfer of non-responsibility

Justin A. Capes[1]

Abstract Arguments for the incompatibility of determinism and moral responsibility sometimes make use of various transfer of non-responsibility principles. These principles purport to specify conditions in which lack of moral responsibility is transmitted to the consequences of things for which people are not morally responsible. In this paper, after developing what I take to be the most serious objections to extant principles of this sort, I identify and defend a new transfer of non-responsibility principle that is immune to these and other objections. This new principle says, roughly, that if you are not morally responsible for any of the circumstances that led to a particular outcome, and if you are not morally responsible for those circumstances leading to that outcome, then you are not morally responsible for the outcome either. After defending this principle against a number of objections, I use it to argue for the conclusion that no one is even partly morally responsible for anything, if determinism is true.

Keywords Moral responsibility · Determinism · Incompatibilism · Transfer of non-responsibility · The direct argument

1 The direct argument

Arguments for the incompatibility of determinism and moral responsibility sometimes employ transfer of non-responsibility principles. These principles purport to identify conditions in which lack of moral responsibility is transmitted to the consequences of things for which people are not morally responsible. In this

✉ Justin A. Capes
   capesj@etsu.edu

[1]   Philosophy and Humanities, East Tennessee State University, Johnson City, TN 37614, USA

article, after developing what I take to be some of the most serious objections to extant principles of this sort, I identify a new transfer of non-responsibility principle that is immune to these and other objections. I then use this principle to argue for the conclusion that no one is even partly morally responsible for anything, if determinism is true.

I begin with an argument of Peter van Inwagen's known as the direct argument.[1] The argument gets its name from the fact that it is an attempt to establish the incompatibility of determinism and moral responsibility "directly", i.e., without first establishing the intermediary conclusion that determinism is incompatible with the freedom to do otherwise. Attention to this argument and to the transfer principle on which it turns will help set the stage for the subsequent discussion.

The direct argument and its progeny are formal elaborations on the following basic line of reasoning. If determinism is true, then everything that happens now, including our present behavior, is ultimately a consequence of the laws of nature and events in the distant past prior to our existence. But we are not morally responsible for what happened before we were born, and neither are we morally responsible for the laws of nature. Therefore, the consequences of these things, including our present behavior, are not things for which we are morally responsible.[2]

In an effort to regiment this basic argument, Van Inwagen uses the following abbreviations: "$\Box$" stands for broad logical necessity, "$\supset$" stands for the material conditional, "P_o" is a proposition describing the intrinsic state of the universe long ago before human beings existed, "L" is the conjunction of the laws of nature, "P" is any true sentence, and "NRp" abbreviates "p, and no one is even partly morally responsible for the fact that p."[3] Van Inwagen also employs two inference rules that, according to him, govern the use of the NR operator:

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

Rule A tells us that if it is broadly logically necessary that p, then we may infer that no one is even partly morally responsible for the fact that p. Rule B is a transfer of non-responsibility principle. It says, in effect, that, necessarily, if no one is even partly morally responsible for the fact that p, and if no one is even partly morally responsible for the fact that p implies q, then lack of moral responsibility for these two facts is transmitted to q.

---

[1] See van Inwagen (1980, 1983, pp. 182–188).

[2] See van Inwagen (1983, p. 16) for a structurally similar argument concerning free will.

[3] In his presentation of the direct argument, van Inwagen uses Np instead of NRp, where "Np" abbreviates "p, and no one is, or ever has been, even partly responsible for the fact that p." It has become common practice in discussions of the direct argument to use NRp instead of Np because elsewhere in van Inwagen's work Np has a different reading as part of an argument for the incompatibility of determinism and the freedom to do otherwise. As far as I can tell, nothing of substance turns on these differences in presentation.

With this formal machinery in place, van Inwagen argues as follows:

0. Determinism is true	assumption for conditional proof
1. $\Box ((P_o \ \& \ L) \supset P)$	formal consequence of 0
2. $\Box (P_o \supset (L \supset P))$	from 1
3. NR $(P_o \supset (L \supset P))$	from 2 by rule A
4. NR P_o	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

If sound, this argument shows that no one is even partly morally responsible for any fact, if determinism is true.

Most philosophers who have discussed the direct argument think that rule B is the argument's weakest link. Interestingly, van Inwagen agrees. He acknowledges, furthermore, that he cannot decisively establish the validity of the principle. Nevertheless, he contends that the rule is intuitively plausible and that it is resistant to non-question begging counterexamples.[4]

Van Inwagen's confidence in the validity of rule B may have been misplaced. In recent years several apparent counterexamples to the principle have emerged, the most incisive of which involve various forms of overdetermination.[5] I discuss one such example momentarily. But first an example of van Inwagen's will serve as a useful introduction to that discussion.

In an effort to highlight the intuitive plausibility of rule B, van Inwagen (1983, p. 187) invites us to consider a couple of examples, including the following case, which has come to be known as Snakebite. Suppose that no one is even partly morally responsible for the fact that a cobra bites John on his thirtieth birthday. Suppose, furthermore, that no one is even partly morally responsible for the fact that if a cobra bites John on his thirtieth birthday then John dies on his thirtieth birthday. It seems to follow from these two suppositions, in keeping with rule B, that no one is even partly morally responsible for the fact that John dies on his thirtieth birthday.

Examples like Snakebite have been called "one-path" cases because there is only one "causal pathway" leading to the relevant outcome.[6] If we focus exclusively on cases like that, rule B can seem unimpeachable. Evidently, though, there can be more than one causal pathway leading to an outcome, an observation many believe points the way to counterexamples to rule B.

---

[4] See van Inwagen (1983, pp. 186–187).

[5] See Ravizza (1994) and Fischer and Ravizza (1998, ch. 6) for the early overdetermination cases. For attempts to provide counterexamples to rule B that do not involve overdetermination, see Shabo (2010a) and Widerker (2002).

[6] Fischer and Ravizza (1998, p. 166). Following Fischer and Ravizza, I use "causal pathways" to refer both to existing causal sequences that actually causally contribute to an outcome, as well as to existing sequences that do not causally contribute to the outcome but would have if they had been allowed to reach completion.

Consider, for instance, the following case, which I will refer to as Double Dose. Kathy stands to gain a substantial amount of money if her husband, Earl, dies before noon. So, at 11:49, she injects Earl with a lethal poison for which there is no antidote with the intention of killing him before noon. The poison with which she injected Earl takes no longer than ten minutes to kill. So, by giving him the poison at 11:49, Kathy ensures that Earl meets his untimely end by 11:59, at the latest. Kathy does all this of her own free will, despite knowing that it is morally wrong. Five minutes later, at 11:54, Earl ingests an even deadlier substance, one that takes no more than five minutes to kill. His ingesting this second substance at 11:54 is sufficient, given the laws of nature and other background conditions, to bring about his death before noon. Both the poison given to Earl by Kathy and this second substance contribute to Earl's demise at 11:59.

This is a "two-path" case, as there are two causal pathways to the relevant outcome, and it seems to provide the materials for a non-question begging counterexample to rule B. It is plausible that no one is even partly morally responsible for the fact that if Earl ingests the second deadly substance at 11:54 then he dies before noon. We may also suppose that no one is even partly responsible for the fact that Earl ingested that second substance at 11:54. So we have

1.   NR (Earl ingests the second substance at 11:54 ⊃ Earl dies before noon);
2.   NR (Earl ingests the second substance at 11:54).

Contrary to what rule B implies, however, it does not follow from 1 and 2 that

3.   NR Earl dies before noon,

For, as I will now argue, Kathy is at least partly to blame for the fact that Earl dies before noon.[7]

Let us add a few additional details to the story. Kathy, we may suppose, was not causally determined to poison Earl, nor was it causally determined, prior to 11:49, that Earl would die before noon, for we may also imagine that, prior to 11:49, there was no causal process up and running that would have deterministically resulted in Earl dying before noon. There was, to be sure, a deterministic causal sequence that led to that outcome, one that was independent of Kathy and her agency, viz., the one involving the second deadly substance Earl ingested at 11:54. However, that causal sequence did not commence until *after* the one initiated by Kathy at 11:49 was

---

[7] Both Shabo (2010a) and Widerker (2002) offer counterexamples to rule B that, unlike Double Dose, do not involve any form of overdetermination. However, none of their cases yield counterexamples to the following temporally restricted version of the principle: NR (p ⊃ q), NRq ⊢ NRq, for all p and q such that the state of affairs described by p obtains prior to the state of affairs described by q. In Widerker's examples, the state of affairs described by p obtains *after* the state of affairs described by q, and in Shabo's example, the two states of affairs obtain *simultaneously*. Notice, however, that Double Dose yields a counterexample to this temporally restricted version of rule B as well. Let p be "Earl ingests the second deadly substance at 11:54," and let q be "Earl dies at 11:59 or thereabouts." No one is even partly morally responsible for p, nor is anyone even partly morally responsible for the fact that p implies q. However, contrary to what the temporally restricted version of rule B implies, it does not follow that no one is even partly morally responsible for the fact that q, since Kathy is at least partly morally responsible for the fact that q (where, again, q is the proposition "Earl dies at 11:59 or thereabouts").

already well underway. The causal sequence involving the second substance thus did not begin to unfold until after it had already been settled by Kathy at 11:49 that Earl would die before noon.[8] The upshot of all this is that we may assume that determinism is false in this story, that, prior to 11:49, it was not settled whether Earl would die before noon, and that it was Kathy's free action of poisoning Earl at 11:49 that sealed his fate. Given these additional assumptions, it is quite plausible that Kathy is at least partly to blame for the fact that Earl died before noon. But if she is at least partly to blame for that fact, then 3 is false, the truth of 1 and 2 notwithstanding, in which case rule B (and thus van Inwagen's direct argument) is invalid.[9]

2 Another direct argument

Related difficulties beset a direct argument of Ted Warfield's (1996). Central to his argument is the following transfer of non-responsibility principle:

Beta $\square$. NRp, $\square(p \supset q) \vdash$ NRq.[10]

This rule says, in effect, that, necessarily, if no one is even partly morally responsible for the fact that p, and q is a logical consequence of p, then lack of morally responsible for p is transmitted to 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 $\square$

---

[8] This is the main difference between Double Dose and other two-path cases thought to be counterexamples to rule B (e.g., those suggested by Fischer and Ravizza). In those other cases, the second causal sequence is already up and running before the featured agent comes on the scene, which could lead some defenders of rule B to conclude that it was already settled prior to the agent's action that the relevant outcome would obtain. But if it was already settled prior to the agent's action that the relevant outcome would obtain, defenders of rule B could argue that the agent is not morally responsible for that outcome. Whatever merit this objection has, it does not apply to the argument against rule B based on cases like Double Dose, for as we have seen, it was not settled in advance that Earl would die before noon. Rather, it was Kathy's free action that settled the matter of whether Earl would die before noon.

[9] Kathy is derivatively responsible for the fact that Earl dies before noon, as her responsibility for that fact derives from her responsibility for poisoning Earl at 11:49. Examples like Double Dose are therefore not counterexamples to rule B, if that principle is restricted to non-derivative responsibility (cf. Widerker 2002, p. 319). Moreover, I cannot think of any case that would provide a clear, uncontroversial counterexample to the principle once we add this restriction to the temporal restriction mentioned in note 8. It could perhaps be argued that the Frankfurt-style cases would provide counterexamples even to that restricted version of rule B. Perhaps they would, though I am skeptical. In any event, there is sufficient controversy over the Frankfurt-style cases that I doubt whether appealing to them in this context is likely to shed any light on whether the restricted version of rule B is valid.

[10] Warfield restricts the principle to cases where "the truth-maker for q obtains after the truth-maker for p" (1996, p. 224, n. 2).

Recall that P is any truth. So if this argument is sound, it shows that no one is even partly morally responsible for any fact, given the assumption that determinism is true.

Warfield contends that his inference rule, Beta □, is not threatened by two-path cases of the sort that seem to make trouble for rule B, and thus that his direct argument improves upon van Inwagen's. To see why Warfield thinks this, consider Double Dose again. To get a counterexample to Beta □, we would need a case in which no one is even partly morally responsible for the fact that p, q is a logical consequence of p, and yet someone is at least partly morally responsible for the fact that q. But cases like Double Dose do not seem to fit the bill, at least not initially, for while it is true that no one is even partly morally responsible for the fact that Earl ingested the second substance at 11:54, Earl dying before noon is not a logical consequence of his ingesting the second substance at 11:54. So, while Double Dose may yield a counterexample to rule B, it does not seem to have the right structure to threaten Beta □.

Two difficulties with Warfield's direct argument merit attention. First, it does not improve upon van Inwagen's, at least not in the way Warfield suggests it does, for given a plausible assumption, one Warfield explicitly endorses, Beta □ is invalid, if rule B is invalid. Consequently, if cases like Double Dose reveal rule B to be invalid, they can be used to show that Beta □ is invalid, too. Contrary to what Warfield maintains, then, Beta □ is threatened by two-path cases like Double Dose, if rule B is. Second, even setting aside the relationship between rules B and Beta □, it turns out that cases like Double Dose provide the materials for a counterexample to Beta □ after all. The upshot of all this is that Warfield's direct argument, like van Inwagen's, is unsound. I develop both of these objections to Warfield's argument in turn.

Warfield is "confident…that if no one is even partly morally responsible for the conjuncts of a conjunction, then no one is even partly morally responsible for the conjunction" (1996, p. 218).[11] He thus appears committed to the validity of the following inference rule:

C. NRp, NRq ⊢ NR (p & q).

Rule C tells us that if no one is even partly morally responsible for the fact that p, and no one is even partly morally responsible for the fact that q, then we may infer that no one is even partly morally responsible for the fact that p & q. Warfield points out that for him to be wrong about the validity of rule C, there would have to be a case in which no one is even partly morally responsible for the fact that p, no one is even partly morally responsible for the fact that q, but someone is at least partly morally responsible for the fact that p & q. Warfield reports that neither he nor

---

[11] Warfield makes this claim in the context of arguing that the lone premise of his direct argument is strictly equivalent to the conjunction of the two premises of van Inwagen's argument.

anyone with whom he has discussed the matter can think of a case fitting that description, and neither can I.[12]

Using rule C, we can show that Warfield's direct argument does not improve upon van Inwagen's in the way Warfield claims it does. Specifically, we can show that, given the validity of rule C, cases like Double Dose are a threat to Beta □, if they are a threat to rule B. To see this, notice that, given rules C and Beta □, rule B is *provably* valid. Here is the proof:

1. NRp, NR (p ⊃ q)	assumption
2. NR (p & (p ⊃ q))	from 1 by rule C
3. □ (p & (p ⊃ q)) ⊃ q	logical truth
4. NRq	from 2 and 3 by Beta □

Now, if rule B is indeed invalid, there must be something wrong with this proof of the principle. Assuming the validity of rule C, the culprit would obviously have to be Beta □. The upshot: given the validity of rule C, Beta □ is invalid, if rule B is invalid, which means that two-path cases like Double Dose are a threat to Beta □ after all, if they are threat to rule B. Warfield's direct argument therefore does not improve upon van Inwagen's in the way it was supposed to.

I have just argued that the fates of rules B and Beta □ are tied together. Specifically, I argued that, given the validity of rule C, the inference rule Beta □ is invalid, if rule B is invalid. However, we have already seen that rule B is invalid. Hence, Beta □ must also be invalid.

This first objection to Beta □ is indirect. It begins with the observation that rule B is invalid, shows that, given the validity of rule C, Beta □ is invalid if rule B is, and so concludes that Beta □ is invalid, too. This conclusion can also be reached directly, i.e., without appealing to the relationship between rules B and Beta □. To see this, consider Double Dose again. Recall that, given the laws of nature and other background conditions, Earl ingesting the second deadly substance at 11:54 *ensured* that he would die before noon. So we have

1. □ ((the actual laws of nature and other relevant background conditions obtain & Earl ingests the second substance at 11:54) ⊃ Earl dies before noon).

We have assumed that no one is even partly morally responsible for the fact that Earl ingests the second deadly substance at 11:54, and we may also assume that no

---

[12] An interesting side note: if rules A and B are both valid, then rule C is provably valid. Here is the proof:

1. NRp and NRq	assumption
2. □ (p ⊃ (q ⊃ (p & q)))	logical truth
3. NR (p ⊃ (q ⊃ (p & q)))	from 2 by rule A
4. NR (q ⊃ (p & q))	from 1 and 3 by rule B
5. NR (p & q)	from 1 and 4 by rule B

one is even partly morally responsible for the laws of nature or other relevant background conditions. So we have

2. NR (the actual laws of nature and other relevant background conditions obtain & Earl ingests the second deadly substance at 11:54).

If Beta $\square$ were valid, it would follow from 1 and 2 that

3. NR Earl dies before noon.

But, as we have seen, 3 is false, the truth of 1 and 2 notwithstanding, as Kathy is at least partly to blame for the fact that Earl died before noon. Beta $\square$ is therefore invalid.

3 Transferring non-responsibility the right way

The basic idea behind transfer of non-responsibility principles like rules B and Beta $\square$ seems to be something like this: if a person is not even partly morally responsible for a certain event or state of affairs, and if that person is also not even partly morally responsible for the fact that that event or state of affairs led to some further outcome, then the person is not even partly morally responsible for that further outcome either. Let us refer to this basic idea as Transfer NR.

Two-path cases like Double Dose are counterexamples to Transfer NR no less than to the formal inference rules B and Beta $\square$ that encapsulate it. Kathy is not even partly to blame for Earl ingesting the second deadly substance at 11:54, nor is she even partly to blame for the fact that his ingesting that substance at that time led to his dying before noon. But Kathy is at least partly to blame for the fact that Earl died before noon, contrary to what Transfer NR implies.

There is, however, a closely related idea that, as we shall see, is not threatened by two-path cases. Here it is: if a person is not even partly morally responsible for *any* of the circumstances that led to a particular outcome, and if that person is not even partly morally responsible for the fact that those circumstances led to that particular outcome, then the person is not even partly morally responsible for the outcome in question either. Let us refer to this related idea as Transfer NR*. Later I will introduce a formal inference rule that encapsulates this new transfer of non-responsibility principle, and I will use it in an argument for the conclusion that no one is even partly morally responsible for anything, if determinism is true. Before doing either of those things, however, I will argue that Transfer NR* is intuitively plausible, that it is not threatened by two-path examples like Double Dose, and that it has important advantages over two other transfer of non-responsibility principles that have been proposed in recent years.

Consider Todd, a toddler who, unfortunately, has contracted malaria. Is Todd's mother even partly to blame for the fact that he is now suffering from this terrible malady? Well, it depends. Is she even partly to blame for any of the circumstances that contributed to his contracting the disease? Suppose not. Is she even partly to blame for the fact that those circumstances led to his contracting malaria? Suppose

not. Then it is hard to see how Todd's mother could be even partly to blame for the fact that he is now suffering from the disease.

Or suppose that a window in Mr. Wilson's house has been shattered to bits. Is the next-door neighbor kid, Dennis, even partly to blame for that fact? Again, it depends. Is Dennis even partly to blame for any of the circumstances that contributed to the window breaking? Suppose not. Is Dennis even partly to blame for the fact that those circumstances contributed to the breaking of Mr. Wilson's window? Suppose not. Then it is hard to see how Dennis could be even partly to blame for the fact that Mr. Wilson's window has been shattered to bits.

Reflection on examples like these helps bring out the intuitive appeal of Transfer NR*. Whatever support such cases provide for the principle is defeasible, to be sure. However, in the absence of a clear counterexample or some other reason to think Transfer NR* is false, I should think examples like these provide us with at least some reason to accept the principle.[13]

Not only is Transfer NR* intuitively plausible, it is also immune to two-path cases like Double Dose. To get a counterexample to Transfer NR*, we would need a case in which a person is not even partly morally responsible for any of the circumstances that led to the relevant outcome, the person is also not even partly morally responsible for the fact that those circumstances led to that particular outcome, but the person is nevertheless at least partly morally responsible for the outcome. A moment's reflection reveals that two-path cases like Double Dose do not fit the bill. Kathy is at least partly morally responsible for the fact that Earl died before noon, to be sure, but she is also at least partly morally responsible for at least one of the events that led to its being the case that Earl died by noon, viz., her act of poisoning Earl at 11:49. Such cases therefore do not provide the materials for a counterexample to Transfer NR*.[14]

Transfer NR* is not the only transfer of non-responsibility principle that is immune to two-path counterexamples, however. Both Michael McKenna (2001) and

---

[13] McKenna (2008) notes that the sorts of cases to which van Inwagen and others typically appeal in an effort motivate their preferred transfer of non-responsibility principle (cases like Snakebite, e.g.) are about responsibility for outcomes, not intentional actions. Notice that I too have appealed to outcome cases. McKenna goes on to argue that unless incompatibilists can produce confirming instances of their preferred transfer principle that involve normal human agency (cases in which a causal path passes from some antecedent set of circumstances, through ordinary deliberation, to some voluntary action), they will have failed to show that the principle is universally valid, as they will have failed to show that the relevant principle is applicable to the cases that really matter. For a response to McKenna that is right on the money, see Schnall and Widerker (2012). They point out that the sorts of confirming cases McKenna insists incompatibilists must produce are easy to come by, and that, even if they were not, this would not undermine attempts to motivate the relevant transfer of non-responsibility principle using outcome cases. As they observe, the point of citing such examples is not to generalize from specific cases to the conclusion that the relevant transfer principle is valid. Rather, the point of the examples is to "elicit a certain logical and conceptual intuition. We realize, through reflecting on an example like Snakebite, that if its premises are true, the conclusion must be true as well, and that therefore, [rule B], which encapsulates this intuition, must be valid" (p. 32). While Schnall and Widerker focus on rule B, their defense of that rule applies, mutatis mutandis, to Transfer NR* as well.

[14] Although I cannot develop the point at length here, it is worth noting that, as far as I can tell, other alleged counterexamples to the likes of rule B (e.g., those suggested by Shabo 2010a; Widerker 2002) pose no threat to Transfer NR*.

Eleonore Stump ([2000](#), [2002](#)) have suggested transfer principles which have this virtue as well, and which, they claim, can be used to show that determinism precludes moral responsibility. How does Transfer NR* compare with these other principles? I believe that it has important advantages over both of them. I will argue, in particular, that there are difficulties in employing these other principles in arguments for the conclusion that determinism precludes moral responsibility, difficulties that do not arise when attempting to employ Transfer NR* in an argument for that conclusion.

Consider, first, the principle suggested (but not ultimately endorsed) by McKenna. He observes that the sorts of cases that lend support to Transfer NR are one-path cases such as Snakebite, whereas the cases that tell most strongly against such principles are two-path cases like Double Dose. This might suggest that these principles are best restricted in such a way that they only apply to one-path cases. According to McKenna, a transfer of non-responsibility principle restricted in this way remains intuitively plausible and yet is immune to two-path cases of the sort that threaten Transfer NR.[15] Let us refer to McKenna's restricted principle as Transfer NR-o.[16] McKenna's formulation of Transfer NR-o is fairly complex, but the details of the principle need not detain us here. The important thing to keep in mind is that it is only applicable to one-path cases. It is this feature of the principle that insulates it from two-path counterexamples like Double Dose. However, that same feature leads to other complications.

John Fischer believes that "it is futile to restrict [the relevant transfer of non-responsibility principle] to one-path cases…precisely because this would render the principle unable to generate incompatibilism about causal determinism and moral responsibility" To illustrate the point, Fischer invites us to consider "a world—unusual as this would be—with all of the morally significant behavior occurring as a result of simultaneous overdetermination." He then points out that principles like Transfer NR-o are "entirely consistent with moral responsibility coexisting peacefully with causal determinism in this world" ([2006](#), pp. 163–164).

The difficulties Fischer raises for using Transfer NR-o in an argument for the incompatibility of determinism and moral responsibility are not obviously insuperable. As Fischer himself acknowledges, "someone might reply that [Transfer NR-o] might be a *first step* in an argument for incompatibilism. On this approach, one would employ [the principle] to rule out moral responsibility in one-path cases, and then supplement it with the contention that merely adding [additional] causally deterministic paths cannot issue in responsibility" (p. 164). Let us refer to the approach described here by Fischer as the two-step approach.

I will not attempt to assess the two-step approach here.[17] Instead, I should like to point out that, whatever merits this approach may have, the problem it is designed to address does not arise in connection with Transfer NR*. Because that principle is not restricted to one-path cases, it applies no less to cases involving

---

[15] See especially McKenna ([2001](#), pp. 45–46).

[16] Shabo ([2010b](#), [2011](#)) also defends versions of Transfer NR-o.

[17] For further discussion of these and related issues, see Shabo ([2011](#)).

overdetermination than to those that do not. Accordingly, any argument based on that principle should apply no less to situations like the one Fischer invites us to consider in which all the relevant outcomes are overdetermined than to situations in which they are not. Transfer NR* thus has at least one important advantage over Transfer NR-o, insofar as it enables us to bypass the sorts of difficulties that arise when trying to use the latter principle in an argument for the incompatibility of determinism and moral responsibility.

Consider next Stump's preferred transfer of non-responsibility principle. According to Stump, counterexamples to the likes of Transfer NR must involve an agent who acts indeterministically. Otherwise, critics of such principles could not claim that the agent is responsible for the relevant outcome without presupposing the compatibility of determinism and responsibility, which, given the present dialectical context, would beg the question at issue.[18] But if an essential feature of the counterexamples is that they involve an agent who is not deterministically caused to act, such cases could not be used to challenge a version of Transfer NR restricted to deterministic contexts. Let us refer to this restricted principle as Transfer NR-d.

Can Transfer NR-d be used to argue for the incompatibility of determinism and moral responsibility? Fischer thinks not. The problem with using the principle in this way, he contends, is that it is difficult to assess the principle independently of the question of whether determinism precludes moral responsibility. Because Transfer NR-d is restricted to deterministic contexts, Fischer grants that it will be impossible for compatibilists to provide non-question begging counterexamples to the principle, since doing so would seem to require providing a case in which someone is morally responsible for an outcome that was causally determined entirely by states and events over which the agent had no control. At the same time, though, he contends that it will be equally impossible for incompatibilists to provide the necessary support for the principle without begging the question against compatibilists, since the only contexts in which the principle is applicable are those in which determinism is true. It is thus unclear whether appealing to principles like Transfer NR-d in an argument for the incompatibility of determinism and moral responsibility is dialectically kosher. At best, Fischer concludes, appealing to such principles will result in a dialectical stalemate, a situation in which theorists can neither defend nor refute a key principle in the debate without begging the question against their opponents.[19]

Fischer may have a point here. Notice, though, that similar difficulties do not arise when using Transfer NR* to argue for the incompatibility of determinism and moral responsibility, since that principle is not restricted to deterministic contexts. It applies to indeterministic situations no less than deterministic ones. When adducing considerations in support of the principle, we therefore need not restrict our attention to deterministic scenarios and so need not risk begging any questions against compatibilists. Notice, in particular, that neither of the examples to which I

---

[18] See especially Stump (2002, pp. 40–42).

[19] See, e.g., Fischer (2006, pp. 166–167).

appealed earlier in support of Transfer NR* presuppose the truth of determinism. The principle thus has an important advantage over Transfer NR-d as well, insofar as it enables us to bypass the sorts of difficulties that arise when trying to use the latter principle in an argument for the incompatibility of determinism and moral responsibility.

A related issue merits attention before moving on. I have occasionally heard it suggested that if we were to find an incompatibilist-friendly transfer of non-responsibility principle that is safe from non-question begging counterexamples, it would be little more than a formalized expression of the fundamental incompatibilist thought that a person is not morally responsible for his actions if they are causally determined by factors beyond the person's control. But it would obviously be dialectically inappropriate to appeal to that fundamental incompatibilist thought in an argument for the conclusion that determinism precludes moral responsibility, as doing so would clearly beg the question against compatibilists. For this reason, some have concluded that transfer of non-responsibility principles cannot make a significant contribution to the overall case for incompatibilism about determinism and moral responsibility.[20]

The suggestion underlying this argument should be resisted, however, at least as it pertains to Transfer NR*. That principle is neither equivalent to nor (in the absence of additional premises) does it entail the fundamental incompatibilist thought. Moreover, as I have already pointed out, the principle applies no less to indeterministic situations than to deterministic ones. It is a perfectly general principle governing the transfer of non-responsibility that, by itself, apart from further argumentation, is entirely neutral when it comes to the debate over the compatibility of determinism and moral responsibility. I therefore see no dialectical impropriety in deploying the principle in an argument for the incompatibility of determinism and moral responsibility.

4 Frankfurt on the transfer of non-responsibility

I will explain how Transfer NR* can be used to argue for the incompatibility of determinism and moral responsibility momentarily. Before doing so, though, I want to briefly respond to one further challenge to the principle suggested by some remarks of Harry Frankfurt's.

Frankfurt reports that principles like the original Transfer NR do not strike him as at all plausible. He writes, "I do not see why it should be thought that if p has some feature, and if the fact that p necessitates q has that feature too, it must follow that q also has it. One might just as well maintain the absurdity that if p has no cause, and if the fact that p causes q has no cause, then q has no cause" (2002, p. 63). Presumably, these remarks could be applied, mutatis mutandis, to other transfer of non-responsibility principles such as Transfer NR* as well.

---

[20] See, e.g., McKenna (2008, p. 379).

Consider the general transfer thesis to which Frankfurt adverts: if p has some feature F, and if the fact that p necessitates q is also F, then q must have F as well. To illustrate the implausibility of this thesis, Frankfurt points to the absurd principle that results if F is the feature of having no cause. Call that absurd principle Transfer NC. It says, "if p has no cause, and if the fact that p causes q has no cause, then q has no cause." I think we can all agree with Frankfurt that both the general transfer thesis and absurdities like Transfer NC to which it leads are false. But what has this got to do with transfer of non-responsibility principles like Transfer NR*?

Perhaps the thought is that proponents of the various transfer of non-responsibility principles are using the more general transfer thesis to derive or otherwise provide support for those principles. Alternatively, perhaps the thought is that once we see how absurd the general transfer thesis is we will no longer be inclined to accept transfer of non-responsibility principles of the sort that figure in direct arguments for the incompatibility of determinism and moral responsibility. Neither thought, though, is especially promising. As far as I am aware, proponents of the various direct arguments have never defended their preferred transfer of non-responsibility principle by appealing to the general transfer thesis identified by Frankfurt. At any rate, I certainly cannot be accused of doing so. Moreover, none of the transfer of non-responsibility principles discussed in this article entail that general transfer thesis. So it is hard to see how the fact that the general transfer thesis is false casts doubt on any of those principles. The fact that some transfer principles are invalid is not a very good reason for thinking they all are. Frankfurt's appeal to the general transfer thesis thus seems to be a bit of red herring. One can consistently embrace transfer of non-responsibility principles of the sort that figure in direct arguments for the incompatibility of determinism and moral responsibility without having to embrace other implausible transfer principles of the sort Frankfurt rightly pillories.

Frankfurt also claims that even if Transfer NR and its ilk were highly plausible, that would not be a very good reason to accept those principles (2002, p. 62). In support of this claim, he cites the fact that many people now confidently reject principles that once seemed to them highly plausible. The implication seems to be that transfer of non-responsibility principles are no different than these other principles and, accordingly, that they too should not be accepted.

It is of course true that we sometimes find ourselves rejecting principles that we find, or once found, appealing. But I see no reason to think that this should lead us to question Transfer NR*. Typically when we reject a claim we find, or once found, plausible, this is because we have (what we take to be) good reasons for thinking the claim is false. In the absence of any reason to reject an otherwise plausible principle, though, the fact that the principle is intuitively appealing and is confirmed by a range of different examples strikes me as a perfectly good, albeit defeasible, reason to accept it. I thus see nothing in Frankfurt's discussion of transfer of non-responsibility principles that should lead us to question the likes of Transfer NR*.

5 A new direct argument

So far I have argued that Transfer NR* is intuitively plausible, that it is immune to two-path counterexamples like Double Dose, and that it has important advantages over other transfer of non-responsibility principles that have been proposed in recent years. I will now use the principle to argue for the conclusion that no one is even partly morally responsible for anything, if determinism is true. I begin with an informal statement of the argument and then develop a more regimented version using the sort of technical machinery introduced by van Inwagen.

Suppose that Jones performs a certain action A at time t and that this is his first action or omission. Assuming determinism is true, it is a necessary truth that certain antecedent states and events, in conjunction with the laws of nature, resulted in Jones A-ing at t. But, necessarily, no one is even partly morally responsible for necessary truths. So, Jones is not even partly morally responsible for the fact that the relevant states and events, together with the laws of nature, resulted in him A-ing at t. Moreover, given that A was Jones's first action or omission, he is surely not even partly morally responsible for any of the states and events that led to him A-ing at t, nor is he even partly morally responsible for the laws of nature. But if Jones is not even partly morally responsible for any of the circumstances that led to him A-ing at t, and if he is not even partly morally responsible for the fact that those circumstances led to him A-ing at t, it follows from Transfer NR* that Jones is not even partly morally responsible for A-ing at t. Therefore, Jones is not even partly morally responsible for A-ing at t, if determinism is true.

By itself this is not an especially interesting conclusion, given that A was Jones's first action or omission. However, if the argument for it is sound, it seems clear that a similar line of reasoning will apply, mutatis mutandis, to every subsequent state or event for which Jones might be morally responsible, including all of his subsequent behavior. To see this, suppose Jones performs another action B at time t2 (where t2 is later than t) and that this is his second action or omission. Assuming determinism is true, it is a necessary truth that certain antecedent states and events, in conjunction with the laws of nature, resulted in Jones B-ing at t2. But, necessarily, no one is even partly morally responsible for necessary truths. So, Jones is not even partly morally responsible for the fact that the relevant states and events, together with the laws of nature, resulted in him B-ing at t2. Moreover, given that B was Jones's second action or omission, it is plausible that he is not even partly morally responsible for any of the states and events that resulted in him B-ing at t2, unless, perhaps, he is at least partly morally responsible for some prior action or omission of his that is among the circumstances that resulted in him B-ing at t2. But since B is Jones's second action or omission, and since he is not even partly morally responsible for his first action or omission (i.e., his A-ing at t), then Jones is not even partly morally responsible for any actions or omissions of his prior to his B-ing at t. Hence, Jones is not even partly morally responsible for any of the antecedent states and events that led to him B-ing at t, nor is he even partly morally responsible for the laws of nature. But if Jones is not even partly morally responsible for any of the circumstances that led to him B-ing at t2, and if he is not even partly morally

responsible for the fact that those circumstances led to him B-ing at t2, it follows from Transfer NR* that Jones is not even partly morally responsible for B-ing at t2. Therefore, Jones is not even partly morally responsible for B-ing at t2, if determinism is true.

Similar arguments could clearly be adduced for the conclusion that Jones is not morally responsible for any subsequent state or event, if determinism is true. But if so, then since Jones is not even partly morally responsible for what happened prior to his first action or omission, we may conclude that Jones is not even partly morally responsible for anything, if determinism is true. And what goes for Jones in this case goes for the rest of us, too. So, we should further conclude that no one is even partly morally responsible for anything, if determinism is true.

A more regimented version of this argument can be constructed using the sort of technical machinery employed by van Inwagen in the original direct argument. Let "α" be the proposition that "Jones A-ed at t," let "C_α" be a proposition describing all the antecedent circumstances that led to its being the case that Jones A-ed at t, including the laws of nature, and let "$NR_S p$" abbreviate "agent S is not even partly morally responsible for p."[21] Here are two inference rules that, I contend, govern the use of the NR_S operator:

A*. $\Box p \vdash NR_S p$
B*. $NR_S (C_p \supset p), NR_S C_p \vdash NR_S p$.

The first of these two inference rules is an agent-relative version of van Inwagen's rule A. It says that if it is broadly logically necessary that p, then we may infer that agent S is not even partly morally responsible for the fact that p. The second rule is a formal version of Transfer NR*. Where "C_p" is a proposition describing all of the antecedent circumstances that contributed to its being the case that p, this rule tells us that if agent S is not even partly morally responsible for the fact that C_p implies p, and if S is not even partly morally responsible for the fact that C_p, then we may infer that S is not even partly morally responsible for the fact that p.

With this formal machinery in place, I argue as follows:

0. Determinism is true	assumption for conditional proof
1. $\Box (C_\alpha \supset \alpha)$	formal consequence of 0
2. $NR_{Jones} (C_\alpha \supset \alpha)$	from 1 by rule A*
3. $NR_{Jones} C_\alpha$	premise
4. $NR_{Jones} \alpha$	from 2 and 3 by rule B*

A few remarks about each of this argument's main steps are in order.

Step 1 follows directly from determinism. In a deterministic universe, if the exact circumstance described by C_α (which, you will recall, include the laws of nature) led to Jones A-ing at t, then any world in which those same circumstances obtain is one

---

[21] Where p is a conjunction, I stipulate that "S is not even partly morally responsible for the fact that p" is to be understood as "S is not even partly morally responsible for p or for any conjunct of p".

in which Jones A-ed at t. In other words, if determinism is true, any world at which C_α is true is one at which α is true.

Step 2 follows from step 1 by rule A*.[22]

Step 3 is also very plausible. Since A is Jones's first action or omission, we may assume, plausibly, that he is not even partly morally responsible for any of the states or events that led to its being the case that α. It is also safe to assume that Jones is not even partly morally responsible for the laws of nature. It is therefore quite plausible to suppose that step 3 is true.

Finally, step 4, which says that Jones is not even partly morally responsible for α follows from steps 2 and 3 by rule B*. As I mentioned earlier, 4 is not an especially interesting conclusion by itself, given that A is Jones's first action or omission. However, as I also mentioned earlier, it should be clear that if the argument for 4 is sound, a similar line of reasoning will apply, mutatis mutandis, to every subsequent fact for which Jones might be even partly morally responsible. And so, since Jones is not even partly morally responsible for what did or did not happened prior to his first action or omission, it seems we can conclude that he is not even partly morally responsible for any fact, if determinism is true. Moreover, we can preplace "Jones" in this argument with any human agent and get similar results. And so I conclude that no one is even partly morally responsible for anything, if determinism is true.

6 The significance of the direct argument

I conclude with some brief remarks about the dialectical significance of direct arguments like the one defended here. Standard attempts to establish the incompatibility of determinism and moral responsibility begin with the claim that the freedom to do otherwise is required for moral responsibility and then attempt to show that no one has that sort of freedom, if determinism is true. By contrast,

---

[22] Kearns (2011) challenges rules A and A*, as does Hermes (2014). A co-author and I take up these challenges in future work. For now I should like to point out that even if rule A* is invalid, the new direct argument can be reformulated without it. Let "$\rightarrow$" stand for entailment. Now consider the following variation on rule B*

B**: $NR_S (C_p \rightarrow p)$, $NR_S C_p \vdash NR_S p$

If B* is valid, there can be little doubt that B** is also valid. Using B** the new direct argument can be reformulated without appealing to rule A*:

0. Determinism is true	assumption for conditional proof
1. $(C_\alpha \rightarrow \alpha)$	formal consequence of 0
2. $NR_{Jones} (C_\alpha \rightarrow \alpha)$	premise
3. $NR_{Jones} C_\alpha$	premise
4. $NR_{Jones} \alpha$	from 2 and 3 by rule B**

Step 1 is equivalent to step 1 in the text. Without appealing to rule A*, we cannot *derive* step 2 from step 1. Nevertheless, step 2 remains highly plausible. That one state of affairs is the logical consequence of another just does not seem like the sort of fact for which a person could be to praise or blame. The upshot of all this is that even if rule A* is invalid, I think this would do little if anything to undercut the appeal of the new direct argument.

arguments like the ones discussed in this article attempt to establish the incompatibility of determinism and moral responsibility directly, without reference to the freedom to do otherwise. An apparent advantage of such arguments is that they seem to provide a means of defending incompatibilism about determinism and moral responsibility without having to take a stand on various metaphysical controversies concerning the freedom to do otherwise. As it turns out, however, things are not quite that simple.

Assume for the sake of argument that the freedom to do otherwise is compatible with determinism. Given that assumption, the compatibilist about determinism and moral responsibility could plausibly claim that an ordinary agent in a deterministic universe who satisfies all the relevant compatibilist requirements for responsibility, and who deliberately performs an action he knows to be immoral, despite believing correctly that he is free to do the right thing instead, is blameworthy for performing that action. But clearly if the agent is blameworthy for his action, the truth of determinism notwithstanding, then determinism and moral responsibility are compatible after all, in which case there must be something wrong with all arguments for incompatibilism, the direct arguments included. To avoid this conclusion, it seems proponents of the various direct arguments we have considered would have to reject either the initial assumption that freedom to do otherwise is compatible with determinism or the assumption that an agent who satisfies all the relevant compatibilist conditions, and who deliberately performs an action he knows is wrong, despite believing correctly that he is free to do the right thing instead, is blameworthy for performing that action. But this latter assumption is extremely plausible. Arguably, then, all incompatibilists about determinism and moral responsibility must insist on the incompatibility of determinism and the freedom to do otherwise. But if so, then the various direct arguments do not enable us to completely bypass controversies about freedom to do otherwise and its relationship to determinism and moral responsibility.[23]

If the freedom to do otherwise is compatible with determinism, then, as the argument I just sketched seems to indicate, so is moral responsibility. But what conclusions should we draw from the truth of this conditional about the dialectical significance of the direct arguments? David Widerker concludes that such arguments do not enable us to bypass debates about the relationship between determinism and freedom to do otherwise, and that such arguments "should not be of concern to compatibilists, since [compatibilists] reject the assumption that in a deterministic world the agent cannot act otherwise" (2002: 323).[24] I draw an importantly different conclusion. While it may be true that proponents of the various direct arguments cannot just ignore the question of whether determinism precludes the freedom to do otherwise, I do not think we should conclude from this that

---

[23] The argument of this paragraph is inspired by that of Widerker (2002).

[24] Widerker's claim that compatibilists about determinism and moral responsibility reject the assumption that determinism precludes the freedom to do otherwise is too broad, for not all compatibilists reject that assumption. A semi-compatibilist, someone who holds that determinism is compatible with responsibility even if it is not compatible with the freedom to do otherwise, need not reject the assumption and, indeed, some semi-compatibilists accept it.

compatibilists have no reason to worry about the direct arguments or that these arguments cannot usefully advance the debate over the compatibility of determinism and moral responsibility. On the contrary, it seems to me that, far from showing that direct arguments are of little dialectical consequence, the truth of the conditional at issue highlights just how useful such arguments can be, for if we append that conditional to the direct argument defended here, we get a new argument for the incompatibility of determinism and the freedom to do otherwise. In its simplest form, that new argument goes like this:

1. If the freedom to do otherwise is compatible with determinism, then so is moral responsibility (premise).
2. Moral responsibility is not compatible with determinism (from the new direct argument).
3. Therefore, freedom to do otherwise is not compatible with determinism.

Unsurprisingly, I think this argument is sound. But whatever you make of the argument, its existence shows that, contrary to what Widerker claims, compatibilists should take seriously direct arguments for the incompatibility of determinism and moral responsibility. While these arguments may not enable us to completely bypass debates about the freedom to do otherwise in the way they initially seemed to, it hardly follows that they are of little dialectical significance.

References

Fischer, J. M. (2006). *My way*. New York: Oxford University Press.
Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: An essay on moral responsibility*. Cambridge: Cambridge University Press.
Frankfurt, H. (2002). Reply to Eleonore Stump. In Sarah Buss & Lee Overton (Eds.), *The contours of agency: Essays on themes from Harry Frankfurt*. Cambridge, MA: MIT Press.
Hermes, C. (2014). A counterexample to rule A. *Philosophia, 42*, 387–389.
Kearns, S. (2011). Responsibility for necessities. *Philosophical Studies, 155*, 307–324.
McKenna, M. (2001). Source incompatibilism, ultimacy, and the transfer of non-responsibility. *American Philosophical Quarterly, 38*, 37–51.
McKenna, M. (2008). Saying good-bye to the direct argument the right way. *Philosophical Review, 117*, 349–383.
Ravizza, M. (1994). Semicompatibilism and the transfer of non-responsibility. *Philosophical Studies, 75*, 61–94.
Schnall, I., & Widerker, D. (2012). The direct argument and the burden of proof. *Analysis, 72*, 25–36.
Shabo, S. (2010a). Against logical versions of the direct argument: A new counterexample. *American Philosophical Quarterly, 47*, 239–252.
Shabo, S. (2010b). The fate of the direct argument and the case for incompatibilism. *Philosophical Studies, 150*, 405–424.
Shabo, S. (2011). What must a proof of incompatibilism prove? *Philosophical Studies, 154*, 361–371.
Stump, E. (2000). The direct argument for incompatibilism. *Philosophy and Phenomenological Research, 61*, 459–466.

Stump, E. (2002). Control and causal determinism. In Sarah Buss & Lee Overton (Eds.), *The contours of agency: Essays in honor of Harry Frankfurt*. Cambridge, MA: MIT Press.

van Inwagen, P. (1980). The incompatibility of responsibility and determinism. *Bowling Green Studies in Applied Philosophy, 2*, 30–37.

van Inwagen, P. (1983). *An essay on free will*. Oxford: Clarendon Press.

Warfield, T. A. (1996). Determinism and moral responsibility are incompatible. *Philosophical Topics, 24*, 215–226.

Widerker, D. (2002). Farewell to the direct argument. *Journal of Philosophy, 99*, 316–324.