TRANSFER PRINCIPLES AND MORAL RESPONSIBILITY

Eleonore Stump
Saint Louis University

John Martin Fischer
University of California, Riverside

It is useful to divide contemporary arguments for the incompatibility of causal determinism and moral responsibility into two types: indirect and direct. The indirect arguments present reasons why causal determinism is incompatible with the possession of the relevant kind of alternative possibilities and conclude from this that causal determinism is incompatible with moral responsibility. It is, of course, a presupposition of the indirect arguments that moral responsibility requires alternative possibilities. The direct arguments contain no such presupposition, although some of their proponents may believe that moral responsibility does indeed require alternative possibilities.

The direct arguments employ what might be called "transfer" principles. These are principles that transfer a certain property; the relevant property here *is lack of moral responsibility*.[1] Let "Np" abbreviate "p and no one is even partly morally responsible for the fact that p." Then this is a transfer principle introduced by Peter van Inwagen:

Rule B: Np and N(p > q) implies Nq.[2]

Van Inwagen's Rule B is a transfer principle insofar as it transfers the property of lack of moral responsibility from one fact to another by the medium of lack of responsibility for the pertinent conditional.

Van Inwagen's direct argument for the incompatibility of causal determinism and moral responsibility can be presented simply as follows. For present purposes, we can understand causal determinism as the doctrine that a complete description of the (temporally nonrelational) state of the universe at a time, and a description of the laws of nature, entail every truth about subsequent times. Let P be a proposition describing the state of the universe before there were any human beings, let L be a proposition describing laws of nature,

and let F be a truth about the way the world is today. Then, if causal determinism is true,

(1) (P and L > F).

Clearly, no one is even partly morally responsible for this fact, and so this is also true:

(2) N[(P and L) > F].

Since [(P and L) > F] is equivalent to [P > (L > F)], this is true as well:

(3) N[P > (L > F)].

Now

(4) NP,

and so by Rule B, from (3) and (4) we can conclude

(5) N(L > F).

Since

(6) NL,

by another application of Rule B, from (5) and (6) we reach the conclusion,

(7) NF.

Since F is an arbitrary truth, this conclusion can be generalized. Consequently, the argument appears to show in a direct fashion that if causal determinism is true, no one is even in part morally responsible for any fact.

But Rule B can be called into question. Mark Ravizza offers the following kind of case to impugn Rule B.[3] At T1, Betty freely detonates explosives as part of a plan to start an avalanche that will destroy an enemy camp; and, in fact, her explosion does succeed in causing an avalanche that is sufficient to destroy the camp at T3. Unbeknown to Betty, however, there is another cause of the camp's destruction by avalanche. At T1, a goat kicks loose a boulder, and it causes an avalanche which is also sufficient to destroy the camp at T3 and which contributes to the actual destruction of the camp at T3. In the story, no one is even partly morally responsible for the goat's kicking the boulder. And no one is even partly morally responsible for the fact that if the goat kicks that boulder at T1, then the camp is destroyed by avalanche at T3. Nonetheless,

Betty is at least partly responsible for the camp's being destroyed by avalanche at T3. Thus, Ravizza's case apparently shows that Rule B is invalid. In cases of simultaneous causation, the rule fails.[4]

In a recent paper, Ted Warfield has suggested a reply on behalf of the incompatibilist.[5] He concedes that Ravizza's case presents a challenge to Rule B. Warfield claims, however, that there is a related but non-equivalent rule—he calls it 'Rule Beta □'—which can play a similar role in an argument for the incompatibility of causal determinism and moral responsibility. According to Warfield, Rule Beta □ is not subject to Ravizza-style counterexamples.

This is Rule Beta □:

[Np and $\Box(p > q)$] implies Nq.

The key difference between Rule B and Rule Beta □ is in the connection between p and q. For Rule B, it must be the case that if p, then q, and no one is responsible for this fact. For Rule Beta □, the connection between p and q is one of logical necessity. Because the connection between p and q in Warfield's Rule Beta □ is so much stronger than the connection between p and q in Rule B, Warfield supposes that it will be much harder to construct scenarios which present a challenge to his rule. For Ravizza's scenario to serve as a counterexample to an inference licensed by Rule Beta □, the connection between the goat's kicking the boulder at T1 and the camp's being destroyed by avalanche at T3 would have to be a logical one; and, of course, it is not. As Warfield says, "The conditional premise (if the goat kicks the boulder at T1, then the avalanche destroys the camp at T3), though not a proposition anyone is even partly morally responsible for, does not express a relation of logical consequence, and so Ravizza's example fails to apply to my argument [for incompatibilism]." (p.222–223)

Contrary to what Warfield claims, his Rule Beta □ *is* subject to Ravizza-style counterexamples, in our view. In what follows, we present two such counterexamples, each of which is sufficient to show that Rule Beta □ is invalid.

Counterexample A. Let it be the case that, necessarily, if the actual laws of nature obtain and the conditions of the world at T2 (some time just before T3) are C, then there will be an avalanche that destroys the enemy camp at T3. Let it also be the case that at T1 Betty *freely* starts an avalanche which is sufficient to destroy the camp at T3 and which contributes to its destruction at T3. Finally, let it be the case that Betty's freely starting an avalanche is the result of some suitable indeterministic process.

Then let r be the conjunction of

(r1) the actual laws of nature obtain

and

(r2) the condition of the world at T2 is C.

And let q be

(*q*) there is an avalanche which destroys the enemy camp at T3.

In this example, r is true. Nr is also true: nobody is even partly morally responsible for the obtaining of the actual laws of nature and the condition of the world's being C at T2. By hypothesis, it is also true that $\Box(r > q)$. Any world in which ($r1$) and ($r2$) are true is a world in which q is true. And yet it seems clear that Nq is false. Insofar as Betty at T1 freely starts an avalanche, she is at least in part morally responsible for the camp's being destroyed by an avalanche at T3.

Warfield anticipates such a case. He says,

> Can a Frankfurt-type case (or a Ravizza overdetermination case) be constructed that is a counterexample to Rule Beta □? I don't see how. To illustrate notice that making...the avalanche a logical consequence of the goat's kicking the boulder requires that we assume that [the avalanche] is a *deterministic* consequence of the arrangement of natural forces. This change would provide a case that is at least of the right form to serve as a counterexample to Rule Beta □. But to be a counterexample to Rule Beta □ the example must be an example in which the Frankfurtian judgment of moral responsibility [Betty's moral responsibility for the camp's being destroyed by an avalanche] holds up. With the additional assumption of determinism that is needed to make the case applicable to Rule Beta □, however, this Frankfurtian judgment is equivalent to the claim that determinism and moral responsibility are compatible. It is hardly of interest to point out that the assumption of the compatibility of determinism and moral responsibility implies that Rule Beta □ is invalid. (p.223)

But note that we have not assumed causal determinism in our example. Contrary to Warfield's claim, such an assumption is *not* "needed to make the case applicable to [Rule] Beta □." This is because, even in an indeterministic world, *some* events and states of affairs can be causally determined. One can suppose that the enemy camp's being destroyed by an avalanche at T3 is causally determined by the goat's kicking a boulder at T1 *without* thereby supposing that Betty's deliberations or actions are causally determined. Even in an indeterministic world, there can be "pockets of local determination".[6] To deny this is to suppose that, for any state of affairs p whatever, the laws of nature and the condition C of the world at T2 is compatible with p at T3 and also compatible with *not-p* at T3. But this is to suppose that absolutely everything in the world is indeterministic, and presumably even libertarians don't want to make so strong a claim.

Counterexample B. For those still inclined to worry about the issue of causal determinism, however, we can construct a counterexample which doesn't depend on there being even local determinism. This time let r be a conjunction of these propositions

($r1$) the actual laws of nature obtain

and

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

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

(8) Nr

and

(9) $\Box$ ($r > q$),

but it isn't the case that

(10) Nq,

for the sort of reasons given in connection with Ravizza's story.

Warfield has an objection to this sort of counterexample, too. He maintains that

(W1) if no one is even partly morally responsible for a conjunction, then no one is even partly morally responsible for either conjunct of the conjunction. (p.218)[7]

This claim calls into question Nr in our counterexample B. It is not the case that no one is even partly responsible for ($r3$). On (W1), then, it isn't the case that no one is even partly morally responsible for r. Consequently, Nr is false.

But is Warfield's claim (W1) right? We think it isn't, because of the connection between conjunctions and conditionals.

To see this, consider again Ravizza's story. It is not the case that if the actual laws of nature obtain, there will not be an avalanche that destroys the enemy camp at T3. So this is true:

(11) not (L $>$ not-q).

Furthermore, it seems odd to think that anyone is even partly responsible for (11). It is peculiar to suppose that a human being is to blame for (11), is the

source of the state of affairs described by (11), could have brought it about that
that state of affairs didn't obtain, and so on. So this also seems true:

(12) N[not (L > not-q)].

Of course, (11) is equivalent to this:

(13) (L and q).

So it seems as if this also has to be true:

(14) N(L and q).[8]

But now we have a problem, if (W1) is correct. In Ravizza's story, Betty is
partly responsible for q. Therefore, it isn't true that no one is even partly re-
sponsible for the conjuncts of (13); Betty is at least partly responsible for q. On
(W1), however, for it to be the case that no one is even partly responsible for
the conjunction, it would also have to be the case that no one was even partly
responsible for either of the conjuncts. Consequently, if Warfield's claim (W1)
is true, (14) is *false*.

In that case, however, Warfield must also hold that (12) is false. But the
claim that (12) is false strikes us as counter-intuitive.

Furthermore, as the preceding discussion shows, if (W1) is true, so is this:

(W2) Given a true antecedent of a conditional,[9] a person is partly respon-
 sible for the conditional's being false if he is partly responsible for
 the falsity of the consequent of the conditional. [10]

That's why commitment to (W1) turns out to require rejecting

(12) N[not (L > not-q)].

But if (W2) is true, it seems that this ought also to be true:

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

Why should we accept that a person is partly responsible for the falsity of a
conditional with a true antecedent because of his responsibility for the falsity
of the consequent, and yet deny that a person is partly responsible for the truth
of a conditional with a true antecedent because of his responsibility for the
truth of the consequent?

Another way to see the connection between (W1) and (W3) is to consider the reason Warfield gives for accepting (W1). To make (W1) seem plausible, Warfield says, "being at least partly morally responsible for a conjunct is a way of being partly morally responsible for a conjunction" (p.218). But if that is right, then it seems that a similar point ought to apply to conditionals: being partly responsible for the truth of the consequent of a conditional with a true antecedent is a way of being partly responsible for the truth of the conditional.

And yet (W3) is clearly mistaken. To see this, consider (9) again:

(9) □ (r > q),

where *r* is

(*r*1) the actual laws of nature obtain,

and

(*r*3) there is an avalanche which destroys the enemy camp at T3,

and *q* is identical to (*r*3). Warfield also accepts this rule of inference, taken from Peter van Inwagen:

Rule A: □ *p* implies N*p*.

Rule A seems entirely uncontroversial. In fact, Warfield says,

van Inwagen's Rule A is (nearly) as trivial and inconsequential as a rule of inference could be. No one has, to my knowledge, challenged this principle nor has anyone challenged any principle closely related to Rule A.[11]

Now, from (9), by Rule A, we get

(15) N (r > q).

On (W3), however, a person is partly responsible for the truth of a conditional with a true antecedent if she is partly responsible for the truth of the consequent of the conditional. So, on (W3), we will have to say that (15) is *false*, just as (12) is, because Betty is partly responsible for *q*. By Rule A, however, it then follows that (9) is false, since by Rule A (9) implies (15). Without doubt, this is absurd. So either Rule A is after all invalid, or (W3) is false. And if (W3) is false, then by parity of reasoning it seems that (W1) is false also.

For these reasons, we think Warfield's claim (W1) should be rejected. The logic of responsibility is more complicated than (W1) implies. Given the rela-

tion between conjunctions and conditionals, it is right to hold that someone can be partly responsible for a conjunct of a conjunction without being partly responsible for the conjunction. Consequently, our counterexample B is also effective against Rule Beta □.

   Finally, we think it is worth pointing out that one of us believes that causal determinism is incompatible with moral responsibility and the other does not. But we unite in thinking that causal determinism cannot be proved incompatible with moral responsibility by Warfield's Rule Beta □.[12]

Notes

1. In the context of the indirect arguments for the incompatibility of causal determinism and moral responsibility, one can have "Transfer of Powerlessness" principles. For a discussion of such principles, see John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control*, (Oxford: Blackwell Publishers, 1994), pp.23–66.
2. Peter van Inwagen, *An Essay on Free Will*, (Oxford: Clarendon Press, 1983), p.184.
3. Mark Ravizza, "Semi-Compatibilism and the Transfer of Nonresponsibility," *Philosophical Studies* 75 (1994), pp.61–93, esp. p.78. For similar examples, see also John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, (Cambridge: Cambridge University Press, 1998), pp.151–169.
4. One of us (Stump), but not the other, thinks it isn't clear that the invalidity of Rule B shown by Ravizza's example renders Van Inwagen's argument irremediably invalid. That is because Rule B fails only in certain cases, and it isn't clear to one of us that cases of moral responsibility can be assimilated to those cases of simultaneous causation in which Rule B fails.
5. Ted A. Warfield, "Determinism and Moral Responsibility Are Incompatible," *Philosophical Topics* 24 (1996), pp.215–226. In addition to the suggestion explored in the text, Warfield also presents other strategies for replying to the compatibilistic strategy of Ravizza; see, especially, pp.221–222. Subsequent references to Warfield's paper will be given by page number in parentheses in the text.
6. Daniel Dennett introduces the term 'local fatalism' to refer to a related but different notion: Daniel C. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting*, (Cambridge, Mass. and London: MIT Press, 1984), pp.104–106.
7. One possible reason for thinking that (W1) is true is the supposition that one is at least partly morally responsible for a conjunction if one is morally responsible for a part of a conjunction. But being morally responsible for a part of a conjunction and being partly morally responsible for a conjunction are not the same thing, as our argument in what follows helps to make clear.
8. This inference is licensed by the fact that if p and q are logically equivalent, then Np if and only if Nq.
9. It's possible to interpret (W1) as applying only to true conjunctions; in that case, this qualification in (W2) is needed.
10. Obviously, we can switch the conjuncts in the conjunction from (L and q) to (q and L), which is equivalent to

   (12*) not (q > not-L).

Since Betty is partly responsible for (q and L), she will also be partly responsible for (12*). Consequently, accepting (W1) requires accepting not only (W2) but also

> (W2*) Given a false consequent, a person is partly responsible for a conditional's being false if he is partly responsible for the truth of the antecedent of the conditional.

11. Pp.218–219. Similarly, Van Inwagen says, "The validity of Rule (A) seems to me to be beyond dispute. 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." (Van Inwagen 1983, p.184).
12. We are grateful to David Widerker, Al Mele, Chris Pliatska, and Ted Warfield for helpful comments on an earlier version of this paper.