

The Direct Argument for Incompatibilism

In their rich and impressive book *Responsibility and Control: A Theory of Moral Responsibility*, John Martin Fischer and Mark Ravizza offer an account of moral responsibility in terms of guidance control. On their view, an agent has guidance control in virtue of acting on a moderately reasons-responsive mechanism which is his own, and guidance control is "the freedom-relevant condition necessary and sufficient for moral responsibility."¹ It is an advantage of this account, they think, that it is compatible with both the truth and the falsity of causal determinism.² All of these claims raise questions which are worth pursuing.³ In this very brief paper, I can consider just one aspect of their account, namely, their rejection of incompatibilism.

As Fischer and Ravizza point out, there are two sorts of arguments that moral responsibility is incompatible with the truth of causal determinism. One is indirect; it maintains that moral responsibility requires alternative possibilities and argues that alternative possibilities are ruled out by causal determinism. The other is a direct argument that moral responsibility is ruled out by causal determinism. Fischer and Ravizza's contention that their own account of moral responsibility is compatible with the truth of causal determinism depends on their success in defusing each sort of argument. Here I want to look just at their attempt to refute the direct argument.

The direct argument is based on a principle of the transfer of non-responsibility -- 'Transfer NR', as Fischer and Ravizza call it.⁴ This principle says,

(*Transfer NR*)

If

¹ John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, (Cambridge: Cambridge University Press, 1998), p.241, ftn.2.

² See, for example, Fischer and Ravizza 1998, p.26.

³ . I have considered some of these other questions in "Persons: Identification and Freedom", *Philosophical Topics* 24 (1996) 183-214; and "Control and Causal Determinism", ***, ed. Sarah Buss and Lee Overton, forthcoming.

⁴ . Fischer and Ravizza take the direct argument from Peter van Inwagen's work; see Peter van Inwagen, "The incompatibility of Responsibility and Determinism", in Michael Bratton and Myles Brand, eds., *Bowling Green Studies in Applied Philosophy 2: Action and Responsibility* (Bowling Green, Ohio: Bowling Green State University press, 1980), pp.30-37, reprinted in John Martin Fischer, ed., *Moral Responsibility* (Ithaca, NY: Cornell University Press, 1986); and Peter van Inwagen, *An Essay on Free Will* (Oxford: Oxford University Press, 1983), pp.182-188. Van Inwagen calls the principle 'Principle Beta'.

(T1) p obtains, and no one is even partly morally responsible for p ;

and

(T2) if p obtains, then q obtains, and no one is even partly morally responsible for the fact that if p obtains, then q obtains;

then

(T3) q obtains, and no one is even partly morally responsible for q .⁵

Fischer and Ravizza attack the direct argument for incompatibilism by trying to show that Transfer NR isn't valid. Their counterexamples to the principle are all cases of preemptive or simultaneous overdetermination. Here is the counterexample they call '*Erosion*', which is a case of preemptive overdetermination.

(*Erosion*) Betty is on a mission to destroy a camp at the base of a mountain by starting an avalanche. She places her explosives accordingly and at t_1 pushes the plunger, detonating the explosives and starting an avalanche which destroys the camp at t_3 . Had Betty not detonated the explosives, at t_2 an avalanche would have been started by natural erosion in a glacier, and it would have destroyed the camp at t_3 .⁶

In this case, Transfer NR is violated insofar as the following claims are true:

"(1) The glacier is eroding and no one is, or ever has been, even partly responsible for the fact that it is eroding; and
(2) if the glacier is eroding, then there is an avalanche that crushes the enemy base at T_3 , and no one is, or ever has been, even partly responsible for this fact"⁷;

and yet it is not true that

"(3) there is an avalanche that crushes the enemy base at T_3 , and no one is, or ever has been, even partly responsible for this fact."⁸

That is, of course, because Betty is responsible for the destruction of the camp at t_3 .

⁵ . I have given the formulation of the principle as Fischer and Ravizza present it when they first introduce it; see Fischer and Ravizza 1998, p.152. When they discuss counterexamples to Transfer NR, Fischer and Ravizza sometimes formulate the principle slightly differently. So, for example, in connection with the counterexample *Erosion*, they formulate the relevant portion of the principle this way: "no one is, or ever has been, morally responsible". In general, I formulate the principle and the applications of the principle as Fischer and Ravizza do.

⁶ . See Fischer and Ravizza 1998, p.157.

⁷ . Fischer and Ravizza 1998, p.157.

⁸ . Fischer and Ravizza 1998, p.157.

In this case there are two paths to the outcome of the camp's being destroyed. One of them, the path through natural forces involving the glacier's erosion, is such that no one is responsible for it. But the other path, Betty's detonating explosives, is different; for this path, no suitable version of (T1) is true. That is, it is not true that

(T1*) Betty detonates explosives, and no one is, or ever has been, even partly responsible for the fact that Betty detonates explosives.

It is entirely reasonable to suppose that in *Erosion* Betty is responsible for her detonating explosives.

So Fischer and Ravizza are right to suppose that this example, as well as others they give, shows that Transfer NR is not a valid principle.

As they themselves recognize, however, these examples might nonetheless not undermine Transfer NR in any sense relevant to the compatibility of responsibility and determinism. It might be that any counterexample to Transfer NR has some feature which can't be found in any case involving responsibility and determinism. In that case, Fischer and Ravizza's strategy for undermining Transfer NR couldn't be employed in the cases relevant to moral responsibility and causal determinism. If Transfer NR is given a restricted formulation which excludes the Fischer and Ravizza cases of overdetermination, then, unless cases of moral responsibility can be assimilated to such cases, the direct argument for incompatibilism which relies on Transfer NR will still be sound. Fischer and Ravizza consider and reject one objection to their argument against Transfer NR which is based on this strategy.⁹ I want to use the same strategy but in service of a different objection.

As Fischer and Ravizza themselves acknowledge, all their counterexamples depend on there being more than one path to the same outcome. In arguing against Peter van Inwagen's support for Transfer NR, Fischer and Ravizza say,

"Van Inwagen focuses exclusively on the one-path cases. But among the two-path cases there is a subclass of cases in which one of the paths to the outcome contains the appropriate sort of control. These cases provide counterexamples to Transfer NR. Van Inwagen's mistake is inappropriately to seek to generalize from a proper subclass (the one-path cases) to the totality of relevant cases."¹⁰

But it isn't nearly so clear as Fischer and Ravizza suppose that we can construct two-path counterexamples to Transfer NR for moral responsibility and causal determinism. Here's why.

Suppose that Cartesian dualism is false and that any mental act or state, such as making a decision, is correlated with some neural state, where by 'correlation' we mean whatever the correct relation between the mental and the neural is, on the assumptions that causal determinism is true and that Cartesian dualism is false.¹¹ Let 'D' stand for some

⁹ . Fischer and Ravizza 1998, pp.159-163.

¹⁰ . Fischer and Ravizza 1998, pp.166-167.

¹¹ . 'Correlation' and its related terms are becoming common in neurobiology as a means of referring to the connection between the mental and the neural without specifying very

mental act or state, such as recognizing a face or making a decision, and let 'N' stand for the neural state correlated with it. For example, D might be the mental state of S's recognizing the face of his daughter. Then N would be the neural state in S's brain in which a certain sequence of neural firings -- from the retina, through the optic nerve to the lateral geniculate nucleus of the thalamus, into various layers of the visual cortex, to the inferior temporal cortex, and so on -- is completed.

Now, on the assumptions that causal determinism is true and that Cartesian dualism is false, we can construct a rejoinder to Fischer and Ravizza, in defense of Transfer NR, along these lines. On these assumptions, the following claims are true:

(T1*) The laws of nature and the conditions at the time of the Big Bang causally determine that S is in neural state N at t₁; and no human being is or ever was morally responsible for this fact.¹²

much about the nature of the connection. So, for example, in a recent article on visual perception in *Scientific American*, the author says, "Only a tiny fraction of neurons seem to be plausible candidates for what physiologists call the 'neural correlate' of conscious perception." (Nikos Logothetis, "Vision: A Window on Consciousness", *Scientific American*, November 1999, p.74)

For the purposes of this example, I am assuming two things about the nature of this correlation.

The first is that (for human beings in this world) the mental is implemented in the neural, so that there is a given mental act or state if and only if the neurons in the neural sequence correlated with that mental act or state have fired. In the formulation of this assumption, the term 'implemented', like the term 'correlation', is vague; but, however exactly 'implemented' is to be understood, the stipulation that the mind is implemented in the brain is intended to rule out the view that the mind and the brain are two distinct substances, each capable of action on its own without the other and each capable of causal interaction with the other.

The second assumption is that the correlation is a one-many relation; one mental act or state is correlated with the firings of many neurons in a neural sequence. The mental act or state doesn't occur or exist unless and until its entire correlated neural sequence is completed.

In my view, other than Cartesian dualism, most theories of the relation of mind and body (including Thomistic dualism) will be compatible with these two assumptions.

¹² . It is important to distinguish (T1*) from a related claim with which it might be confused, namely,

(T2*) If the laws of nature and the conditions at the time of the Big Bang causally determine that S is in neural state N at t1, then S makes decision D at t1; and no human being is or ever was morally responsible for the fact expressed in this conditional.

Neural states are states of material objects, namely, neurons; and if causal determinism is true, then the states of neurons, like the states of any other material objects, are causally determined by initial conditions and the laws of nature. But no human being is responsible for the fact that the laws of nature and the conditions at the time of the Big Bang determine S's being in neural state N at t. So (T1*) is true, given the presupposition that causal determinism is true.

(T2*) seems unimpeachable as well, on the presupposition that Cartesian dualism is false. If there is no separate soul isolated in its acts from events in the brain, then mental acts are correlated with neural states. But that mental states and neural states are correlated in this way is clearly not something that any human being is even partly responsible for.

From (T1*) and (T2*) it apparently follows that

(T1**) S is in neural state N at t1, and no human being is or ever was morally responsible for this fact.

A compatibilist might well want to argue that (T1**) is false. If mental and neural states are correlated, then an agent S who has a certain mental state ipso facto has the correlated neural state. Consequently, if moral responsibility is compatible with causal determinism, a compatibilist might suppose that S is responsible for his being in a particular neural state, even if that state is causally determined, just in virtue of the fact that S is responsible for his being in the correlated mental state. Whatever one thinks of this strategy for defending the falsity of (T1**), it will not be effective against (T1*). That is because there is a difference in the two claims in the nature of the fact for which responsibility is being denied. (T1**) denies that an agent is responsible for the fact that he is in a certain neural state. But (T1*) denies that an agent is responsible for the fact that the laws of nature and conditions at the time of the Big Bang determine a particular neural state of his. Whatever one might think of compatibilism, it seems clearly true that no human agent is responsible for this cosmic connection between the laws of nature and initial conditions of the world, on the one hand, and a neural state, on the other. I am grateful to Scott MacDonald for calling my attention to the need to address this point.

(T3*) S makes decision D at t1, and no human being is or ever was morally responsible for this fact.

In order to use the Fischer and Ravizza strategy to show that this conclusion does not follow and that causal determinism poses no threat to moral responsibility, we would have to show three things: (a) there is another path to the same result of S's making decision D, (b) this path is one for which someone is at least partly responsible, and (c) that someone is S himself. If there is no other path to the same result, if there is another path to the outcome of S's making decision D at t1 but it is only one for which no one is responsible, or if there is another path to the outcome for which someone is responsible but it is only one in which the responsible person isn't S, then the Fischer and Ravizza strategy will not succeed in undermining the direct argument for the incompatibility of causal determinism and moral responsibility.

Given the plasticity of the brain, it is clearly possible that there be a different path to the outcome of S's making decision D. So, for example, it might be the case that, as things are, S's neural state N is in the left hemisphere of his cerebral cortex. But if some illness had destroyed S's left hemisphere in early childhood, then S's brain would have reorganized itself so that the neural state correlated with S's mental act of making decision D would have been not neural state N but rather some neural state R in S's right hemisphere. In that case, there would be a different path to the outcome of S's making decision D, namely, the path that goes through neural state R.

This alternative pathway is no help for Fischer and Ravizza's case, however, since a version of (T1*) and (T2*) could obviously be constructed for the alternative pathway as well.¹³

Now suppose that, although the world is causally determined in the normal course of things, there is a God who can override the laws of nature. Then God can bring it about directly, just by willing it, that S is in neural state N and therefore that S makes decision D. In this case, apart from miracles, causal determinism is true, and yet there is another path to the outcome of S's making decision D. Furthermore, in this case, it is not possible to construct a version of (T1*) and (T2*) for the alternative pathway. In addition, the alternative pathway is one for which someone -- namely, God -- is responsible. The problem with this case, of course, is that the alternative path isn't one for which *S himself* is responsible. On the contrary, since agents who can avail themselves of this alternative pathway have to be able to abrogate the laws of nature, this alternative pathway isn't open to human beings.

In fact, if causal determinism is true and if Cartesian dualism is false, then any path to the outcome of S's making decision D which is a path involving *human* agency will be a path for which some version of (T1*) and (T2*) can be constructed. Suppose, for example, that S arrives at decision D by considering different evidence, accepting different reasons, and basing his deliberations on different considerations from those he did in the original case. Any mental route S takes to reach his

¹³ . As Carl Ginet has pointed out to me in correspondence, there is also an additional problem about whether such alternative pathways are actually available for most decisions.

decision, no matter what it is, will be a route which is correlated with some neural state N^* , on the presupposition that Cartesian dualism is false; and this neural state will be causally determined, on the presupposition that causal determinism is true. Consequently, some analogue of (T1*) and (T2*) will also always be true.

It is clear therefore that any alternative pathway to the outcome of S's making decision D will run into the same trouble. If causal determinism is true (and Cartesian dualism is false), then for any such alternative pathway to a decision, it will be true that

(T1*a) the laws of nature and the conditions at the time of the Big Bang causally determine that S is in neural state N^* at time t^* .

And to make this alternative pathway analogous to the Fischer and Ravizza counterexamples such as *Erosion*, one would have to go on to deny the truth of

(T1*b) no human being is or ever was morally responsible for its being the case that the laws of nature and the conditions at the time of the Big Bang causally determine that S is in neural state N^* at time t^* .

But, on the presupposition of causal determinism, (T1*b) is so clearly true that even the most ideologically committed compatibilist couldn't seriously deny it. To deny it would be to hold that in a causally determined world some human being S is responsible for the causal determination of S's neural states by natural laws and conditions at the time of the Big Bang.

Consequently, insofar as any alternative path goes through material objects (as it must if human beings are not taken to be immaterial substances) and material objects are taken to be causally determined (as they must be if causal determinism is true), there is no alternative pathway for which any human being is or ever was responsible.

Therefore, the Fischer and Ravizza strategy fails. Their counterexamples to Transfer NR all require a certain feature for their success -- namely, an alternative pathway in which someone is clearly responsible -- which isn't in the cases involving moral responsibility and causal determinism.¹⁴ Although Fischer and Ravizza are right that

¹⁴ . Of course, this objection to Fischer and Ravizza's attempted refutation of the direct argument for the incompatibility of moral responsibility and causal determinism would fail if Cartesian dualism were true (and if 'causal determinism' is taken to mean just causal determinism of the physical). There is obviously something at least mildly disconcerting about making a *refutation* of an argument for incompatibilism rest on an *acceptance* of Cartesian dualism. But the important thing to see is that, though accepting Cartesian dualism would undermine this objection to Fischer and Ravizza's case, it would not bolster their attempted refutation of the direct argument against compatibilism. If human beings are immaterial

Transfer NR isn't valid, the cases in which it fails don't include cases of decisions which are causally determined. If we restrict the scope of Transfer NR so that it doesn't apply to two-path cases of the *Erosion* sort, cases of causally determined decisions are not outside its scope.

Consequently, Fischer and Ravizza have not succeeded in showing that the direct argument for incompatibilism fails. To that extent, their claim that the account of moral responsibility which they give is compatible with causal determinism is undermined, whatever the other virtues of their powerful and interesting account of moral responsibility may be.¹⁵

substances, then they are not causally determined by initial physical conditions and the physical laws of nature. But if we could build a counterexample to Transfer NR only by assuming that human beings are not causally determined, the result would hardly trouble incompatibilists since the argument employing Transfer NR is intended to show that a *causally determined* agent is not morally responsible.

¹⁵ . I am grateful to William Alston, Carl Ginet, Peter Graham, Scott MacDonald, Michael McKenna, and Alfred Mele for helpful comments on an earlier draft. This paper was done at the National Humanities Center, where I was a Lilly Foundation Fellow for the academic year 1999-2000. It is hard to imagine a more pleasant environment in which to work, or one more conducive to research, than the National Humanities Center, and I am grateful to the Center and to the Lilly Foundation for my year there.