Aporia vol. 16 no. 2—2006

Why the Direct Argument for Incompatibilism is not Actually Direct

STEPHEN MAHAFFEY

In recent years proponents of the view that causal determinism and moral responsibility are incompatible have increasingly employed a so-called "Direct Argument" in support of their position. This argument bears this name because it contrasts with other incompatibilist arguments that begin with an attempt to show that causal determinism is incompatible with any person having the freedom to act otherwise than they actually do. These arguments can be deemed 'indirect' because, in order to yield the incompatibilist conclusion, they require additional argumentation to establish the truth of the Principle of Alternate Possibilities (PAP). This means that indirect incompatibilist arguments only work if it can be demonstrated that people must have the freedom to act otherwise than they actually do in order to be judged morally responsible for their performed actions. This requirement is a concern for incompatibilists because Frankfurt-type objections to PAP have effectively shown that the truth of this principle is problematic. The Direct Argument, on the other hand, is designed to be a proof of the incompatibilist thesis that does not rely upon contentious arguments about what moral responsibility requires. This paper will question whether the Direct Argument can establish the incompatibilist thesis in the way it is claimed to. To do this I will first examine John Martin Fischer and Mark Ravizza's criticism of the Direct Argument and detail the potent incompatibilist response to it. I will then argue that one can craft a counterexample to the Direct Argument that is impervious to the response directed at Fischer and Ravizza. Using this result, I will argue that the counterexample can only be answered by incompatibilists if it can be established that

Stephen Mahaffey is a student at the University of Calgary studying philosophy. He plans to pursue a PhD in moral philosophy upon graduation..

PAP is true. Hence, I will attempt to show that the Direct Argument is no less 'indirect' than the arguments which normally bear that description; a result which entails that the Direct Argument is highly vulnerable to Frankfurt-type objections to PAP.

It is, of course, necessary to begin this essay by laying out the Direct Argument. This argument relies upon a proposed-to-be-valid inference principle which is very similar to some employed by incompatibilists in the aforementioned 'indirect' arguments. This principle is called "Transfer NR" and can be expressed as follows:

> i) If p obtains, and no one is even partly morally responsible for p; and
>
> ii) if p obtains, then q obtains, and no one is even partly morally responsible for the fact that if p then q; then
>
> iii) q obtains, and no one is even partly morally responsible for the fact that q obtains. (Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* 152)

Proponents of the Direct Argument simply suppose that causal determinism is true and utilise Transfer NR1 to formulate their argument in the way detailed below. For purposes of brevity I shall use the expression "NR" as an abbreviation aid so that when this term is applied to a true proposition, p, describing some occurrence at a time, t, the completed statement means:

NR(p): p, and no person is even partly morally responsible for the fact that p.

The Direct Argument for incompatibilism can now be portrayed in three steps (with E referring to any action/event at some time t):

> 1) NR(The distant past and the laws of nature obtain).
>
> 2) NR(If the distant past and laws of nature obtain, then E occurs).
>
> 3) Therefore, NR(E occurs), (by application of Transfer NR to 2 and 3).[1]

Incompatibilists claim that this argument is the uninterrupted, self-contained proof that no one is morally responsible for any action if causal determinism is true. Clearly, the argument's essential thrust is the claim that Transfer NR is a valid rule of inference, for (3) follows from (1) and (2) purely on the grounds that this principle is valid. The second thrust of the

[1]I am deeply indebted to Dr. Ishtiyaque H. Haji for his help with formalizing many of the arguments mentioned in this essay.

argument is simply the claim that (1) and (2) are true for any event[2] if causal determinism is true, which entails that (3) is true for any event if causal determinism is true. It is in this manner that the Direct Argument argues for incompatibilism, for if the conclusion, "Therefore, NR(E occurs)" is true for every event, then incompatibilism is undeniable. Quite appropriately, there are two ways in which the argument can be criticized. One can try to show, in a bid to exclude Transfer NR from being utilised by the argument, that one of (1) or (2) (or both) can be false when causal determinism is assumed to be true, or one can attempt to show that Transfer NR is not a valid rule of inference.

However, the first way in which the argument can be attacked is entirely fruitless. To see that this is so, it is useful to note that Causal determinism can be roughly expressed as the theory that the state of the world at some time t entails, in conjunction with the laws of nature, all future events after t. This means that any current event is understood as being a causal entailment of the state of the world in the distant past and the laws of nature. Thus, the conditional contained in (2) must be true if causal determinism is assumed to be true because it is just an expression of the assumption that causal determinism is true. Therefore, since it cannot be plausibly denied that no one (at least no person) could be morally responsible for the truth of causal determinism, the complete premise (2) appears unimpeachable. For similar reasons, the proposition contained in (1) (the sentence sans NR qualifier) also cannot be false if causal determinism is assumed to be true, as the theory of causal determinism assumes that there are long-past events (which occurred before the dawn of humanity) and laws of nature. Consequently, since clearly no one (at least no meek and mortal being) can be morally responsible for the fact that past events and the laws of nature are as they are, (1) also seems to be unquestionable. If one draws these thoughts together, what is clear is that both (1) and (2) are beyond reproach.

This then is the challenge for compatibilist theorists: can it be shown that Transfer NR is invalid? One attempt to defeat the Direct Argument by meeting this challenge was made by John Martin Fischer and Mark Ravizza.

[2] In this paper I will generally use "event" synecdochically and so refer with it to events, actions, consequences, omissions etc. In other words, I shall use "event" to refer to all those occurrences in the world that agents can be morally responsible for.

They propose a case, called "Erosion", where a secret agent, living in a world where causally undetermined choice is possible, is trying to destroy an enemy base nestled at the base of a huge glaciated mountain face. To perform her mission the agent places explosives in the glacier so that when she detonates them at time t1 an avalanche occurs and flattens the enemy base at time t3. However, the glacier is naturally eroding and had she not detonated the explosives at t1, a natural avalanche would have occurred at time t2 and demolished the enemy base at t3. Using this scenario, Fischer and Ravizza craft the following two premises (modified with the abbreviation "NR" for sake of conciseness again):

1a) NR(The glacier is eroding).

2a) NR(If the glacier is eroding, then there is an avalanche that crushes the enemy base at t3). (Eleonore Stump, "The Direct Argument for Incompatibilism" 460)[3]

Fischer and Ravizza argue that these two premises are correct given the scenario in play. The first premise is beyond reproach and the second premise expresses a true causal relation for which it is clear that no one is morally responsible. Given the truth of these two premises, it seems that Transfer NR would entail the conclusion:

3a)Therefore, NR(There is an avalanche that crushes the enemy base at t3).

However, the argument is that this conclusion is clearly false because the secret agent chose to detonate the explosives at t1 and so it is only plausible to think that she is morally responsible for the camp's annihilation. Consequently, it seems as if Transfer NR cannot be a valid rule of inference because it yields this obviously false conclusion from true premises. As a result, the direct argument for incompatibilism appears to have been defeated by Fischer and Ravizza's example.

There is, however, there is an important difference with the case proposed by Fischer and Ravizza. This difference is that their example is a case of overdetermination where two entirely independent pathways[4] are concurrently working to bring about the occurrence of the same event. To put it differently, in Erosion there is overdetermination of the avalanche

[3] This article will be cited by the author's last name followed by the page number.

[4] By "pathway" or "path" I simply mean a way or manner in which an event can come to pass. Thus, I use the words to refer to any process, procedure, chain of events etc. that causes or leads to some resultant event.

because at t1 the secret agent was causing the slide while the natural avalanche causing process of the glacier's movement was in progress. Thus, the secret agent's actions and the glacier's natural movement were simultaneously operating to cause the avalanche. In addition, cases like Erosion are distinguished by the fact that the overdetermination is unique because one path is such that examination of it in isolation would indicate to one that no agent is morally responsible for the resultant event, while the other path is such that after similar examination in isolation it seems only plausible to suppose that some agent is morally responsible for the same resultant event (Fischer and Ravizza, "Replies" 447). In such cases (called "two-path" cases by Fischer and Ravizza) the general strategy being employed against the Direct Argument is to express the path which indicates that no one is morally responsible for the result in the premises (1a and 2a) of the argument, while leaving the path which indicates that someone seems clearly to be morally responsible for the result unmentioned. Given transfer NR, the conclusive event which culminates the path expressed in the argument should be such that no one is morally responsible for it, but because this event is concurrently or pre-emptively overdetermined by the unmentioned path, it seems that the only plausible thing to say is that it is true that someone is morally responsible for this event. In contrast, one-path cases are those cases that do not have two paths overdetermining an event such that only one path indicates that someone is morally responsible for the event. To put it succinctly, in one-path cases there is no overdetermination of the very specific sort imagined by Fischer and Ravizza.

The fact that Fischer and Ravizza's apparent counterexample to Transfer NR is a two-path case is important because Eleonore Stump has voiced a very interesting and effective defence of the Direct Argument which hinges upon this feature of Erosion (and its counterparts). To put it in schematic form, her argument is that these examples fail to defeat the direct argument because all cases when causal determinism is assumed to be true will be one-path cases (Stump 465–466). Hence, while Erosion is an effective refutation of Transfer NR, this inference principle can be easily modified to yield a new version that only applies to one-path cases. This modification, she believes, will exclude all cases exhibiting the particular overdetermination Fischer and Ravizza imagine, while still driving the

Direct Argument by including all cases where causal determinism is assumed to be true. Her argument relies upon the idea that, in a causally determined world, all paths leading to an event can be placed in a causal chain leading back to occurrences in the distant past (Stump 465). Thus, all pathways and their culminating events should be understood as causally issuing from a conjunction of past events and natural laws. The conclusion which Stump claims to follow from this realization is that there is no event about which one can think that someone is clearly morally responsible for its occurrence. This is because to assume that someone is morally responsible for an event in a causally determined universe is to beg the question against incompatibilists. As a result, the special sort of overdetermination present in two-path cases cannot be thought to exist when causal determinism is assumed to be true, for one of the pathways required for such two-path overdetermination cannot be supposed to be present without begging the question against incompatibilists. Indeed, since all paths will bear the mark of causal determination, there can only be one-path cases when causal determinism is assumed to be true.

Stump maintains that all this argumentation results in the realization that two-path counterexamples to Transfer NR are irrelevant to the Direct Argument because they do not address cases in which causal determinism is assumed to be true. Such counterexamples may invalidate Transfer NR, but it is easy to create a restricted form of Transfer NR that only applies to one-path cases (Stump 466). Using this restricted inference principle the Direct Argument can still run because cases where causal-determinism is true are one-path cases. This new and restricted principle is known as Transfer NR1 and is as follows:

> (i) If p obtains, and no one is even partly morally responsible for p; and
> (ii) if p obtains, then q obtains, and no one is even partly morally responsible for the fact that if p obtains, then q obtains; and
> (iii) if the pathway implicated in (ii) is the one pathway (in the rele vant sense) of q's obtaining; then
> (iv) q obtains, and no one is even partly morally responsible for this fact.

This new principle is employed in exactly the same way the original one was to yield the incompatibilist conclusion. The only difference is that this time two-path counterexamples are entirely spurious. Consequently, it seems as if the Direct Argument still stands as a very simple way of proving incompatibilism.

To defeat the Direct Argument then, one must invalidate Transfer NR1 by providing a one-path counterexample to it. Happily for compatibilists, Transfer NR1 seems to suggest that an invalidating example can be devised. To demonstrate this, imagine a world where casual determinism is false and people have the ability to make undetermined choices. Furthermore, consider a scenario where an outwardly ordinary man, Bob, who possesses a choice faculty like the one described, has had the misfortune to be unknowingly exposed to a very strange form of cosmic radiation. This radiation has, unbeknownst to him, caused his brain to develop a very bizarre abnormality. If lightning strikes within 200m of Bob at some time t1, then the electrical field can fully 'mature' the abnormality and so begin a causal process in Bob's brain that turns him into a choiceless automaton and causes him to murder the next person he comes across at some time tn. In fact, let us suppose that the lightning strike will 'mature' the abnormality, and thereby turn Bob into a murdering automaton, unless, at t1, Bob chooses to murder the next person he meets at tn. In other words, due to some peculiar aspect of Bob's brain abnormality, his choice of this murderous action at the moment of the lightning strike results in a brain state that somehow nullifies the lightning's effect upon the abnormality and prevents this malformation from initiating the causal process that turns him into murdering automaton. Let us also suppose that there are no subsequent lightning strikes after t1 that could affect his brain abnormality. Let us now make this situation of Bob's even more sinister and suppose that, having made such a choice at t1, if he later recants his decision and decides not to murder the next person he encounters, then the brain state corresponding to this undetermined recantation somehow reactivates and 'matures' his brain abnormality so that a causal process that will force him to commit a murderous action at tn is once more initiated.

Unfortunately for Bob, we must make the case even more twisted and insidious by supposing that when Bob was a child, a rogue neurologist

identified him as being at great risk to one day decide to murder the next person he met. In response to this finding, the neurologist covertly installed a small nano-device in Bob's brain (unbeknownst to him of course) which prevented him from choosing such a murderous action. Now, imagine that this device is effective but has one major flaw. The device is very sensitive to strong positive charges, just like the very strong positive charge exhibited by the area of a lightning strike moments before the actual bolt descends. Let us imagine that this device is so vulnerable to positive charge that if lightning were to strike a point at some time t, this device would succumb to the strong positive charge present at t-1 if the device were within 200m of the lightning's strike point. This malfunctioning would thereby leave the device's host free to choose a homicidal act at the next moment in time; which happens to be t. Thus, at a time t (if we assume that Bob has never been close to a lightning strike before) Bob is able to choose to murder the next person he sees only if lightning strikes within 200m of him at t.

The summation of this scenario is that Bob has no alternative but to murder the next person he encounters if lightning strikes within 200m of him. Interestingly though, this murderous action can be the result of either his undetermined choice or an unavoidable causal process in his brain which is itself caused by the lightning and his abnormality. Using this imagined scenario we can create a very interesting case. In this case suppose that lightning strikes within 200m of Bob for the first time at t1. We shall also imagine is that at t1, Bob makes an undetermined choice to murder the next person he comes across at some time tn; a decision he never recants. With these suppositions and the aforementioned scenario, the following argument can be crafted (with the term "NR1" performing the same role as "NR" did in the previous arguments, except this new term expresses that Transfer NR1, not Transfer NR, is in use):

1b) NR1(Lightning strikes within 200m of Bob at t1).

2b) NR1(If lightning strikes within 200m of Bob at t1, then Bob murders the next person he comes across at tn).

If Transfer NR1 is valid, the following conclusion would seem to issue from these premises:

3b) Therefore, NR1(Bob murders the next person he comes across at time tn).

However, this conclusion is false. Bob made an undetermined choice at t1 to murder the next person he came across; so it is only plausible that Bob is morally responsible for his murderous action. Thus, Transfer NR1 is invalid because it yields this wrong conclusion from true premises. However, one must consider the possible objections to this case before declaring a victory over the Direct Argument. If the premises of the case can be proven false or the entire case shown to be irrelevant[5], then the case of Bob would fail to block incompatibilists.

To begin this discussion of possible responses to this case, the truth of the two premises is difficult to dispute. Premise (1b) is undeniably true because the lightning strikes and it is clear that no one is morally responsible for the lightning strike. Bob might well be responsible for building his house in its location and for sitting down to read where he does, but he is certainly not responsible for the lightning striking within 200m of him. The truth of premise (2b) is perhaps less straightforward, but is no less certain. When the lightning strikes at t1, Bob cannot avoid murdering (or attempting to murder) someone at tn. It is true that he can freely choose to murder or be forced as a mindless automaton to murder the next person encounters, but he cannot avoid this action once lightning strikes within 200m of him. This is because if, as the lightning strikes, he does not freely choose his murderous action, then he is caused to perform it by the lightning's influence on his brain abnormality. The only way he can avoid being forced to perform such an action is by choosing to perform this same murderous action. Therefore, the conditional contained in (2b) is true because it is an expression of the fact that Bob has no alternative to murdering once the lightning strikes. With this point established, one can also see that no one is morally responsible for the truth of the conditional contained in (2b). To be morally responsible for (2b) a person would have to be responsible for the fact that Bob must murder the next person he comes across after the lightning strikes. This is impossible, though, as Bob cannot but perform this homicidal action because of his brain abnormality. Therefore, the truth of (2b) cannot be denied and, as a result, it is not possible to reject the case of Bob by denying the truth of its premises.

[5] If one could demonstrate that the conclusion yielded by Transfer NR1* in the case of Bob is true, then one would surely have defended the Direct Argument. However, I see no way in which this could be done without begging the question by taking the quite unjustified view that people cannot be morally responsible for their actions even if these actions are not causally determined.

The next potential objection is the argument that this case is a spurious two-path case just like Erosion. If this is so, then clearly the case of Bob is just another failed attempt to defeat the direct argument which only manages to defeat Transfer NR. Bob's case, however, is not a two-path case. There is no overdetermination of Bob's murderous action like there was of the camp's destruction in Erosion. Admittedly, there seems to be two possible pathways that lead to his murderous action, for Bob can choose this action or be forced by his brain abnormality to perform it. Moreover, one path indicates that no one is morally responsible for the resultant event; while the other path indicates that someone is morally responsible for the result. However, in the case of Bob these two pathways cannot be concurrently active in the way two paths are assumed to be in Erosion. Remember that if Bob freely chooses at t1 to murder, then the causal process issuing from his brain abnormality never begins and he is in no way caused or forced to perform the action. Similarly, if his brain abnormality causes him to murder, then he does not choose this action. In other words, if he does not freely choose his action at t1, then the causal process at work in his brain renders him an automaton unable to make any choices after this time. What should also be clear is that, before t1, the nano-device in Bob's brain simply prevented him from ever choosing to murder the next person he encountered. Thus, these two pathways cannot concurrently be in effect and only one actually leads to Bob's action. In any imagined scenario then, there can be no overdetermination of Bob's action.

The absence of overdetermination in this case may be enough to differentiate it from a two-path case. However, it might still be argued that this scenario is not a one-path case because while these two pathways cannot concurrently operate, they are nonetheless two bona-fide paths which can lead to Bob's murderous action and one pathway indicates that Bob is morally responsible for his action. Consequently, the case of Bob is a two-path case regardless of whether these paths concurrently operate. As a second objection, critics might also question the relevancy of this case because neither of the two pathways is mentioned in premise (2b); which appears to be a direct breach of condition (iii) of Transfer NR1. Both of these counter-arguments, however, misunderstand the nature of Bob's case. Both objections are faulty because they fail to realize that the lightning strike at

t1 is ultimately what necessitates Bob's murderous action. Indeed, the circumstances of the case are such that Bob could not have chosen to murder the next person he encountered at any time before the lightning strike allowed him to make such a choice at t1. To respond to the first objection, when recognition of the primacy of the lightning strike is combined with the already demonstrated fact that there is no overdetermination in this case, it becomes clear that only one pathway is present in the case of Bob's homicidal act. This is because any pathway leading to his action must begin with the lightning strike and the two seeming pathways cannot both begin with such because they cannot be active concurrently. Hence, the lightning strike either makes it possible for him to choose this action and he does so choose, or the lightning causes his action via his brain abnormality. Thus, there are not two pathways present in the case of Bob at all, just one pathway originating with the lightning strike that can take one of two possible forms, depending upon Bob's undetermined choices at t1. The illusory thought Bob's case is a two-path case springs from the fact that the single pathway which leads to Bob attempting homicide can simply be different (having two different routes of expression) in cases with different circumstances. Therefore, what this arguments shows is that the case of Bob is actually a one-path case and it cannot be objected to using Stump's arguments against two-path cases.

The second objection was that the case is deficient because neither of the two pathways which lead to Bob's action are expressed in premise (2b). However, the argument above established that these two pathways were simply an illusion and that only one pathway can lead to Bob's attempted murderous act. Moreover, this single pathway issues from or begins with the lightning strike at t1, for the lightning strike is what necessitates Bob's action. Consequently, premise (2b) of the argument does express the pathway leading to Bob's act when it states that if lightning strikes within 200m of Bob at some time, then Bob will murder the next person he encounters. This is because, while the details of the pathway are not fixed, it is nevertheless undeniable that the lightning strike is ultimately the source or beginning of the pathway leading to the act in question. As a result, (2b) does sufficiently express the pathway at work in the case of Bob, which means that this case does comply with condition (iii) of Transfer NR1 and is a relevant counterexample to this principle.

The case of Bob may be resistant to criticism so far, but there is one very interesting and potentially problematic feature of this scenario. This feature is that the lightning strike at t1 ensures that Bob has no freedom to perform an act other than murdering the next person he comes across at some time tn. He truly has no alternative possibilities once the lightning strikes within 200m of him. Clearly then, if the Principle of Alternate Possibilities (PAP) is true, then it is obvious that Bob is not morally responsible for his murderous action. Hence, a new inference principle, Transfer NR1p, which includes the assumption that PAP is true, could be proposed. If valid, this principle would still drive the Direct Argument, while also entailing that Bob is not morally responsible for his action. In other words, it could be argued that this case is a totally ineffective response to the direct argument because it addresses an unnecessary inference principle. Of course, this possible method of responding to cases like Bob's would only work if PAP was unimpeachably true. Thus, it seems that the question of whether the Direct Argument can be successful depends on whether or not PAP is true. Unluckily for incompatibilists, this question is by no means answered in their favour as Frankfurt-type examples form a collection of potent and not successfully refuted arguments against PAP. Therefore, incompatibilists must somehow demonstrate that PAP is true before they can hope to think that the direct argument is successful. Until such is proven, the Direct Argument is as vulnerable to Frankfurt-type examples as the previously mentioned 'indirect' incompatibilist arguments are. With this fact in mind, if one considers why the Direct Argument has become popular, then it becomes clear that it is already a failure. This is because it was originally designed to prove the incompatibilist thesis without requiring additional argumentation in support of the thesis that moral responsibility requires alternate possibilities. What I have argued is that the direct argument does not actually escape such concerns, for only through positive argumentation for PAP can potent counterexamples like the case of Bob be nullified. This should not really be a surprise. It was always strange to think that an argument could establish the incompatibility of causal determinism and moral responsibility without attempting to deal with the question of what is required for a person to be morally responsible for an action she performs.

Works Cited

Fischer, John Martin and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, 1998

—— "Replies." *Philosophy and Phenomenological Research* 61:2 (2000): 467–480.

Stump, Eleonore. "The Direct Argument for Incompatibilism." *Philosophy and Phenomenological Research* 61:2 (2000): 459–466