

Free Will and Moral Responsibility: A Reply to Fischer

I. Introduction

At the center of much contemporary debate regarding the metaphysics of free will and moral responsibility is a simple, intuitively plausible principle known as the “Principle of Alternative Possibilities” (or PAP, for shorthand). PAP embodies the common intuition that in order for us to be held morally responsible for a choice or act of will, we must be able to choose or act otherwise. So, according to PAP, the presence of alternative possibilities—the freedom to will or choose otherwise—is a *necessary condition* for moral responsibility. If a person doesn’t have alternative possibilities open to her at the time of choice or at some relevant time earlier, then she is not morally responsible for her choice. Before I present what I want to argue for, I think it would be helpful to quickly survey how the debate surrounding PAP has progressed in recent years.

Beginning with Harry Frankfurt’s seminal paper in 1969, many philosophers have sought to offer counterexamples to PAP and thus refute it. Named and modeled after their founding father, such counterexamples are often referred to as Frankfurt-style examples, and purport to be cases in which it is true both that (1) a person is morally responsible for a choice or act of will and (2) that person *lacks* alternative possibilities. Thus, Frankfurt-style examples purport to show that PAP is false – the presence of alternative possibilities is *not* a necessary condition for moral responsibility. I will present a concrete Frankfurt-style example later on.

Not surprisingly, this isn’t the end of the story. Recently, many defenders of PAP (whom I shall call *PAPists*) have taken a closer look at the metaphysical assumptions and ambiguities in typical Frankfurt-style examples. Perhaps, most importantly, PAPists have noted that typical Frankfurt-style examples do not make *explicit* whether causal determinism obtains or not.¹ What is causal determinism? Roughly, causal determinism is the thesis that the facts of the past, in conjunction with the laws of nature, entail every truth about the future. For ease of reference, I shall be referring to this thesis simply as *determinism*, though it is important to keep in mind that there are a variety of supposed determinisms (i.e. not just causal, but also theological, logical, etc). So, if determinism (as I have defined it) is true, then, given the actual past, and holding fixed the laws of nature, only one present and future is physically possible at any moment of

¹ I have in mind the likes of David Widerker (1995), Carl Ginet (1996), and Robert Kane (1996).

time. Notice that an implication of determinism is that there are (causal) conditions of a person's internal and external actions located in the remote past, prior to her birth, and that these conditions are sufficient for each of her actions (McKenna).

With this said, the main PAPist challenge to Frankfurt-style examples can be stated. It is useful to present the challenge in the form of the following *dilemma*:

- (A) If determinism is assumed, then Frankfurt-style examples *assume what they are trying to prove* (or *beg the question*) since the very point at issue is the relationship between determinism, alternative possibilities, and moral responsibility;² however,
- (B) If determinism is not assumed, then it is *not* the case that no alternative possibilities are open to the agent.

So, according to the PAPist, either the agent is *not* morally responsible (if determinism obtains) or the agent *does* have alternative possibilities (if determinism does not obtain). Either way, PAP is not refuted. Since the PAPist is unaware of any case in which it is true that the agent has no alternative possibilities and is also morally responsible, Frankfurt-style examples fail to cast doubt upon the original intuition underlying PAP.

Of course, this has not been the last word either. In particular, John Martin Fischer (1999, 2004) has recently offered an important response to both “horns” of the dilemma. He remains convinced that Frankfurt-style examples “help to establish that it is very plausible that moral responsibility does not require alternative possibilities” and he argues for the compatibility of moral responsibility and determinism – a position he calls *semi-compatibilism* (1999, 112). In this paper I will focus on critically evaluating his reply to the first horn of the dilemma and seek to show that he does *not* manage to provide a satisfactory answer – more specifically, I think the first step in his argument still begs the question against the PAPist.³ So the conclusion will be that Fischer's response to the PAPists' dilemma is (at least) partially unsatisfactory.

II. A Test Case and the Dilemma

² The charge of “begging the question” can often be misleading and is used in different ways by philosophers. Perhaps my objection can be better understood as charging Fischer with not using theoretically neutral premises in his argument (i.e. not using premises that would be acceptable to all major parties involved). That is, I think his reply to the first horn involves a “theoretically loaded” premise.

³ Although I believe Fischer's reply to the second horn is also unsatisfactory, I will leave it to the side for the purposes of this paper.

The general structure of Frankfurt-style examples usually consists in there being a “failsafe mechanism” that (a) makes it the case that the agent has no alternative possibilities and yet (b) plays no *actual* role in the agent’s chain of action (i.e. deliberations, choices, behavior). So-called “prior-sign” Frankfurt-style examples include some involuntary sign or indication of the agent’s future intentions, choices, or actions, and it is on the basis of this sign that the mechanism intervenes. It will be helpful to have a concrete example explained. Since I am aiming to criticize Fischer, I will use his own “favorite” Frankfurt-style example, slightly modified:

Imagine a person named Jones is in a voting booth deliberating about whether to vote for Al Gore or George W. Bush. Unbeknownst to him, Black, a liberal neurosurgeon, has implanted a device in Jones’ brain which monitors Jones’ brain activity. If Jones is about to choose to vote for Gore, the device does not intervene in the process in any way. If, however, Jones is about to choose to vote for Bush, the device intervenes by electronically stimulating the brain in a way that is sufficient to produce a choice to vote for Gore and, thus, a subsequent vote for Gore. How does the device tell whether Jones is about to choose to vote for Bush or Gore? Well, suppose that...

- (A) If Jones is about to choose at T2 to vote for Gore at T3, he shows some involuntary sign (N), such as a neurological pattern in his brain, at T1
- (B) If Jones is about to choose at T2 to vote for Bush at T3, he shows some involuntary sign (N*), such as a neurological pattern in his brain, at T1
- (C) If (B), Black’s device intervenes; but if (A), Black’s device does not intervene

The story continues: After serious reflection, Jones chooses to vote for Gore and, since the device does *not* intervene, Jones does vote for Gore. Given that the device plays no role in Jones’ deliberations and act of voting, it seems that Jones acts freely and is morally responsible for voting for Gore. Given the presence of Black’s device, it is plausible to think that Jones does not have alternative possibilities with regard to his choice and action since if he had tried to vote for Bush, he would not have been able. Thus, it is plausible to think that PAP is false – alternative possibilities are *not* necessary for moral responsibility (1999, 109-110).

This is a typical Frankfurt-style example. At first glance, such an example does indeed seem to cast serious doubt on PAP. How is the PAPist to respond? Well, it is precisely in response to examples such as the one just described that the PAPist poses the dilemma mentioned earlier. As you may recall, the PAPist will complain the Fischer’s example (like most other Frankfurt-style examples) is *under-described* – it is not clear whether determinism is supposed to obtain or not.

Well, consider the options again. First, suppose determinism obtains. Then the PAPist would argue that such an assumption is illicit because it begs the question against the PAPist by assuming what it is trying to prove. Represented schematically [HANDOUT]:

The point at issue is the following “natural intuition” (PAP):

1. Moral responsibility requires alternative possibilities (If **A**, then **B**)

An assumption (for both Fischer and the PAPist) is:

2. Alternative possibilities are *incompatible* with determinism (If **B**, then *not-C*)

It follows that:

3. Moral responsibility is *incompatible* with determinism (If **A**, then *not-C*)

The semi-compatibilist (i.e. Fischer) conclusion is supposed to be:

4. Moral responsibility is *compatible* with determinism *not*-(If **A**, then *not-C*)

So assuming that determinism is true, it follows that the relationship between the prior-sign (i.e. the neural pattern) at T1 and the decision or choice at T2 is also causally determined. That is, the state of world at T1 and the laws of nature are causally sufficient to bring about Jones’ choice at T2. But, given the argument that determinism is incompatible with alternative possibilities, it follows that Jones does *not* have alternative possibilities (i.e. Jones has no choice but to choose the action he does). So one cannot assume determinism and also claim that Jones is obviously morally responsible in Fischer’s example. The PAPist would surely reject such an assertion and no ground would be won by person who denies PAP.

Now let’s turn to the other option and suppose determinism does *not* obtain, but rather causal *indeterminism* is true. The problem here is that if one assumes that causal *indeterminism* obtains, that means that no determinate relationship exists between the prior sign at T1 and the agent’s choice at T2. Thus, the agent would have alternative possibilities open to her at or just prior to T2. That is, given the state of affairs at T1 and the laws of nature, it would *not* follow that Jones chooses to vote for Gore at T2. The prior sign would be (at most) a *reliable* indicator – but since it is not “failsafe”, it follows that Jones’ choice is *not* unavoidable. So the PAPist concludes that if determinism obtains, then Jones is *not* morally responsible, and if determinism does *not* obtain, then alternative possibilities (i.e. choices) are open to Jones. Thus, the scenario fails to be a case in which it is true both that Jones is morally responsible and Jones has not

alternative possibilities open to him. The Principle of Alternative Possibilities remains unscathed and the PAPist untroubled.

III. Fischer's Response (Contra the First Horn)

Fischer, of course, recognizes the force of the question-begging charge.⁴ He agrees that, on the assumption of determinism, one cannot “simply and precipitously conclude, from consideration of the examples, that the agent is morally responsible” (1999, 113). However, Fischer offers another, slightly more complex argument to the conclusion that Jones *is* morally responsible in Frankfurt-style examples even when determinism *is* assumed. It will be helpful to have a short-hand label for Frankfurt-style examples in which determinism is assumed to obtain. Let us call them “CD-examples” (CD for “causal determinism”). So Fischer argues that after reflecting on CD-examples, one should conclude that the lack of alternative possibilities does not “in itself” ground a claim that the agent is not morally responsible for his choice and action. More formally, Fischer's first step is:

1. If Jones is not morally responsible for his choice and action, then it is not simply because he lacks alternative possibilities.

Fischer supports this step by noting that, after all, “everything that has any causal (or any other kind of) influence on Jones would be exactly the same, if we ‘subtracted’ Black entirely from the scene. And Jones's moral responsibility would seem to be supervenient on what has an influence or impact on him in some way” (1999, 113). He does not think his first step begs the question against the PAPist, even if determinism obtains. In a later article (2004), Fischer makes the first step slightly stronger: based on consideration of CD-examples, “intuitively it is plausible that alternative possibilities [alone] are irrelevant to ascriptions of moral responsibility” (197). The proponent of the Frankfurt-style CD-examples should *not* say Jones is obviously morally responsible – that would beg the question. Rather she should just say, “I don't know at this point whether the agent is morally responsible for his behavior, but *if* he is not, then it is *not* because he lacks alternative possibilities” (197). In this way, Fischer thinks that Frankfurt-type examples serve the important function of “*shifting the debate* away from considerations pertinent to the relationship between determinism and alternative possibilities” (197). Whether or not Jones is

⁴ As Fischer puts it, if determinism is “explicitly presupposed, it does not seem that someone could say that Jones is obviously morally responsible for his actual choice and action in a context in which the relationship between determinism and control (freedom) and moral responsibility are at issue” (1999, 112).

morally responsible now depends on whether determinism in the actual sequence can plausibly be thought to directly rule out moral responsibility, independently of considerations about alternative possibilities. So Fischer's *second* step in his argument is:

2. Determinism in itself and apart from ruling out alternative possibilities does not threaten moral responsibility.

Fischer supports this latter step by noting that of all the various possible reasons why someone might think that determinism *directly* threatens moral responsibility (i.e. apart from ruling out alternative possibilities), he finds none convincing. He admits that he does not have any sort of decisive or "knockdown" argument for such a conclusion, but thinks that since the reasons for thinking that determinism rules out moral responsibility directly are *not* as strong as the reasons for thinking determinism rules out alternative possibilities, genuine progress has been made (2004, 198).

V. Why Fischer Fails

When I first read Fischer's response to the first horn, my reaction was one of confusion. In fact, it still is. His defense of both steps [in (1999) and (2004)] is very brief. My objection is simply that the first step in Fischer's argument does *not* manage to avoid begging the question against the PAPist. As noted previously, Fischer thinks that upon reflection, one should conclude that in CD-examples, the lack of alternative possibilities does not *in itself* ground a claim that the agent is not morally responsible for his choice or act of will. That is, *if* Jones is not morally responsible for his choice and action, then it is not *simply* because he lacks alternative possibilities. The first step seems to be ambiguous and permits of at least two possible interpretations.

First, it may be interpreted as a denial of the claim that alternative possibilities are required (i.e. necessary) for moral responsibility. After all, if the presence of alternative possibilities are a *necessary* condition for moral responsibility, then the *lack* of alternative possibilities is a *sufficient* condition for the lack of moral responsibility. On this interpretation, Fischer is arguing that consideration of CD-examples should lead one to the conclusion that the lack of alternative possibilities is *not* sufficient for the lack of moral responsibility (and so alternative possibilities are not necessary for moral responsibility). This amounts to an outright rejection of PAP. We're supposed to think that, "intuitively it is plausible that alternative

possibilities [alone] are irrelevant to ascriptions of moral responsibility” (2004, 197). We’re supposed to arrive at this conclusion on the basis of “reflection”, but Fischer does not expand on *how* exactly this key step is apprehended. Sure, perhaps a *semi-compatibilist* like Fischer will arrive at the rejection of PAP upon mere “reflection” on Frankfurt-style example that assumes causal determinism, but I highly doubt this mysterious insight will appear to the PAPist (or any other person for whom the CD-examples are supposed to pose problems). Unless the first step can be made more plausible, it would seem to beg the question against the PAPist insofar as it amounts to bare rejection of PAP, which is what the argument is supposed to *prove*, not assume in its premises.

Unfortunately, Fischer provides little support for the first step, but his main reason (which I mentioned previously) is that in the example described above, “everything that has any causal (or any other kind of) influence on Jones would be exactly the same, if we ‘subtracted’ Black entirely from the scene. And Jones’s moral responsibility would seem to be supervenient on what has an influence or impact on him in some way” (1999, 113). Well, I have two things to say. First of all, it is unclear how we are supposed to take Fischer’s last claim – that moral responsibility essentially depends upon on what has an *influence* or *impact* on him in some way. This claim also seems to be an outright rejection of PAP, and it is not argued for. After all, PAP states that the ability to choose otherwise or have alternative possibilities open to you is required for moral responsibility, yet such an ability or presence of alternative possibilities would not seem to be things that have an influence or impact on you. So, again, the PAPist should have deep suspicions about Fischer’s reasoning behind the first step.

Secondly, Fischer tries to support the first step by arguing that “everything that has any causal (or any other kind of) influence on Jones would be exactly the same, if we ‘subtracted’ Black entirely from the scene”. So, if I understand him correctly, Fischer wants us to think about Black’s role in the CD-example and notice that were Black and his device to be removed, nothing in the actual sequence of events would be changed. Our assessment of Jones’ moral responsibility would also likely be unchanged. Since Black’s mechanism plays no *actual* causal role, Fischer wants us to conclude that it’s presence and thereby elimination of alternative possibilities is *irrelevant* to ascriptions of moral responsibility. This point is then taken to support the stronger claim that *any* elimination of alternative possibilities (by itself) is irrelevant to ascriptions of moral responsibility. So even though determinism rules out alternative

possibilities, this fact *by itself* should not have any bearing on whether or not Jones is morally responsible.

What should the PAPist say? Well, I think the PAPist should *agree* that Black (and his device) are entirely irrelevant to Jones' moral responsibility in Fischer's CD-example. But I think that this is true *because determinism obtains*. That is, the PAPist should say that since the past together with the laws of nature logically necessitate the future, Jones was *determined* to choose to vote for Gore *prior* to and *irrespective of* the presence of Black's device. If determinism did *not* obtain in the actual sequence, then (as I noted previously) Black's device would *not* be able to guarantee what Jones' decision would be and so would *not* be a "failsafe" prevention mechanism (that is, Jones *would* have alternative possibilities open to him at the time of choice). But since determinism *is* assumed to obtain, Black's device plays no role at all in thwarting Jones' alternative possibilities. Since the device is *irrelevant* in explaining why Jones lacks alternative possibilities (determinism seems to do all the work *by itself*), *that* is why the mechanism is irrelevant to moral responsibility. Black and his device are just *altogether irrelevant*, if determinism obtains.

The original "natural" intuition at the center of the debate is, in Fischer's words, that "if we have only one option that is genuinely available to us, then we *have* to do what we actually do, and that if we have to do what we actually do, we are *compelled* so to behave. But if we are compelled to behave as we actually do, then surely we cannot legitimately be held morally responsible for what we do" (2004, 190). So, according to this PAP-esque intuition, if determinism is assumed in the actual sequence of events and thereby entails a lack of alternative possibilities, it follows that Jones is *not* morally responsible. To repeat, Fischer wants the PAPist to consider Black's presence and note that, even though his device seems to abolish alternative possibilities, these facts about the alternative sequence are irrelevant to our moral judgment. Fischer thinks that this may then recommend the preliminary conclusion that alternative possibilities *as such* are irrelevant as well. But if, as I maintain, Black's counterfactual intervention is not only irrelevant to whether Jones' is morally responsible but also irrelevant regarding Jones' lack of alternative possibilities, then Fischer's reasoning behind his first step (i.e. his preliminary conclusion) is undermined. Frankfurt-style examples are intended to challenge the PAPist intuition. But a Frankfurt-style example that assumes determinism

accomplishes nothing to this end since all “external” factors to the causal chain (such as Black & Co.) are simply beside the point.

I imagine that someone like Fischer might try to resist this conclusion by arguing that Black’s lack of alternative possibilities is *not* explained by determinism alone. She might point out that in Fischer’s CD-examples, it seems that Jones’ choice is “*over-determined*” by *both* determinism *and* Black’s device (Mele and Robb, 1998). That is, this objector would agree that determinism guarantees the lack of alternative possibilities, but she would argue that Black’s device does *as well*. That is, if determinism did *not* obtain (i.e. counterfactually), then the mechanism *would* intervene and thereby deprive Jones of alternative possibilities. So Black’s device does play *some* role in accounting for Jones’ lack of alternative possibilities – determinism is not alone. She might then argue along similar lines as noted previously that since Black’s device which helps to guarantee a lack of alternative possibilities plays no role in our moral assessment of Jones (i.e. it is irrelevant to our ascriptions of moral responsibility to Jones), *neither* should determinism. In this way, she might continue, it seems that the lack of alternative possibilities is *completely* irrelevant to moral responsibility.

I don’t find a response along these lines very convincing. I think there is a clear difference between the role played by determinism in the actual sequence of events and the role (or lack of a role) played by Black in the alternative or counterfactual sequence of events. The former determines Jones’ actual choices and action *from the inside*, so to speak, and renders all alternative possibilities impossible by itself, while the latter is *external* to the chain of events that constitutes Jones’ action and (as noted previously) *does not* render all alternative possibilities absent by itself. Only determinism ensures that no alternative possibilities are present. Again, I think the PAPist can respond by noting that the *real* reason why Black’s device (that is, counterfactual intervener) is irrelevant to our moral assessment of Jones is because Black’s device, on the assumption of determinism, is irrelevant to Jones’ lack of alternative possibilities. Since a deterministic chain of events entails that it is (causally) *impossible* for something to happen other than how it *does* happen, a counterfactual (i.e. not actual) intervener is rendered unnecessary; that is, contrary to initial appearances, Black’s device does *not* play any additional role in “helping” to bring it about that Jones lacks alternative possibilities.

To put the point another way, Fischer thinks that determinism guarantees the impossibility of alternative possibilities in the actual sequence, while Black guarantees the

impossibility of alternative possibilities in the alternative or counterfactual sequence. Yet, as I've pointed out before, while the former is true, the latter is *false - unless determinism also obtains*. This was the point of the original PAPist dilemma. So it seems that determinism alone suffices to rule out alternative possibilities—and determinism is *precisely* what the PAPist thinks is *incompatible* with moral responsibility. No additional reasons have been presented by Fischer that would lead the PAPist to reconsider her position.

However, perhaps Fischer means something else by his second step. As a reminder, he claims that, upon reflection, CD-examples support the thought that *if* Jones is not moral responsibility for his choice and action, then it is not *simply* because he lacks alternative possibilities. Well, perhaps Fischer just means that alternative possibilities alone are not sufficient to *fully explain* Jones' moral responsibility. If this is what he means by the first step, then the PAPist may very well agree! After all, PAP does *not* imply that the presence of alternative possibilities or choices is a *sufficient* condition for moral responsibility, only a *necessary* one, and no PAPist (that I'm aware of) would argue that the presence of alternative possibilities is the *only* consideration relevant to an agent's moral responsibility. There will be *many* other facts about the particular case that are morally relevant (i.e. relevant to moral ascriptions of praise, blame, etc) – for example, it seems that *mere* alternative possibilities or choices is not enough for moral responsibility, and so the alternative possibilities open to the agent have to be sufficiently robust, under the agent's control, free from compulsion, etc; also, in order to be morally responsible it seems the agent must be aware of the alternative possibilities and their moral consequences, and so forth. The point is just that moral responsibility is a complex, tricky thing and more than just alternative possibilities is usually required to fully explain its presence or absence in a particular case. However, I will not belabor this point, since I do not think this is actually Fischer's objection.

VI. Conclusion⁵

⁵ I would like to note that although the arguments and conclusions of this essay were arrived at independently, many of the same points are argued for in a similar manner by Stewart Goetz (2005) in the first half of his "Frankfurt-Style Counterexamples and Begging the Question", published in *Midwest Studies in Philosophy*, 29. After writing the first draft of this essay for a Winter 2005 Metaphysics course, I continued my research and came across a reference to Goetz' forthcoming paper. Though it had yet to be published, I requested a copy from Dr. Goetz himself and he graciously provided me with one. Despite many differences, the strong overall similarity in terms of argument is stunning, a fact in which I take consolation – it's nice to have a greater mind than yourself thinking the same thoughts!

In this paper I have sought to critically evaluate John Martin Fischer's reply to the first horn of the PAPist dilemma. I have tried to show that he has *not* managed to provide a convincing case for the conclusion that, even when determinism is assumed, alternative possibilities are irrelevant to moral responsibility. So I conclude that Fischer's response to the PAPists' dilemma is (at least) partially unsatisfactory because, contrary to Fischer, I think Frankfurt-style examples that assume the truth of determinism tell us nothing about how alternative possibilities are supposed to be irrelevant to moral responsibility. In other words, no illumination or "insight" appears to be gained by adding Black to the story – Black and his device are *superfluous*, if determinism is assumed. I think most of the plausibility behind Frankfurt's original examples derived from the presumption that although Black took away Jones' alternative possibilities, he did not figure into the *actual* sequence of events and it was presumed that determinism did not obtain (although Frankfurt never explicitly mentioned it). Determinism, however, both rules out Jones' alternative possibilities *and* figures into the actual sequence of events. In this important way, the hypothesis of Black and the obtaining of determinism are very different. Once the assumption of determinism is made explicit, and not just Black, I think the Frankfurt-style examples lose much, if not all, of their intuitive support.

This shows that, under the assumption of determinism, Frankfurt-style examples are not successful in their avowed purpose of casting doubt on the strong "natural" intuition lying being PAP. They lack any sort of persuasive force for the conclusion that alternative possibilities are irrelevant to moral responsibility – attention remains focused (at least partially) on what happens in the *alternative* sequence. So the debate remains exactly where it was *before* the introduction of CD-examples like Fischer's – they have not furthered the debate and, in the absence of further argument, the dialectical stalemate between competing "intuitions" continues.

Bibliography

1. Fischer, John M. The Metaphysics of Free Will. Massachusetts: Blackwell, 1994.
2. Fischer, John M. "Recent Work on Moral Responsibility." Ethics 110 (October 1999): 93-139.
3. Fischer, John M. "Frankfurt-Style Compatibilism." Free Will. 2nd ed. Ed. Gary Watson. New York: Oxford, 2004.

4. Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." Free Will. 2nd ed. Ed. Gary Watson. New York: Oxford, 2004.
5. Kane, Robert. The Significance of Free Will. New York: Oxford, 1996.
6. Kane, Robert, ed. The Oxford Handbook of Free Will. New York: Oxford, 2002.
7. Widerker, David. "Libertarianism and Frankfurt's Attack on the Principle of alternative Possibilities." Free Will. 2nd ed. Ed. Gary Watson. New York: Oxford, 2004.