

XII Free Will II

FRANKFURT

Frankfurt's basic contention is simple: contrary to what we have suggested, it is not true that you are not responsible if you could not have done otherwise. That is, he wants to reject:

The principle of alternate possibilities: A person is morally responsible for their act only if they could have done otherwise.

The principle is a problem for the compatibilist, since, if determinism is true, no one could have done otherwise. Frankfurt aims to show the principle is false by counterexample: by showing that you can be responsible for doing something even though you could not have done otherwise. The basic move is to provide examples that draw apart the two properties.

First example: Coercion. Perhaps PAP gets its appeal from the idea that a person who is coerced is not morally responsible. Suppose that Jones is threatened by Black that he will suffer dire consequences if (and only if?) he does not do something that he has already decided to do. Is he still responsible for doing it?

Three different cases:

- (i) the threat made no difference to him whatsoever.
- (ii) the threat swamped all thought of his previous intention: it made all the difference
- (iii) the threat impressed him, but didn't change what he intended to do.

The case of (iii) seems to be one in which he is coerced and yet is still responsible. So perhaps the principle of coercion is that a coerced person is not responsible for their action if and only if it was the coercion that led them to perform that action. Is this already a counterexample to PAP? The problem here is that arguably a coerced person could still have done otherwise. So change the example. Suppose that Black wants Jones to perform some action that he knows Jones already intends to perform. But Black wants to be sure that he will do it. So he implants a device in Jones's brain. Should Jones change his mind about performing the action, the device will kick in and make him do it. (How will it do this? this will become important later. Perhaps it simply takes control of his body away from him. Perhaps it takes control of his mind away from him, so that he is suddenly overwhelmed by a desire to perform that action.) Suppose that Jones does go ahead and perform the action, without the device being used. Now we seem to have a case in which Jones is responsible for what he does, even though he couldn't have done otherwise.

Reflection on what was said about coercion above might suggest an alternative to PAP, which we can call PAP*:

The principle of alternate possibilities:* A person is not morally responsible for their act if they did it *because* they could not have done otherwise.

This principle seems to still be problematic for the compatibilist, since if a person is determined, then the reason that they act as they do is because of their causal antecedents. But Frankfurt rejects that principle too (what is his argument here?) and instead accepts

*The principle of alternate possibilities***: A person is not morally responsible for their act if they *only* did it because they could not have done otherwise.

In particular, he thinks that when we excuse someone because he couldn't have done otherwise, we must be taking it that their desires didn't cause them to do it. (There is a tricky issue here with overdetermination; and what if the reason that entailed that they couldn't have done otherwise does so by affecting their desires?) What is Frankfurt's argument for this?

SOME INITIAL DISTINCTIONS AND CLARIFICATIONS

- (i) Frankfurt is here talking about moral responsibility, not about freedom.. Would we say that Jones freely performed the action, even though he couldn't have done otherwise. Would we say, more broadly, that he is a free agent?
- (ii) Frankfurt says nothing about the experience of freedom, and whether the compatibilist can accommodate that.

LOCKE, AND THE FLICKER STRATEGY

Back in the Seventeenth Century, John Locke considered a case rather similar to Frankfurt's. Suppose that you go to visit your friend. Whilst you are sitting talking to him, someone locks the door of the room you are in, so that you cannot leave. You don't realize this and remain talking happily. Surely we want to say that you freely stay with your friend, even though you are not free to leave. So isn't the crucial thing whether or not you try to leave. Similarly then, mightn't the crucial thing for Jones be whether or not he tries not to do the thing that Black wants him to do? And won't this show a general response to Frankfurt cases? The device can only start to work after the agent has done something that shows they are trying to do otherwise: a 'flicker of freedom'. In response to that, some (e.g. Fischer) have claimed that Black might pick up on clues as to how Jones will try to act that come even before he tries.

DISPOSITIONS

Consider a fair coin: there is nothing in the coin itself that prevents it falling heads or tails. But suppose that there is a genie who looks after the coin, and who hates seeing it come down tails. If ever it is about to come down tails he intervenes and tips it over so that it comes down heads.

Suppose the coin is tossed, and it comes down heads without the genie intervening. Then we might say that it is a fair coin that has fairly come down heads. The facts about what would happen if it were to fall differently are irrelevant. (Worry: does this mean that someone who had bet that it would come down tails should pay up?) Now we might say the same about Jones. He is a free actor, who acts freely. Issues about how things would have gone are beside the point. If that is right then the lesson is that being free is a capacity, and that capacity should not be analyzed in terms of counterfactuals.

The point here is one that is widely recognized in the dispositions literature: capacities can be seen as a kind of disposition. A disposition like fragility cannot be analyzed by means of the counterfactual 'will break if dropped' since there might be an intervener who will catch it if it is dropped. How do we analyze it then? Perhaps in terms of certain intrinsic properties which are defeasibly connected to the truth of the counterfactuals.