FREE WILL IX

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

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 is 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.

A preliminary worry: Locke, and the Flicker Strategy

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; to which others (e.g. Widerker) have responded that this assumes determinism (does this matter: isn't the point that there can still be moral responsibility even if determinist is true?)

VIHVELIN

Vihvelin is a compatibilist. But she doesn't want to use Frankfurt's strategy. She wants to say that even if Black has implanted a device into Jones that would operate if Jones tried to do other than he currently intends to do , it is still true that, if Jones does do as he currently intends to do, he could have done otherwise.

Initial clarifications

(i) Vihvelin distinguishes Frankfurt's PAP from PAP'. The distinction is not altogether obvious. PAP' says that a person is morally responsible for performing an action X only if they could have not done X. PAP (as understood by Vihvelin) says that a person is morally responsible for performing an action X only if they could have done *something* else different; it needn't be the action X itself. (The clearest statement of this is in the paragraph beginning at the bottom on p. 5.) The crucial difference between them is that it is easy to show that PAP' is false: you are not responsible for doing something if you try not to do it, even if that attempt does not succeed. But it is much harder to refute PAP. You need a case in which the person is morally responsible, but in which there is *nothing* that they could have done differently. In effect, this is the flicker strategy. Only showing that PAP is false would undermine the traditional worry for compatibilists.

(ii) Conditional and counterfactual interveners. I don't understand why these are so named. The crucial thing is just that conditional interveners intervene at the moment when the agent starts to form the intention not to perform the action. The counterfactual intervener intervenes before that: they pick up on some sign that the agent is going to form that intention, and they act to preempt it.

The Argument

Conditional interveners cannot serve to show that PAP is false, since by the time they intervene the morally innocent agent has already done something that the morally guilty has not done: they have tried to form the intention not to perform the bad act. (Conditional interveners can however serve to show that PAP' is false.) So Vihvelin's focus is on whether the counterfactual intervener cases remove the possibility of doing otherwise.

Vihvelin's presentation of the example of the coin and the counterfactual intervener is clear. But then the argument becomes hard to follow. Remember what Vihvelin is trying to do: to show that even with the presence of the counterfactual intervener, the coin, when it falls heads without intervention, could have fallen tails. But her main argument for this works by saying that someone who denies this (i.e. someone who thinks that the coin couldn't have fallen tails) is committed to affirming :

(1) If Black had not existed, the coin might have landed tails

Whereas she wants to affirm what he takes to be the negation of (1), namely (2):

(2) If Black had not existed, the coin would still have come up heads

I have a number of worries with this part of the argument:

(i) On a natural reading, (1) isn't the negation of (2). We only get the result that it is if we believe what is said in footnote 26, i.e. that 'might' is the negation of 'would not'. There's another (more?) natural reading of 'might' (and of 'could') in which we say things like: 'It didn't happen, but it might have done'. From this, we get a natural reading of 'It wouldn't have happened, but it might have done.' On this alternative reading of 'might', Vihvelin herself would want to affirm (1).

(ii) Why would Vihvelin's opponent want to affirm (1), even with 'might' read in her sense? Surely what the opponent wants to affirm is:

(3) If Black exists, the coin could not have come up tails.

But that isn't the same as (1). Indeed, once (1) is read as the negation of (2), it seems to me that the opponent will not want to affirm (1); for why should the non-existence of Black make any difference to what *would* have happened, given that Black didn't intervene? It simply makes a difference to what *could* have happened.

Despite the problems with the argument, there seems to be something to be said for the claim that Vihvelin makes on p. 18 (also, confusingly, called '(1)'; I'll call it '(A)'):

(A) EITHER the coin comes up heads even though it could have come up tails, OR the coin comes up heads and could not have come up tails.

(Note that here the talk is of 'could' and not of 'might'; I take it that this is to be understood in standard way, so that something could have happened that wouldn't have happened.) The responses that Vihvelin makes to the responses to this seem to have something going for them. for instance, the fallacy she identifies on p. 20 is indeed a fallacy. Nonetheless, there seems something odd about the argument.

Counterfactuals and Possibility Claims

One feature of Vihvelin's approach is that she argues using counterfactuals: i.e. statements about what would have happened if … But 'can' and 'could' sentences are usually understood in terms of straight possibility. ('Could' is the simple past of 'can'; there is also a present tense version of 'could', but it is not easy to see how to analyze that.) Possibility here should be understood as relative possibility: one holds certain conditions fixed (the presence of Black, with his current intentions) and then asks what is possible relative to them. Thought of in this way, it is obvious why (A) is false: there is no possible world in which Black exists doing what he does, and the coin comes up tails. Does Vihvelin have a response to that?

Salvaging something

Still, isn't there something right about the idea that the cases in which Black does intervene are very different from the cases in which he doesn't. The point seems to be this: in the cases in which Black doesn't intervene, it is those features of the coin and surroundings that would come into play in a fair toss that determine how it falls. In short: the coin has a certain disposition (to fall fairly) and that disposition is what causes it to fall as it does, Black's presence notwithstanding. To see this, note the difference with a case in which Black has interfered with the coin, so that it is weighted, or has two heads.

The same can be said of an agent in a Frankfurt case; they have a certain capacity to choose freely, and when the intervener doesn't intervene this capacity is exercised. So we might think that this is why they are responsible in such cases. The crucial point is the exercise of a capacity, not the ability to do otherwise. that gives responsibility. typically these go together, but Frankfurt cases show that they need not.

The point here is one that is recognized in the dispositions literature (see Vihvelin's note 34): 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.