

20. FRANKFURT ON ALTERNATIVE POSSIBILITIES

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 (PAP): 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 moral responsibility and the ability to do otherwise

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:

the threat made no difference to him whatsoever.

the threat swamped all thought of his previous intention: it made all the difference

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:

The principle of alternate possibilities* (PAP*): 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** (PAP**):* 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 have claimed that Black might pick up on clues as to how Jones will **try** to act that come even before he tries.

DISPOSITIONS

Suppose that I have a fair coin; and a genie who will step in whenever it is about to land tails and *flip* it over the heads. I bet you that *it will* fall heads. Clearly I am *cheating*. But suppose it happens that the genie doesn't need to step in. Isn't it true that in that case the coin has freely, fairly fallen heads, even though it couldn't have fallen tails?

Rough idea: the coin falls freely heads *iff* it has an intrinsic disposition to fall *equally* heads or tails, and that disposition is realized in this case. The idea is that dispositions cannot be reduced to dispositions. To see this, note the difference with a case in which I have interfered with the coin itself, 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: 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.

AN ALTERNATIVE APPROACH TO MORAL RESPONSIBILITY: STRAWSON

Strawson argues that there are two different sorts of attitudes that we can take towards a person: reactive attitudes and objective attitudes. As an example of the first, consider resentment. Suppose someone hurts you: pushes you over for instance. Then you will be apt to feel, not just angry, but resentful towards them. And this ties into a network of attitudes that goes much further than simply holding someone responsible: you might expect them to be sorry; you might forgive them if they do; your affection for them might be lost if they do not; and so on. However, if it turns out that they pushed you by accident these feelings will be diminished; perhaps they will go altogether; it depends on the nature of the case. If they had been rushing to try and help you, and had fallen into you they probably will; indeed they might even be replaced by positive reactive attitudes like gratitude. (Unless they have a history of clumsy over-solicitous interference.) If, on the other hand, they accidentally pushed you because they were recklessly trying to get past to see someone else, your feelings of resentment might be somewhat reduced, but they are unlikely to go completely. We require certain standards of good will and concern from those around us (how much depends on how close we are to them); and we feel resentment when we think these are not met. (Note: Strawson is rather unclear on what is required: normally he speaks of goodwill, but also of regard, affection and esteem.)

The objective attitude, in contrast, requires stepping back from such involvement. We move to it when we truly think that the harm was caused accidentally; here we view the action objectively. In these cases we go on treating the person reactively. But sometimes we no longer take the reactive stance to a person in just about everything they do. Consider, for instance, someone suffering from serious mental illnesses like schizophrenia, or severe depression. Here we adopt the objective stance to the person. We no longer feel resentment to them; we rather treat them as people needing to be treated or managed. Why do we suspend the reactive stance when dealing with the very young or the mentally ill? Again because we do not think that their actions indicate a lack of good will towards us.

How should we understand these different sorts of attitude? Strawson's first point is that they are attitudes on the part of the person who is reacting, not different facts about the person they are reacting to. This is shown by the fact that we sometimes take both attitudes towards the same person for the same acts. Thus: suppose you are involved in a torrid love affair. Most of the time you are passionately in the grip of reactive attitudes to your beloved: affection, resentment as you think they scorn you, affection again as they ask your forgiveness, jealousy as you think they are unfaithful, and so on. Sometimes though it all gets too much. To escape from the unbearable strain you fall back into the objective stance. You remind yourself of the terrible childhood they had; you think of the earlier affairs they had that left them so wounded; you explain what they are doing as a result of the insecurities that these experiences have inflicted upon them. You can't maintain this suspension of the reactive stance for very long (or at least you can't suspend it and maintain the relationship with them; once it's over you might think of their actions, and indeed your own, almost entirely objectively); but the fact that you can suspend it at all shows that taking the stance is something that isn't just dictated by them; it's something that comes, at least partly, from you.

Strawson now asks: why should we think that acknowledging the truth of determinism will make us give up the reactive stance across the board? He thinks that it is absurd to think that it would. In the first place, we are just not able to give it up; it is too deeply ingrained. It would be to give up on our humanity. Secondly, even if we ask whether the acknowledging the truth of determinism should rationally make us give it up (whether or not we actually would), the answer is: no. For the decision as to whether we should rationally take the reactive stance is a practical one; whereas the truth of determinism is a theoretical question. Practical questions are answered by considering how we would benefit from the Afferent answers to them. And there is no question that our lives would be hugely impoverished if we were to give up on the reactive attitudes. (The difference between practical and theoretical questions is clearly related to the difference between is-statements and ought-statements. Roughly theoretical question as questions about how things are; and practical questions are questions about what we ought to do; and there are good reasons for thinking that we can never derive an ought-statement from an is-statement; the contention sometimes known as Hume's Law.)

All of this has addressed the compatibility of determinism and our practice of ascribing moral responsibility. But what about the compatibilism of determinism with the claim that people are free? (What we might call freedom compatibilism, to distinguish it from responsibility compatibilism.) There are two ways to go here. One is to accept that the claim that people are free is a theoretical claim, but one which is detached from the practical question of how we should treat people. So even if freedom compatibilism is false, this won't have the terrible consequences for our practice of ascribing blame that was suggested in the first lecture. (But what should we make of the phenomenology of freedom?)

There is another way to understand Strawson. This is to understand him as saying that the claim that a certain person is free is not really a descriptive theoretical claim at all. Rather it is simply something we say when we are prepared to take the reactive stance towards them. (Compare emotivism in ethics: emotivists say that the statement that a given thing is good isn't really a descriptive statement at all. It's just something that we say that expresses our attitude to it; it's like saying 'Hurrah') Understood this way, Strawson comes out as a kind of sophisticated freedom compatibilist: the claim that determinism is true is quite compatible with the claim that people have free will, since the latter is just a way of indicating that we are prepared to take the reactive stance towards them. It is not a descriptive statement at all. (But again, what now becomes of the phenomenology of freedom?)

Independently of this, we might wonder how successful Strawson's approach is on its own terms. And the big worry here is whether or not we can isolate the theoretical from the practical in the way he suggests. One aspect of this is the following: don't we think that there is a real question of whether we are right in thinking that we are justified in holding someone responsible? We don't just want to say: "This is what we do"; we want to say "This is what we are justified in doing". One way of thinking about this is to compare it with other practices. thus suppose someone made a parallel defense of religion. Suppose it could be shown that there was a natural human tendency to believe in a god, and that human beings flourished when they did. Would that justify belief in god? Wouldn't we want to say: whether or not it is unavoidable, or it benefits us, we want to know whether the belief is justified.

MIT OpenCourseWare
<http://ocw.mit.edu>

24.00 Problems in Philosophy
Fall 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.