

24.231 Ethics – Handout 24 Harry Frankfurt, “Alternate Possibilities and Moral Responsibility”

Classic Incompatibilist Argument:

- (1) If determinism is true, then every human action is causally necessitated.
- (2) If every action is causally necessitated, no one could have acted otherwise than they do.
- (3) Someone has free will only if she could have acted otherwise than she does.
- (4) Determinism is true.
- (5) (*Conclusion*) No one has free will.

- *Hard determinism:* embraces the premises and thus the conclusion of this argument.
- *Libertarianism:* Accepts the incompatibility of free will and determinism, but rejects (4) – determinism doesn't hold.
- *Compatibilism:* Rejects the incompatibility of free will and determinism, usually by rejecting either (2) or (3); that is, compatibilists either embrace a compatibilist understanding of “could have done otherwise,” or reject the thesis that free will requires the ability to do otherwise.

Frankfurt isn't really challenging (5); he's challenging a parallel argument concerning *moral responsibility*, rather than *free will*. He rejects a version of premise (3) that concerns moral responsibility, which he calls

The Principle of Alternate Possibilities (PAP): A person is morally responsible for her act only if she could have done otherwise than she does.

Why should we be tempted to accept PAP in the first place?

It, or rather, part of it, seems to be entailed by the Ought Implies Can Principle”

Ought Implies Can (OIC): A person ought to do something (or refrain from doing something) only if she can do (or refrain from doing) that thing.

Let's divide each of these principles into two part:

PAP(i): A person is *praiseworthy* for doing X only if she could have done otherwise than X.

PAP(ii): A person is *blameworthy* for doing X only if she could have done otherwise than X.

OIC(i): A person ought to *do X* only if she can do X.

OIC(ii): A person ought to *not do X* only if she can *not do X*.

Now, it seems very plausible that

- (1) A person is blameworthy for doing X only if she ought to not do X.

Now add our ought-implies-can premise:

- (2) OIC(ii): A person ought to not do X only if she can not do X.
- (3) Not doing X constitutes doing otherwise than X.

This gives us:

- (4) PAP(ii): A person is blameworthy for doing X only if she can do otherwise than X.

So the OIC principle seems to entail a part of the PAP. And Frankfurt rejects, on the basis of his examples, not just the PAP as a whole, but also each of its parts. So it seems that Frankfurt is committed to rejecting the OIC principle, or at least the part of it (OIC(ii)) that concerns omissions. Since the OIC principle (including the part about omissions) is itself strongly supported by our intuitions, this is a significant consequence of Frankfurt's argument, if it is sound.

Let's look at Frankfurt's argument against the PAP:

Frankfurt suggests that PAP looks plausible when we're thinking about cases where the very circumstances that bring it about that someone does something also make it impossible for him to do otherwise: e.g., some cases of coercion, hypnotic suggestion, or inner compulsion. But, he says, the principle looks much less plausible in cases where what motivates the agent is *not the same thing* as what makes it impossible for him to act otherwise. It's cases like this that Frankfurt asks us to consider.

Imagine three cases of coercion: in all three, Jones decides for reasons of his own to do something, and the someone threatens him with a very harsh penalty unless he does precisely the thing he has already decided to do. Now:

- (1) In the first case, we're to imagine that the threat actually exerts no effective force on Jones₁ whatsoever. It forms no part of his motivation for doing what he does. Once he's made up his mind to do something, he simply doesn't care about threats of this sort. In this case, Jones₁ is fully responsible for his action. But, says Frankfurt, the threat in this case is also not sufficient to prevent Jones₁ from doing otherwise: if he'd intended differently, we can imagine he would have acted differently, despite the threat.
- (2) In the second case, we're to imagine that Jones₂ is completely "stampeded" by the threat: it upsets him so much that he entirely forgets his earlier decision and is motivated to perform the act only by the threat itself. In this case, says Frankfurt, it's true that he could not have done otherwise, and also that he should not be held responsible.
- (3) In the third case, we're to imagine that Jones₃ would have been motivated to perform the action by the threat had he not already had a sufficient motive for performing it. The threat "impresses" him, but is motivationally irrelevant. Frankfurt acknowledges that cases of this kind may be hard to recognize (how are we to know which motive Jones is acting on?). But, he says, if it's apparent to us that Jones₃ acted on his own decision, not on the basis of the threat, we should conclude that Jones₃ is morally responsible for what he does, even though, since he would have submitted to the threat, he could not have avoided doing what he did.

Jones₃ is Frankfurt's first counterexample to the PAP. It shows, more particularly, that PAP cannot be supported by appealing to the idea that people are not morally responsible for

doing what they are coerced to do, since we take coercion to absolve people of morally responsibility only if the coercive threat is what motivates them to act.

But we might doubt that Jones₃ provides a decisive counterexample to the PAP: after all, he *could* have chosen to defy the threat – he *could* have done otherwise than perform the action (although at great cost to himself).

So Frankfurt comes up with a fourth example:

(4) Suppose someone – Black, let us say – wants Jones₄ to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones₄ is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones₄ is going to decide to do something other than what he wants him to do. If it does become clear that Jones₄ is going to decide to do something else, Black takes effective steps to ensure that Jones₄ decides to do, and that he does do, what he wants him to do.⁸ Whatever Jones₄'s initial preferences and inclinations, then, Black will have his way. (Frankfurt, p. 835)

We can imagine that Black implants a device in Jones₄'s brain that will be triggered if Jones₄ shows the initial signs of a decision not to perform the action. And imagine that the device is never triggered: Jones reaches the decision to perform the act, and performs it, just as he would have done had Black never entered the picture. In this case, Frankfurt maintains, we have a clear counterexample to PAP: Jones could not have done otherwise than he did, but we should still hold him morally responsible. (As in the coercion case, Black's trigger, while securing that he could not have done otherwise, was not the reason why he acted as he did.)

Questions:

- Is this genuinely a case in which Jones₄ could not have done otherwise than he does? Let's call the action Jones₄ intends to perform, and Black wants him to perform, X.
 - It seems like a case where Jones₄ could not have done otherwise than do X (though this may depend on how we individuate actions – if we individuate them by their motives, the action Jones₄ performs when Black doesn't intervene may not be the same as the one he performs when Black *does* intervene, depending also on the form Black's intervention takes...)
 - But it may not seem like a case where Jones₄ could not have done *anything* other than what he did do. After all, some difference in his behavior would have to trigger Black's intervention. (This is sometimes called a "flicker of freedom"...))
 - Perhaps we're not supposed to think of this as 'trigger' as something Jones₄ does? Perhaps it's some process that occurs before Jones₄ acts at all that determines whether or not he sets off the trigger (perhaps Jones always twitches before he decides to do X)? But in that case, we might worry that incompatibilists might balk at the suggestion that Jones₄ is nonetheless responsible for what he does – after all, his

choice what entirely determined by some non-voluntary process that occurred before he made it.

- Thinking more carefully about how Black's intervention is supposed to be triggered raises further problems for Frankfurt: Frankfurt claims to have identified a case in which Jones₄ is morally responsible for some act despite the fact that he could not have done otherwise than he does, because the fact which causally explains his inability to do otherwise is not the fact that causally explains his acting as he did: "there may be circumstances that make it impossible for a person to avoid performing some action without those circumstances in any way bringing it about that he performs that action." (p. 837) But is Black's intervention an example of this? Let's say Black will intervene only if Jones₄ does not twitch. Now there seem to be two possibilities:
 - Either Jones₄' twitch, or whatever causes him to twitch, is causally sufficient for his doing X – in that case the same thing that makes it impossible for Jones₄ to avoid Xing (the process that leads to the twitch) is also what causes him to X. We don't have a case where the two come apart.
 - Or the twitch (or what causes it) is not causally sufficient for Jones₄' doing X, but merely accurately predicts that he will do it. But now it seems that, once Jones₄ has twitched (and the trigger is not released), he still has the possibility of doing otherwise: again, the conditions Frankfurt wants his example to meet do not seem to have been met... (This is sometimes referred to as the "Dilemma Defense" of the PAP.)
- What alternative principle can Frankfurt appeal to to capture our intuitions about coercion and moral responsibility?

- He starts by considering

PAP2: A person is not morally responsible for what he has done if he did it because he could not have done otherwise.

But, as Frankfurt points out, PAP2 is also sufficient to determine the incompatibility of moral responsibility and determinism: after all, if we're causally determined to act as we do, this fact both explains why we do what we do, and why we could not have done otherwise.

- So Frankfurt wants to reject PAP2, and replace it with:

PAP3: A person is not morally responsible for what he has done if he did it only because he could not have done otherwise.

Why think *PAP3* is preferable to *PAP2*? And is *PAP3* compatible with compatibilism about moral responsibility and determinism?

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

24.231 Ethics
Fall 2009

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