

forthcoming in *Analysis*

Frankfurt counterexample defended

Ezio Di Nucci

In *Alternate Possibilities and Moral Responsibility*, Frankfurt sets out to refute the principle according to which “a person is morally responsible for what he has done only if he could have done otherwise” (1969: 829). Frankfurt devises a counterexample in which an agent is intuitively responsible even though she could not have done otherwise:

Suppose someone - Black, let us say - wants Jones 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 Jones 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 Jones is going to decide to do something other than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. Whatever Jones's initial preferences and inclinations, then, Black will have his way... Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it (1969: 835-36).

As things turn out, Jones's decisions and actions are autonomous and unconstrained; he acts “for reasons of his own”. Jones is intuitively responsible for what he does, but he could not have done otherwise: Black would have intervened to prevent it and “to ensure that Jones decides to do, and that he does do, what he

wants him to do". So Jones has no alternatives; still, he appears to be responsible. Therefore, concludes Frankfurt, the principle is false: a person can be morally responsible for what she has done even though she could not have done otherwise.

Larvor has recently argued in this journal against Frankfurt's counterexample by pointing out that in the counterfactual scenario where Black intervenes "it would be false to say that Jones kills Smith" (2010: 1). Therefore it would not be true that Jones is responsible for killing Smith even though Jones did not have any alternatives: if in the counterfactual scenario Jones does not kill Smith, then Jones does have an alternative, not killing Smith. And then we can ground Jones's responsibility for killing Smith on this alternative. So Frankfurt would not have offered a case in which an agent is clearly responsible despite not having alternative possibilities: Jones's case would not be such a case because Jones did have an alternative, as shown by Jones not killing Smith in the counterfactual scenario.

Larvor's powerful idea is that what happens in the counterfactual scenario does not count as Jones killing Smith because of Black's intervention. In the counterfactual scenario Black is controlling Jones, and it is therefore Black who kills Smith: "This is Black's deed. The fact that the causal chain passes through parts of Jones's body does not make it Jones's deed. Jones is merely Black's unwilling instrument" (2010: 1). Here it might be objected that to say that Black's control is incompatible with Jones's agency begs the free will question by assuming that agency is incompatible with external determination. But I actually agree with

Larvor on this point: the counterfactual scenario is such that Jones does not act, and therefore he does not kill Smith (see Alvarez 2009 and Di Nucci 2010).

Denying that Jones acts in the counterfactual scenario, though, is not enough to reject Frankfurt's counterexample. Claiming that Jones does not kill Smith is not equivalent nor does it imply claiming that Jones avoids killing Smith. If the planet had imploded the moment before I wrote the last sentence, I would not have written the last sentence. But that does not mean that I would have avoided writing the last sentence. It would have been true to say of me that I did not write the last sentence. But it would have been false to say of me that I avoided writing the last sentence. For something to count as avoiding ϕ -ing rather than just not ϕ -ing, it has to be up to me that I do not ϕ , while in the case in which the planet implodes just before I ϕ it is not up to me that I do not ϕ .

It is not enough that Larvor argues that Jones does not kill Smith: Larvor needs to argue, also, that Jones avoids killing Smith. And in order to argue that Jones avoids killing Smith we must say something more than just that Jones does not kill Smith. We must show that not killing Smith was up to Jones. What Larvor says here is that "by (unwittingly) forcing Black to intervene [Jones] avoids killing Smith" (2010: 2). So Jones can avoid killing Smith by unwittingly forcing Black to kill Smith. There is something that Jones can do to bring it about that he does not kill Smith, and therefore there is a way for Jones to avoid killing Smith: that is, to unwittingly force Black to intervene, so that Black, and not Jones, kills Smith.

The crucial point upon which Larvor's argument against Frankfurt depends is then, whether Black's intervention is up to Jones. But, as it has been already recognized (Fischer 1999), Frankfurt had anticipated this point in his original formulation of the argument, clarifying that whether Black intervenes is not actually up to Jones: "We can imagine that Jones has often confronted the alternatives – A and B – that he now confronts, and that his face has invariably twitched when he was about to decide to do A and never when he was about to decide to do B. Knowing this, and observing the twitch, Black would have a basis for prediction" (1969: 835). Involuntary twitches aren't up to agents, therefore it isn't up to Jones whether his face twitches; therefore it's not up to him whether Black intervenes.

Even if we accept that Jones does not kill Smith in the counterfactual scenario, Frankfurt's counterexample is still safe because showing that Jones does not kill Smith in the counterfactual scenario does not show that Jones avoids killing Smith, because whether Black intervenes is not up to Jones. In conclusion, I have not just shown that Larvor's case fails; also, more generally, I have argued that Frankfurt's counterexample does not depend on the agent acting (let alone doing the same thing) in the counterfactual scenario.

Universität Duisburg-Essen
45117 Essen, Germany
ezio.dinucci@uni-due.de

References

Alvarez, M. (2009), 'Actions, Thought-Experiments, and the Principle of Alternate Possibilities', *Australasian Journal of Philosophy* 87: 61-81.

Di Nucci, E. (2010), 'Refuting a Frankfurtian Objection to Frankfurt-type Counterexamples', *Ethical Theory and Moral Practice* 13: 207-213.

Fischer, J.M. (1999), 'Recent Work on Moral Responsibility', *Ethics* 110: 93-139.

Frankfurt, H. (1969), 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy* 66: 829-39.

Larvor, B. (2010), 'Frankfurt counter-example defused', *Analysis* (forthcoming).

Advance Access: analysis.oxfordjournals.org/cgi/content/full/anq038