

Gesollten hand in hand gehen, eröffnet sich der Mensch daher die Chance, aus der Selbstverständlichkeit seines sicheren Unterwegsseins, das sich entlang der gewohnten Bahnen der Konvention und der herrschenden Moral bewegt, stolpernd herauszutreten, den Panzer seiner selbstvergessenen Unversehrtheit einzureißen und sich für die Widerständigkeit eines Einstehens zu entscheiden.

Davor Pećnjak

Alternate Possibilities and Responsibility: A Discussion of PAP

The problem of free will and the problem of moral responsibility are two different problems, generally speaking, but they are in some respects closely connected and have very important junction where our notions of moral responsibility are strongly dependent about how we construe our abilities or inabilities to decide, intent and perform actions. On the other hand, we do not have to have a definite solution to the problem of free will (and, accordingly, of our freely or not freely deciding, intending and acting) to be in a position to discuss moral responsibility. A solution to the free will problem does not have to be given prior to settling the problems concerning moral responsibility. We can discuss and argue about various possible solutions for the problem of moral responsibility just under the assumptions what *would be* the case if this or that position about free will, free deciding or free action is true. We can formulate some general principles of moral responsibility and test them under various possible assumptions and scenarios.

One of the most important principles, or, indeed, the most important principle, in discussions of responsibility is given by the phrase „could have done otherwise.“ Namely, both intuitively and more technically or philosophically, we think that someone is responsible for something that person did, only if that person could have done otherwise. Of course, it is meant that he or she could have done otherwise in just the same antecedent conditions. If circumstances were only slightly altered and person had acted differently, then situation is in fact completely different so it does not matter for the considerations of moral responsibility under the same specification of states of affairs antecedent to act or acts in question.

One brief general remark before we go into the main discussion. There would be no question of moral responsibility if hard determinism or incompatibilistic determinism is the case. This type of determinism holds that everything that happens – no matter whether it is the flow of a river, development of the whole universe, billiard play, falling of the rock, writing an article by a philosopher, playing tennis match by the amateur or professional, just rising the hand or anything else – is completely determined solely by the initial conditions of the universe and laws of nature. Nothing could be different from what it is, no thing, no process could be different, no action could be done otherwise than it is done, including all human actions. So, there would be no real agents in such a world; no free will for humans; and, all human beings have done or will do, do not depend on their own acting but only on antecedent conditions, states and laws of nature and everything is inevitable. Because these conditions are out of influence and control of every human being and they are sufficient for everything that happens, there is no place for any kind of responsibility. In such a world, moral responsibility simply would not exist. I think that this is the right discussion of the moral responsibility if hard determinism is the case.

But, since we don't know yet for sure what kind of world is our own world, let us discuss possibilities of moral responsibility. As I said earlier, we shall focus our discussion on the principle expressed by the phrase „could have done otherwise.“

This principle is called „the principle of alternate possibilities“, PAP for short hereafter, and it is formulated as follows:

(PAP) A person is morally responsible for what he has done only if he could have done otherwise.

For some time this principle seemed so obvious that it was generally firmly held and it was unchallenged. Attack on this principle was launched in a well-known article by Frankfurt.¹ He

1 H. G. Frankfurt, „Alternate Possibilities and Moral Responsibility“, *Journal of Philosophy*, vol. 66., (1969), pp. 829–839.

tried to show that this principle is false. Let us review his counter-example and argument. Since Frankfurt is best in his own words, let me first cite the most pregnant part of his article, as well as others do:

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 wait 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 something other than what he wants him to do. If it does become clear that Jones is going to decide 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.²

Let us be a bit more concrete about this. I borrow here a slightly modified story from Ginet's article.³ Suppose that Jones is considering to kill a certain person Smith. There is also another person – Black – who also wants that Jones kills Smith. But Black prefers to avoid showing his hand if not necessary. He can force Jones to decide to kill and then kill Smith, if Jones decides not to kill Smith; he installs a mechanism which will sufficiently causally influence Jones to kill Smith in a case Jones decides not to kill Smith. Of course, Jones is always completely unaware of the existence of the mechanism and does not know anything about the mechanism.

This example is construed in such a way that it is obvious that person in question, Jones, could not have done otherwise (if he decides not to do what Black wants him to do, namely to kill Smith, then Black takes steps that /inevitably/ Jones nevertheless in fact does it). Whatever the initial decision, Jones does A; he cannot fail

2 *ibid.*, p. 835.

3 C. Ginet, „In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing“, *Philosophical Perspectives*, vol. 10., (1996), pp. 403–417.

to do A; he cannot avoid doing A, namely, killing Smith. But, in fact, Jones could kill Smith in two slightly different ways: one way is acting only on his own reasons without the intervening of Black's mechanism, and the second way is acting influenced sufficiently causally by Black's mechanism. In the case where mechanism intervenes, then Jones is forced to kill Smith and he is not responsible for that. But, though Jones cannot do otherwise, Frankfurt argues that if Jones kills Smith only on his own reasons without the intervention of Black's mechanism, he is responsible for doing that. Though Jones cannot do otherwise, he is responsible in the last case (because he kills Smith for his own reasons only). So, the PAP is false.

Namely, Jones, when he kills Smith only for his own reasons without the intervention of the mechanism, he would do this action in the same way even if there is no any such mechanism. Since he is clearly responsible for killing Smith in ordinary case where he kills him and there exists no any mechanism that could possibly force him to do so, when he kills Smith in the same way even if mechanism is there but it does not activate and intervene so Jones kills Smith only for his own reasons, he is responsible for killing him.

It seems fairly strong and convincing that Principle of alternate possibilities thus is not generally valid. But it does not seem so for everyone. Some people were trying to save PAP by showing that Frankfurt's case is not all that convincing. I would like, through examining the attack on Frankfurt made by Carl Ginet⁴ to consider this more fine-grained case and reconsider the application of PAP and its possible falsity.

The crucial thing is as follows. Case is refined by Ginet in the way which supposedly does not alter Frankfurt's original case. Let us see how the story goes. As we have already seen, there are three different points in time. Black sets a mechanism which will cause Jones to kill Smith by t_3 if Jones has not killed Smith by some deadline at t_2 . So, if Jones does not kill by that deadline, at t_2 the mechanism is activated and causally necessitates Jones to kill Smith

4 *ibid.*

by t_3 . This activating renders Jones unable to avoid killing Smith by t_3 . The mechanism is not activated because Jones kills Smith at t_1 . Now, let me have a bit longer citation from Ginet's article which is supposed to serve, from his point of view, to show that Frankfurt is mistaken and to defend PAP:

Now for the crucial question. Is Jones morally responsible for its being the case that he does B by t_3 , despite the fact that he could not avoid the obtaining of that state of affairs? It is Frankfurt's intuition that he is responsible for it, since Jones does B by t_3 on his own without the intervention of the mechanism and would have done B by t_3 in the same way for the same reasons even if the mechanism had not been there and he could have avoided doing B by t_3 . I do not share that intuition. I do, however, have the intuition that there is something here for which Jones may be responsible, something that entails his doing B by t_3 – namely, his doing B at the precise time he actually did it, t_1 . This latter state of affairs is one whose obtaining Jones could have avoided (though, if he had avoided it, he would have done B at some later time between t_2 and t_3). Jones's doing B at t_1 differs from Jones's doing B by t_3 , only in being temporally more specific. My intuition is that Jones is not responsible for obtaining of the temporally less specific state of affairs, because, owing to the presence of Black's mechanism, Jones could not have avoided it, but he may be responsible for the obtaining of the temporally more specific state of affairs which he could have avoided. This phenomenon of being responsible for the obtaining of a more specific state of affairs while *not* being responsible for the obtaining of an entailed less specific state of affairs is not at all uncommon. I am, for example, responsible for my being now in the particular room I'm in but I am not responsible for my being now within a one-million-mile radius of the center of the earth. I'm inclined to think that the attraction of Frankfurt's intuition about this example arises from failure to distinguish the two states of affairs, Jones's doing B at t_1 and Jones's doing B by t_3 , which differ only in their temporal specificity.
(...)

Notice that I'm not saying that what Jones may be responsible for is its being the case that he does B *on his own rather than as a result of Black's mechanism*. It would be a mistake to say this. Since Jones was completely unaware of Black's mechanism, he neither knew nor should have known the fact expressible by his saying „I am doing B on my own, rather than as a result of Black's

mechanism;" but he could have been responsible for making the fact obtain only if, at the time, he knew or should have known that he was doing so. What I say Jones may be responsible for is its being the case that he does B at t_1 . He did know that he was making that fact obtain.⁵

First, we do not assess, in this kind of examples, someone's responsibility according to what someone knows or what someone is aware of. Responsibility is assessed according to the facts of the matter or state of affairs which obtain or not, irrespective of what someone subjectively knows or is aware of. So, Jones's responsibility is not to be judged on what he is aware and could say. It is to be judged exactly on what Ginet thinks it is mistake to say, namely whether he did it on his own or not (it is really irrelevant whether Jones knows or is aware of that he did something on his own or some force outside him took complete control over him). So, this manouvre of Ginet does not succeed to show that it is wrong to assess Jones's responsibility on whether he did something on his own or not.

Notice also that Jones is in Ginet's words always excused exactly and only in the same situations as in Frankfurt's – namely, only when Black's mechanism is activated. Ginet had not provided an explanation why temporally more and less specific state of affairs would be relevant in assesing moral responsibility. Ginet would refute Frankfurt only if he conclusively showed that if Jones kills Smith by t_3 in the presence of Black's mechanism but without activating that mechanism (which means that Jones kills on his own reasons so this is a situation where he would kill even if mechanism is not there) he is not responsible for doing that by t_3 . But nowhere Ginet says something like that and not even tries to show this. But mere presence of the mechanism, though it makes the situation in which Jones could not have acted otherwise, if not also activated, does not bring Jones killing Smith *whenever* he kills Smith. So it seems that temporally more and less specific state of affairs do not play some important role in assessing moral responsibility.

5 *ibid.*, pp. 406–407.

To make things clearer let us once again consider the applicability of PAP. Here follows slightly altered scenario but I hope that it could help to see things a bit clearer.

Assume the following situation:

Jones at t_1 decides to kill Smith and proceed to do this; but at t_2 before he actually killes Smith, mechanism is activated regardless of Jones's already made decision to do that; mechanism, of course completely determines that Jones kills Smith at t_3 . After the activation of the mechanism Jones certainly could not decide and do otherwise; he is fully deprived from changing the decision and deprived from refraining to kill Smith after t_2 – the time of the activation of the mechanism. Until t_2 Jones could decide to do otherwise but after t_2 he could not. At t_3 he kills Smith. If there were no mechanism, Jones could decide to act otherwise and act otherwise in all the moments between t_1 and t_3 . So, from t_2 he is deprived of possibility to decide and act otherwise which he would have if there were no mechanism activated. His time for deciding to do otherwise would be longer. But is he responsible in such a situation for killing Smith? It seems that answer depends considering *how* (or under what conditions) Jones killed Smith and not *when* or *when more specifically* or *less specifically*. Strong intuitions lead us to consider how much he was forced or causally determined by some force outside him to do this or did he do it without being forced or to say whether he did it by himself or to say only for his own reasons.

So, since it seems that Frankfurt-style counterexample to PAP is not refuted, I would answer that question with – yes, he is responsible but maybe only partial. How is it so? In the time before t_2 , before mechanism activates, he was on his way to kill Smith by his own reasons. After t_2 , the activation of the mechanism is causally sufficient to bring Jones to kill Smith. From t_2 Jones could not change his decision and could not avoid killing Smith. But it does not absolve Jones from partial responsibility at least. Why? Because from t_1 to t_2 he could have done otherwise and could have decided not to kill Smith. After t_2 , Jones's killing Smith would be necessitated by the mechanism, irrespective of his prior decisions made by his own. Then he would not be responsible. But since he already had decision to kill Smith (for his own reasons only), he

bears a partial responsibility for killing Smith. He is not, of course, responsible for the time interval between t_2 and t_3 (at t_3 he kills Smith in revised story) because he was totally deprived of the possibility to change his decision, namely, he was deprived by the mechanism, so by means outside him and his control, of possibility to decide otherwise and carry the action according to such a changed decision (namely, not to kill Smith). So, though he could not have done otherwise, he is still partially responsible even in such a situation. He is partially responsible just because he proceeded for killing Smith even in the period when he was able to decide not to kill him and thus he deprived himself on his own to be absolved from responsibility when mechanism activates at t_2 . At t_2 , then, situation could be different: Jones enters t_2 either with the decision, made on his own, to kill Smith, or enters t_2 with pure heart, namely without any decision to kill Smith. It is a different situation whether mechanism activates when Jones has not made any decision and does not have any intention to kill Smith or whether Jones has made such a decision and have an intention to kill Smith. Surely he is responsible for the decision and having intention to kill Smith made on his own.

We would not judge someone responsible for what one did if suddenly machine took completely over that person and if that person had not made any decisions about action which that machine caused her to do, but we would look different if person prior to activating the machine already had made a decision to make the very same action which that machine caused her to do (of course, regardless of knowing by that person whether or not machine would activate or is actually activated).

So, Jones's entering t_2 with already made decision on his own to kill Smith confer at least partial responsibility on him, even though from t_2 on, machine was causally sufficient to cause Jones to kill Smith regardless of his own prior decisions.

I hope that these reflections, though short, show that it still seems that Frankfurt is right – namely that PAP is not generally valid. But I think also that applicability of some kind of such a principle is not completely eliminated, because of the importance of notions like „being forced to do something“, „coercion“, „causally necessitated to do something“ etc. They also play role in deter-

mining responsibility in a significant way. Combining these with the PAP, perhaps we can give more precise meaning of the PAP: when and how it could be applied though principle is not universally applicable. But this is a matter for another article.