CRÍTICA, Revista Hispanoamericana de Filosofía. Vol. 35, No. 104 (agosto 2003): 109–120

BLOCKAGE CASES: NO CASE AGAINST PAP*

CARLOS J. MOYA ESPÍ
Facultad de Filosofía y Ciencias de la Educación
Universidad de Valencia
Carlos.Moya@uv.es

SUMMARY: According to the Principle of Alternative Possibilities (PAP), an agent is morally responsible for something she has done only if she could have done otherwise. Harry Frankfurt held that PAP was false on the basis of examples ("Frankfurt cases") in which a counterfactual, and unactivated, device ensures that the agent will decide and do what she actually decides and does on her own, if she shows some sign that she is going to decide and do something else. Problems with these cases have led some thinkers to design examples in which the counterfactual factor is replaced by a device that actually blocks alternative possibilities. I argue that, even if these cases did not illicitly assume determinism, they are not successful against PAP anyway, for they violate a plausible condition on moral responsibility that Fischer has called "reasons-responsiveness".

KEY WORDS: alternative possibilities, moral responsibility, reasons, decision

RESUMEN: Según el Principio de Posibilidades Alternativas (PPA), un agente es moralmente responsable de algo que hizo sólo si podría haber actuado de otro modo. Harry Frankfurt sostuvo que el PPA era falso sobre la base de ejemplos ("casos Frankfurt") en los que un dispositivo contrafáctico, y no activado, asegura que el agente decidirá y hará lo que de hecho decide y hace por sí mismo, en el caso de que muestre algún signo de que va a decidir y hacer algo distinto. Los problemas que plantean estos casos han llevado a algunos pensadores a diseñar ejemplos en los que el factor contrafáctico es reemplazado por un dispositivo que bloquea de hecho las posibilidades alternativas. Sostengo que, aun cuando estos casos no asumieran ilícitamente el determinismo, no tienen éxito frente al PPA, porque violan una condición plausible de la responsabilidad moral que Fischer ha denominado "capacidad de respuesta a razones".

PALABRAS CLAVE: posibilidades alternativas, responsabilidad moral, razones, decisión

*This paper is part of the research I carried forward at the University of Sheffield during the academic year 2002–2003, with a scholarship for Stays in Foreign Research Centres awarded by the Spanish Ministry of Education. I am grateful to Christopher Hookway for his wise remarks about previous drafts of this essay.

Harry Frankfurt's path-breaking article, "Alternate Possibilities and Moral Responsibility" (Frankfurt 1969) made a strong case against the widely assumed view that alternative possibilities are required for moral responsibility for a certain action. Frankfurt himself called this assumption the "Principle of Alternate Possibilities" (PAP). According to this principle, in Frankfurt's own words, "a person is morally responsible for what he has done only if he could have done otherwise" (Frankfurt 1969, p. 829). Frankfurt's criticism of PAP rested mainly on a counter-example to it. Following Frankfurt's steps, many other putative counter-examples to PAP have been produced. Until recently, they have shown the same basic structure as Frankfurt's original example. We shall call them 'classical Frankfurt cases'. Classical Frankfurt cases feature an agent who, on her own, deliberates, decides to perform a certain action and does so; however, unknown to her, if she showed some inclination towards an alternative way of acting, she would be prevented from acting in such an alternative way by a factor which would then be activated; but, since she shows no such inclination, this factor remains causally inert.

An important feature of classical Frankfurt cases, then, is that the factors that prevent the agent from doing otherwise are purely counterfactual. This feature confers on these cases a significant advantage, namely that the intuition that the agent is morally responsible for what she does is very strong and natural, for we feel that she would have decided and acted in exactly the same way if the counterfactual factors, which ensure that she could not do otherwise, had been absent. This clearly distinguishes these cases from typical coertion or compulsion cases, in which the coercive or compulsive factor causally affects the agent's decision, so that we strongly feel that she does not decide and act in a sufficiently autonomous, self-determined way to ascribe her full responsibility. However, there is a price to pay for this important advantage of classical Frankfurt cases, namely that, since the counterfactual factors' activation is contingent upon the agent's showing a certain relevant sign, the agent is bound to have alternatives of *some* sort. She must be

able to show or not to show the relevant sign. The nature of these alternatives, then, depends on what this sign is supposed to be in the particular case at hand. So, the presence of alternatives of some sort is a *structural* feature of classical Frankfurt cases. And this means that they contain a crack in which defenders of PAP can insert a wedge.

Some thinkers have thought of construing Frankfurt cases with no such crack. In these cases, any alternatives, however thin, are ruled out because the counterfactual factors have been replaced by actual blocking mechanisms. These prevent any alternatives from arising *without, however, causing them not to arise*. The general idea is to get the agent to decide and do on her own something which, owing to the blocking device, is nonetheless the only thing she can actually decide and do. This means that there need be *no sign* that the agent could show, and the alternatives of showing that sign or not are simply not available. This advantage over classical Frankfurt cases, however, does not go for free. As one may expect, one difficulty is to convincingly show that a mechanism that is actually blocking alternatives is not thereby exerting any causal influence on the actual process of decision making. A related difficulty is that the intuition that the agent is morally responsible is likely to be much less firm and stable than in classical Frankfurt cases.

Both classical and blockage Frankfurt cases have to respect some adequacy conditions if they are to be plausible. One of them is that determinism has not to be assumed, even implicitly, for this will beg the question against incompatibilists, and these will not judge that the agent is morally responsible. Besides, there are also some requirements related to rational control. The process of practical reasoning and decision making has to meet some minimal standards: the agent must be able to consider reasons and to decide according to them while the process develops. A defective process or an impaired capacity of decision making will induce, in incompatibilists and compatibilists alike, the judgment that the agent is less than fully responsible.

Blockage cases have been recently proposed by David Hunt (Hunt 2000) and by Alfred Mele together with David Robb (Mele and Robb 1998).

Hunt thinks it is worth exploring whether "the unavoidability essential to a Frankfurt scenario" necessarily has to rest "on a counterfactual device" (Hunt 2000, p. 217). Hunt suggests it does not have to. In fact, he thinks that the classical Lockean example of a man who remains gladly in a room of which he cannot actually get out provides support for this answer. Of course, Locke's example as it stands cannot yield the wanted results, for it still leaves open many alternatives to the agent, such as his trying to leave the room, which are relevant to his moral responsibility. Hunt then conceives of some other ways in which "unavoidability does not wait upon a counterfactual trigger and so can extend to all the agent's actions, leaving no alternate possibilities to ground moral responsibility" (Hunt 2000, p. 217). Of these ways, by far the most promising against expected rejoinders by PAP defenders is to show how to construct blockage cases. The general structure of a blockage case is nicely outlined by Hunt himself in the following text:

> Imagine then a mechanism that blocks neural pathways rather than doorways [...] The mechanism blocks alternatives in advance, but owing to a fantastic coincidence the pathways it blocks just happen to be all the ones that will be unactualized in any case, while the single pathway that remains unblocked is precisely the route the man's thoughts would be following anyway (if all neutral pathways were unblocked). Under these conditions, the man appears to remain responsible for his thoughts and actions [...]. (Hunt 2000, p. 218)

A fantastic coincidence, indeed. But we are supposed to play with conceptual possibilities, and this would seem to be one. Let us examine this new route against PAP by using a concrete example of this rather general structure. The example was designed by Mele and Robb, in an article published in fact before Hunt's:

> At $t1$, Black [a neurosurgeon] initiates a certain deterministic process P in Bob's brain with the intention of thereby causing Bob to decide at $t2$ (an hour later, say) to steal Ann's car. The process, which is screened off from Bob's consciousness, will deterministically culminate in Bob's deciding at $t2$ to steal Ann's car unless he decides on his own at $t2$ to steal it [. . .]. The process is in no way sensitive to any 'sign' of what Bob will decide. As it happens, at $t2$ Bob decides on his own to steal the car, on the basis of his own indeterministic deliberation about whether to steal it, and his decision has no deterministic cause. But if he had not just then decided on his own to steal it, P would have deterministically issued, at $t2$, in his deciding to steal it. Rest assured that P in no way influences the indeterministic decision-making process that actually issues in Bob's decision. (Mele and Robb 1998, pp. 101–102)

The conclusion seems to be, again, that Bob is responsible for deciding to steal Ann's car (and for stealing it, provided that he carries out his decision). One may wonder what would happen if, at $t2$, the decision arrived at by Bob's indeterministic decision-making process (call this process 'x') is the decision *not* to steal Ann's car. Mele and Robb address this issue. Their response, in general terms, is as follows. In case of conflict, P, the deterministic process initiated by Black in Bob's brain would prevail. But if there is no conflict, the decision to steal Ann's car will be caused by x, Bob's indeterministic decision-making process, which will then prevail over P. But this looks a bit too nicely arranged and again one may wonder why in the first situation P prevails over x while in the second it is the other way around. Mele and Robb provide an answer in the form of a story, where $N2$ is a 'decision node' in Bob's brain whose 'lighting', in being 'hit' by P or x, represents his decision not to steal Ann's car at $t2$ ($N1$'s lighting would represent his decision to steal the car). According to this story, by $t2$ P has already blocked $N2$ without affecting the unfolding of process x. So, if x were to 'hit' $N2$ at $t2$, $N2$ would not light up. More exactly, by $t2$ P has blocked all decision nodes in Bob's brain that are incompatible with a decision at $t2$ to steal Ann's car. Again, of course, this blockage does not cause Bob's decision

at all, since at $t2$ his own indeterministic decision process 'hits' $N1$ and he decides on his own to steal Ann's car.

Now, are we forced to accept the intended conclusion, namely that Bob is morally responsible for deciding to steal Ann's car even if he had no alternatives at all to that decision? As we anticipated, I do not think that our intuitions here speak up as clearly for this conclusion as in classical Frankfurt cases. Though we may accept that P does not cause Bob's decision, still it is *actually*, not merely counterfactually, blocking any alternative to it. How to exclude the possibility that an actual, active blocking mechanism in Bob's brain is not, in any way at all, influencing Bob's decision-making process? Stipulating that it is not does not seem to be enough in this case to dispel our doubts.

But Mele and Robb have another way of presenting his case "from an intuitively appealing perspective", a perspective which, in a note, they acknowledge it was recommended to them by John Martin Fischer. Here it is:

> Subtract Black and P from our scenario and imagine that what happens at Bob's indeterministic world is that x, Bob's indeterministic decision-making process, indeterministically issues at $t2$ —in some way favored by libertarians— in his decision to steal the car. Plainly, there is no deterministic cause of Bob's decision in this case. Now add Black and P to the scenario in just the way we have done. At $t2$, process x issues in the same indeterministic way in Bob's decision: by hypothesis, Black and P do not influence x. Although at $t2$ Bob cannot do otherwise than decide to steal the car, nothing warrants the claim that his decision is deterministically caused. (Mele and Robb 1998, p. 108)

So, the idea seems to be this. We certainly are willing to accept that Bob is morally responsible, if someone ever is, for his decision in the first scenario, where no blocking mechanism is operating. Our intuitions are quite clear in this case. Now, suppose that we refuse, or at least are reluctant, to accept that Bob is morally responsible for his decision to steal Ann's car in the original example, when the blocking mechanism is in

operation. This is the second scenario. Then the challenge is for us to find a *relevant* difference between these two scenarios to justify a difference in our respective judgments about Bob's moral responsibility. The difference has to be relevant, for difference there is: a blocking mechanism in the second case which is absent in the first. But this difference, Mele and Robb clearly think, is not relevant to justify a difference in our respective judgments, for the actual causal history is the same in both cases.

Mele and Robb claim that their example avoids the objections that may affect classical Frankfurt cases: there is no sign showing which or not would introduce some alternative possibilities into the picture; determinism is not question-beggingly assumed; and, as they also insist, the example "is what Widerker [1995, p. 248] calls an 'IRR-situation': there are 'circumstances in which' Bob decides to steal Ann's car that 'make it impossible for him to avoid' deciding to do this but 'in no way bring it about that' he decides to do this" (Mele and Robb 1998, p. 108). So, we may add, what should prevent one from accepting Bob's moral responsibility except perhaps arcane libertarian prejudices?

Derk Pereboom (cf. Pereboom 2001, p. 17) has reacted to a similar two-scenarios challenge, devised around a Hunt-inspired example, claiming that it is not clear that determinism has not been illegitimately assumed in blockage cases. To see that it might be, he constructs another two-scenarios case, this time involving an atom. Imagine a universe correctly described by Epicurean physics, in which all that ultimately exists is atoms and frictionless void. As is known, atoms, according to this physics, do not have completely deterministic trajectories: from time to time they suffer uncaused swerves. Suppose they naturally fall downwards. Now, in the first scenario, a spherical atom falls downward between instants $t1$ and $t2$; though it does not swerve, it could do so. The second scenario is just like the first, except that the atom falls downward through a vertical tube, whose interior is frictionless and fits exactly the atom's size. Pereboom comments on this case as follows:

> One might initially have the intuition that the causal history of the atom from $t1$ to $t2$ in these two situations is in essence the same. However, [...] since the tube prevents any alternative motion, it would seem that it precludes any indeterminism in the atom's causal history from $t1$ to $t2$. And if the tube precludes indeterminism in this causal history, it would appear to make the causal history deterministic. Whether this line of argument is plausible is difficult to ascertain, but it is not obviously implausible [...] My own view is not that actual causal histories in blockage cases are clearly deterministic, but only that these considerations suggest that they may be. (Pereboom 2001, p. 18)

Pereboom's considerations throw serious doubts about whether blockage cases beg the question of determinism and I am sympathetic with them. The problem can be put in other terms if we think of the "fantastic coincidence" Hunt talks about, according to which the agent uses only those neural paths that can actually be used by her. A better explanation of this fact, in that it does not have resource to coincidences, would seem to be that the agent only uses the neural paths that can actually be used by her *because*, the rest of paths being blocked, she cannot do otherwise.

So, there is room for suspecting that blockage cases may be begging the question of determinism. But I think a much stronger case can be made against the blockage strategy by taking into account rational control requirements. I think that these requirements can be made to weigh heavily, in fact decisively, against 'blockage' attempts to reject the necessity of alternative possibilities for moral responsibility. We shall focus on Mele and Robb's example, but the argument generalizes to other versions, such as Hunt's.

First of all, though our objection does not essentially rest on this point, it might seem that the way in which Mele and Robb depict decisions in their story is a bit too simple. They talk about neural 'decision nodes'. The blockage affects all decision nodes incompatible with Bob's decision at $t2$ to steal Ann's car. Decisions, however, and especially decisions to which moral responsibility ascriptions paradigmatically apply, are preceded

by consideration of reasons and based on them. But in Mele and Robb's story, while some decision nodes are blocked, nothing is said about *P*'s blocking the 'reasons pathways', as we may call them, that speak for, and eventually lead to, those decisions. Now imagine the following case. Close to *t*2, when decision nodes incompatible with Bob's decision at *t*2 to steal Ann's car have already been blocked by *P* (remember the story) and Bob's decision to steal Ann's car is imminent and unavoidable, Bob is told by a friend of his that he has put in Ann's car a bomb which will explode if someone tries to get into the car. (One may substitute for this any decisive, overwhelming reason one can think of for Bob not to steal Ann's car.) What happens then to Bob? He clearly sees that he is going to die if he decides to steal Ann's car and carries out this decision, he clearly sees that he has a decisive reason to decide not to steal it, but when he tries to decide not to steal the car, he finds himself unable to do it. Poor Bob, then. He sees how his reasons and his decisions come dramatically apart.

So, this is our story. What it shows is that the blockage is not without consequences for Bob's decision making capacity. It affects his dispositions to decide according to his reasons. Fischer and Ravizza have proposed one way, which they call 'weak reasons-responsiveness', of carving the rationality condition (cf. Fischer and Ravizza 1998, pp. 41–46). But whether or not this way is fully satisfactory, there is little doubt that something like this must be a necessary condition on moral responsibility. Fischer's and Ravizza's idea is roughly as follows: weak reasons-responsiveness holds just in case, keeping constant the agent's actual deliberative and decision-making mechanism, there are some possible scenarios, or possible worlds, in which there is a sufficient reason to decide and do otherwise, she recognizes this reason and she decides and does otherwise. Think of someone who decides to steal a book and does so (the example is Fischer's). Fischer writes:

> If (given the operation of the actual kind of mechanism) he would persist in stealing the book even if he knew that by so acting he would cause himself and his family to be killed, then the

> actual mechanism would seem to be inconsistent with holding him morally responsible for his action. (Fischer 1994, p. 167)

But something very close to this seems to be the case with Bob. And this undermines the judgement that, when the blockage is in operation, he is morally responsible for his decision.

Our example shows that moral responsibility can be undermined, not only by features of the actual causal history of a decision or action, but also by the actual dispositions of the agent to decide and act upon reasons. Even if determinism can be shown not to be question-beggingly assumed, and Bob's actual process of decision-making is not deterministic, it cannot be said to be free, in the sense relevant for moral responsibility, for it issues from impaired, non-reasons-responsive capacities for deliberation and decision. We see that, even if Bob were given an absolutely decisive reason for not deciding in a certain way, he still would decide that way. His practical rationality is, then, seriously impaired and this, for incompatibilists *and* compatibilists alike, undermines an agent's moral responsibility. The considered judgment, then, about Bob's case is that he is not morally responsible for his decision. This may explain our reluctance to accept, in actual blockage cases, unlike counterfactual intervener cases, that the agent is fully morally responsible.

We said above that Mele and Robb's view of decisions looks too simple, in that their connection to reasons is overlooked. So, one may take this connection into account and modify the example so that the blockage does not only affect those 'decision nodes' incompatible with Bob's decision to steal Ann's car at *t*2, but also all corresponding 'reasons pathways' connected to them. Then Bob will become insensitive to any reason that goes against that decision. In this case, were he presented, close to *t*2, with our (or other) decisive reason not to steal Ann's car, Bob would not feel the split between that reason and his decision, for he would not even see the force of this reason. But of course this does not solve the problem concerning his moral responsibility; it rather makes it worse, for the impairement of his deliberation and decision-making capacities extends even

further than in the former case. Now he is not only unable to decide as a decisive reason recommends, but he is even unable to appreciate the force of such a reason.

These considerations suggest an important point, which goes beyond the limits of the issue we are discussing. Notice that, in ruling out any alternatives of decision, the blockage has also affected the agent's capacity for practical reasoning. This strongly suggests that alternative possibilities and practical rationality are not independent of one another. Or, more precisely, that alternative possibilities are an essential aspect of rationality. This deserves further exploration, which, however, cannot be pursued here.

We can now answer Mele and Robb's challenge. There is a difference between the two scenarios they depict which may explain why our judgment about the agent's moral responsibility becomes unstable, to say the less, when blockage is added to the picture. The difference is that practical rationality capacities and dispositions in this latter case are seriously impaired. And this should affect our judgment about Bob's moral responsibility. He is responsible in the first scenario, where no blockage is at work. In the second, however, he is not, or not fully so.

In the light of the preceding arguments, it seems fair to say that the blockage strategy against PAP is a dead end. The counterfactual character of the circumstances that make it impossible that the agent does otherwise seems to be, against Hunt's contention, an essential feature of plausible Frankfurt scenarios. Note that classical Frankfurt cases do not face the rational control problem we raised for blockage cases, for in the former, when the actual mechanism of deliberation and decision operates, the agent is supposed to be adequately sensitive to reasons. In the alternative sequence he may not be, but the mechanism has changed, since the counterfactual factor has taken over. We think, then, that Hunt is wrong when he writes that "the unavoidability essential to a Frankfurt scenario does not have to rest on a counterfactual device" (Hunt 2000, p. 217). We are left, then, with classical Frankfurt cases, which feature a counterfactual factor and which, by their very structure, contain

alternatives of *some* sort. Frankfurt theorists would be better advised, then, to follow Fischer's so-called 'robustness' strategy against PAP: they should accept that there are alternatives in Frankfurt cases, but insist that they are not robust enough to ground moral responsibility.

REFERENCES

Fischer, J.M., 1994, *The Metaphysics of Free Will*, Blackwell, Oxford.
———and M. Ravizza, 1998, *Responsibility and Control. A Theory of Moral Responsibility*, Cambridge University Press, Cambridge.
Frankfurt, H., 1969, "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy*, vol. 66, pp. 829–839.
Hunt, D.P., 2000, "Moral Responsibility and Unavoidable Action", *Philosophical Studies*, 2000, pp. 195–227.
Mele, A.R. and D. Robb, 1998, "Rescuing Frankfurt-Style Cases", *The Philosophical Review*, vol. 107, pp. 97–112.
Pereboom, D., 2001, *Living Without Free Will*, Cambridge University Press, Cambridge.
Widerker, D., 1995, "Libertarianism and Frankfurt's Attack on the Principle of Alternate Possibilities", *The Philosophical Review*, vol. 104, pp. 247–261.

Received July 9, 2003; accepted November 5, 2003.