

A Problem for Frankfurt Examples

Samuel Kahn

Indiana University-Purdue University, Indianapolis

In this paper I intend to raise a problem for so-called Frankfurt examples. I begin by describing the examples and what they are used for. Then I describe the problem.

1. Background

In 1969, Harry Frankfurt published a seminal attack on what he dubbed the Principle of Alternate Possibilities (PAP):

PAP: an agent can be morally responsible for performing some action A only if there was some alternate possible action s/he could have performed (Frankfurt, 1969).

Frankfurt's attack on this principle consists in an example in which there is a counterfactual intervener, someone (or something) waiting at the sidelines should the agent show any sign of doing otherwise than s/he actually does. The idea is that, on the one hand, sincere reflection on the example will elicit the judgment that the agent is responsible for what s/he does but, on the other hand, such reflection also will reveal that the counterfactual intervener makes it impossible for the agent to do otherwise.

For instance, suppose Smith is going to the voting booth.¹ Smith plans to vote for candidate A. But Jones is monitoring Smith's every movement and, in fact, Smith's every thought. Should there be even the slightest indication that Smith is going to vote for candidate B, Jones will intervene and trigger a neural sequence that will cause Smith to vote for candidate A. But there is no such indication that Smith is going to vote for candidate B and, thus, Jones does not intervene. In this case, according to Frankfurt, Smith is responsible for voting for candidate A notwithstanding the fact that Smith cannot do otherwise, whence it follows that the PAP is false.

Frankfurt's attack on the PAP, as those who wrote in its wake have pointed out, is inspired from the work of John Locke (Hunt, 2000). In Locke's example, a man is trapped in a locked room, but the man does not know the room is locked and he stays in the room voluntarily (Locke, 1975, section 2.2.10). According to Locke, the man is responsible for staying in the room even though he cannot do otherwise. On the basis of this, Locke concludes that free will should be analyzed in terms of volition rather than the counterfactual ability to do otherwise. Frankfurt followed suit.

Attacks on Frankfurt have come in three different waves. The first wave of attacks conceded the PAP but took issue with Frankfurt's move from the falsity of the PAP to a compatibilist account of free will (Blumenfeld, 1971 or van Inwagen, 1978). The second wave focused on the metaphysical premise underlying Frankfurt examples, on the very possibility of a counterfactual intervener who renders an agent unable to do otherwise (Kane, 1985, 51n25; Heinaman, 1986; Ginet, 1996; and Widerker, 1996). The third focused on the moral premise underlying these examples, on the idea that these agents are genuinely responsible for their actions (Kahn, 2019, chapter 6 section 2).

Meanwhile those who defend Frankfurt against these attacks have developed new examples to do the trick: Fischer examples, in which an agent is incapacitated should s/he attempt to do something other than what the intervener wants him/her to do (Fischer, 2006); buffering examples, in which an agent can do otherwise but would have to cross some psychological “buffer,” typically involving some sort of reflection, in order to do so (Pereboom, 2005); and more.

The debate thus continues, and this paper is meant to be a part of that: I intend to argue on the side of those who defend the PAP. To be clear, I do not want to offer a positive argument in favor of the PAP. Rather, I want to argue that there is something awry with Frankfurt examples and their derivatives. However, I want to do so in a way that will upset much of the extant debate about the PAP. In particular, I want to argue that the idea that the deontic status of an action can be determined independently of considerations about whether an agent can do otherwise, the idea that there is a moral premise that can be separated cleanly from a metaphysical premise to begin with, is misguided.²

2. The problem

Here, in outline, is my argument:

1. If Frankfurt examples and their derivatives are coherent, then it is possible to come up with such an example for a merely permissible action.

2. It is not possible to come up with such an example for a merely permissible action.
3. Therefore, Frankfurt examples and their derivatives are not coherent.

I shall discuss the two premises of this argument in turn.

Most Frankfurt examples involve an agent who performs a conventionally impermissible action like murder or breaking a promise. The kind of responsibility that is therefore associated with the agent is blame. One reason why this is notable is that some philosophers have suggested that there might be an asymmetry in judgments regarding blameworthy actions and judgments regarding praiseworthy actions (Fischer and Ravizza, 1991). If there is such an asymmetry, then perhaps the PAP is too broad a principle: perhaps we should distinguish between the PAP-blame, which concerns solely moral responsibility for blameworthy actions, and the PAP-praise, which concerns solely moral responsibility for praiseworthy actions. Perhaps other deontic categories, like the merely permissible, also deserve their own version of the PAP. This would undermine the first premise of my argument for it would render *prima facie* implausible an inference from the coherence of Frankfurt examples in general, most of which coincidentally involve blame, to the coherence of Frankfurt examples involving mere permission: in the context of the PAP, a move from Frankfurt examples in general to Frankfurt examples involving mere permission no longer would be a move from genus to species.

A second reason why the fact that most Frankfurt examples involve blame is notable has to do with the principle that ought implies can (OIC). An impermissible action is one that an agent ought not to perform. Thus, if these examples work against the PAP, they also work

against OIC: the agents in the examples are morally responsible for something that they ought not to do even though they are unable to do otherwise (falsifying the PAP), and this presupposes that they ought not to do whatever it is even though they are unable to do otherwise (falsifying OIC).³ This is relevant for current purposes because a Frankfurt example involving an action performing a merely permissible action would have no obvious bearing on OIC.

Now it might be objected that if agents ought not to perform impermissible actions, then agents ought to perform permissible actions. Thus, it might be objected that what I said in the previous paragraph is mistaken, that my project in this paper does have a bearing on OIC. This objection builds on ideas about action types and action tokens. That is, it builds on the idea that if an agent is responsible for performing some action token that is merely permissible, s/he also is responsible for performing the type <a permissible action>, an action type that agents might be said to have a standing duty to instantiate.

I would like to say two things about this. First, I think that these ideas about action types and action tokens require more refinement. For example, it certainly is *not* the case that an agent responsible for performing some action token is responsible for performing all the action types instantiated by that token. This bumps up against the action description problem, difficult territory that I do not want or need to enter into here. Second, this objection overlooks the fact that a Frankfurt example in which an agent is responsible for performing a permissible (or a merely permissible) action is *not* doing something that s/he ought *not* to do, and thus the fact that s/he is unable to do otherwise, if it is a fact, would be irrelevant for OIC: OIC says that an agent *can* do what s/he *ought* to do and thus always is able to avoid behaving impermissibly; but OIC is silent regarding whether an agent can do what s/he ought *not* to do. So I do not think that my project in

this paper has an immediate bearing on OIC, and I think that should be a welcome result insofar as it enables discussion of the PAP to take place without invoking OIC.

But given the potential asymmetry between blame-judgments and praise-judgments, how can I defend the first premise of my argument? I would like to say three things about this.

First, although most Frankfurt examples do involve blameworthy actions, not all do; some involve merely permissible actions. Indeed, both of the examples considered in the previous section, the voting example and the room-staying example, involve merely permissible actions. Perhaps there are some instances, like some of the votes in Nazi Germany, in which voting one way rather than another is obligatory. But not all instances of voting are like that, and there is nothing in the example to suggest that something along these lines is going on. Similarly, it would be possible to imagine a case in which staying in a room is impermissible, perhaps because by doing so the agent is breaking a promise to show up somewhere. But not all instances of staying in a room are like that, and there is nothing in the example to suggest it should be filled out in that manner. So in a way the work for my first premise already has been done for me: there already *are* Frankfurt examples that involve merely permissible actions, and in fact the first example from which all subsequent examples are inspired was so.

Second, I find it *prima facie* implausible that responsibility judgments would splinter so that moral theorists would need to investigate separately the responsibility conditions for each deontic category. Moreover, this *prima facie* implausibility is not based on brute intuition. The PAP creates a simple binary in this dialogue: either alternate possible actions are necessary for moral responsibility or they are not, so it is not as if each deontic category could go in its own bin. And given that Frankfurt examples involving merely permissible actions are already out

there, it seems to me that some positive argument is needed for separating the PAP-mere-permissibility from the PAP and, thus, for denying an inference from the coherence of Frankfurt examples in general to the coherence of Frankfurt examples involving mere permissibility.

Third and finally, should I turn out to be wrong regarding the first premise of my first argument, the second premise of the argument still would show that the PAP-mere-permissibility is true and, thus, the argument still would give an interesting result. Furthermore, if, as I am now going to argue, the second premise is not merely true but analytically true, this would suffice to show that the PAP-mere-permissibility is an analytic truth, which would be an even more interesting result. So let me turn to the defense of premise 2.

Premise 2 says that it is impossible to come up with a Frankfurt example for a merely permissible action. This is a very strong premise, stronger still because the kind of possibility I have in mind is logical possibility. And this might seem initially puzzling given that my defense of premise 1 turns, in part, on the fact that there already have been Frankfurt examples involving merely permissible actions: my own defense of premise 1 seems to undermine premise 2, the premise that I in effect just argued is the more interesting of the two. What is going on?

I would like to say two things about this. First, the conclusion of my argument is that Frankfurt examples in general, not merely those involving merely permissible actions, are incoherent. But obviously I do not want to deny the existence of Frankfurt examples. From this it may be inferred that I do not think there is any absurdity in affirming the existence of X while nonetheless asserting that X is incoherent. There are proofs that $0=1$ and that Santa Claus exists. The proofs rely on fallacious reasoning and/or false premises, and sometimes the errors can be

difficult to spot. The fact that these proofs do not work does not impugn their existence. The point is that if my argument works, then something similar is going on with Frankfurt examples.

Second, I remarked above that premise 2 of my argument is the more interesting of the two because even in the absence of premise 1 it would yield an interesting result, the PAP-mere-permissibility. Now I add that the PAP-mere-permissibility is a more interesting principle than the PAP-blame or the PAP-praise. This is because most of daily life is a matter of the merely permissible. So not only is the PAP-mere-permissibility more relevant for daily life than the PAP-blame or the PAP-praise, but also the PAP-mere-permissibility would suffice on its own to undermine the Lockean/Frankfurtian compatibilist analysis of free will.

The reason premise 2 is correct is that part of what makes an action merely permissible is, by definition, that if an agent is able to perform one merely permissible action, then there must be at least one other permissible action that the agent is able to perform. That is, both the category of the merely permissible and the category of the obligatory are species of the genus category the permissible. And what distinguishes these two categories is that an action is obligatory only if it is the only permissible action that an agent can perform whereas an action is merely permissible only if it is one of a plurality of actions that an agent can perform.

Using this account of the merely permissible, premise 2 can be proved by *reductio*. Suppose that there were a Frankfurt example involving a merely permissible action. Then there would be a scenario in which an agent is able to perform a merely permissible action but unable to do otherwise and, thus, unable to perform any other permissible action. But this straightforwardly contradicts the account of the merely permissible just given, whence it follows that any Frankfurt example involving a merely permissible action must be incoherent.

Now it might be objected that my account of the merely permissible is mistaken. In particular, it might be objected that what distinguishes a merely permissible action from an obligatory action is blame: a merely permissible action can be omitted without blame whereas an obligatory action cannot. This entails that *if* a merely permissible action is omitted, *then* an agent will not be blamed for this omission. It does *not* entail, however, that a merely permissible action *can* be omitted. Thus, the impossibility of a Frankfurt example involving mere permissibility cannot be demonstrated by appeal to the defining marks of mere permission.

But if, as many seem to think, there can be excusing conditions, then this objection cannot be right. That is because excusing conditions are conditions in which an agent infringes on an obligation but is not to blame for this infringement. So if excusing conditions are at least logically possible, then an obligatory action cannot be distinguished from a permissible action by virtue of the fact that an obligatory action cannot be omitted without blame.

However, this response is easily circumvented. The original objection might be retooled: the objector might contend that what distinguishes a merely permissible action from an obligatory action is that if a merely permissible action is omitted, no excusing conditions are needed to explain why the agent is not to blame whereas if an obligatory action is omitted, then either the agent is to blame or there are excusing conditions. This preserves the core of the objection, which is that it is not the case that, necessarily, a merely permissible action can be omitted.

But now the converse problem arises. That is, just as excusing conditions might make the performance of an impermissible action non-blameworthy, so condemning conditions might make the performance of a merely permissible action blameworthy. Here one might imagine a merely permissible action being performed by an agent with a bad heart, or one might construct a

thought experiment involving inverse akrasia. The latter is usually reserved for praiseworthy actions performed under the guise of the bad (e.g., Huck Finn's preserving Jim from the slave-catchers even though he believes that he has a duty to do otherwise), but it would be easy enough to tweak the examples to make them serve the purpose at hand. There are then further problems that arise from the fact that blame seems to come in degrees and as such might not be a suitable kind of mark for distinguishing the merely permissible from the obligatory. To put the point rather bleakly: it seems to me that this line of objection is headed in the wrong direction. Thus it seems to me that my original distinction between the merely permissible and the obligatory stands and, with it, so does premise 2.

3. Conclusion

In this paper I posed a problem for Frankfurt examples. I began by setting out the background, explaining what Frankfurt examples are used for and giving two such examples for illustrative purposes. I then turned to the problem. I argued that if Frankfurt examples in general are coherent, then Frankfurt examples involving merely permissible actions are coherent, but I also argued that Frankfurt examples involving merely permissible actions are not coherent. If these two claims are correct, it follows immediately that Frankfurt examples are, in general, incoherent. If my first claim is incorrect but my second claim stands, then it nonetheless follows that the PAP is true for merely permissible actions, and because most of daily life involves merely permissible actions, this is an interesting result in itself, and it also has broader implications for attempts to give a volitional analysis of free will.

Bibliography

Blumenfeld, David. (1971) The Principle of Alternate Possibilities. *The Journal of Philosophy* Vol. 68 No. 11: 339-345.

Cohen, Yishai. (2016) Leeway Compatibilism and Frankfurt-Style Cases. *Thought* Vol. 5 Issue 2: 89-98.

Fischer, John and Mark Ravizza. (1991) Responsibility and Inevitability. *Ethics* Vol. 101 No. 2: 258-278.

Fischer, John. (2006) Responsibility and Agent-Causation. In David Widerker and Michael McKenna (ed.s), *Moral responsibility and alternative possibilities* (pp. 235-250). England, UK: Ashgate Publishing.

Frankfurt, Harry. (1969) Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy* Vol. 66 No. 23: 829-839.

Ginet, Carl. (1996) In Defense of the Principle of Alternative Possibilities. *Philosophical Perspectives* Vol. 10: 403-417.

Heinaman, Robert. (1986) Incompatibilism without the Principle of Alternative Possibilities. *Australasian journal of philosophy* Vol. 64 No. 3: 266-276.

Hunt, David. (2000) Moral Responsibility and Unavoidable Action. *Philosophical Studies* Vol. 97 No. 2: 195-227.

Kahn, Samuel. (2019) *Kant, Ought Implies Can, the Principle of Alternate Possibilities, and Happiness*. Maryland: Lexington Press.

Kane, Robert. (1985) *Free will and values*. Albany: SUNY Press.

Locke, John. (1975) *An Essay concerning Human Understanding*. Oxford, UK: Oxford University Press.

Pereboom, Derk. (2005) Defending hard incompatibilism. *Midwest studies in philosophy* Vol. 29 Issue 1: 228-247.

van Inwagen, Peter (1978) Ability and Responsibility. *The Philosophical Review* Vol. 87 No. 2: 201-224.

Widerker, David. (1991) Frankfurt on 'Ought Implies Can' and Alternative Possibilities. *Analysis* Vol. 51 No. 4: 222-224.

Widerker, David. (1995) Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities. *The Philosophical Review* Vol. 104 No. 2: 247-261.

Widerker, David. (2006) Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities. In David Widerker and Michael McKenna (ed.s), *Moral responsibility and alternative possibilities* (pp. 53-73). England, UK: Ashgate Publishing.

Yaffe, Gideon. (1999) 'Ought' implies 'can' and the principle of alternate possibilities. *Analysis* Vol. 59 No. 3: 218-222.

¹ Smith and Jones are the names of the agents in Frankfurt's original article. For a Frankfurt example involving voting, see (Cohen, 2016).

² The distinction between the "metaphysical premise" and the "moral premise" is due to (Widerker, 2006).

³ Some, like (Widerker, 1991), argue for a conceptual connection between the PAP and OIC. This connection is contested (Yaffe, 1999), but I am not relying on it above. My assertion that the examples that falsify the PAP also falsify OIC is weaker than an assertion of a conceptual connection and is not impugned by attacks on this connection.