

Frankfurt cases, alternative possibilities and agency as a two-way power

Helen Steward

To cite this article: Helen Steward (2022) Frankfurt cases, alternative possibilities and agency as a two-way power, *Inquiry*, 65:9, 1167-1184, DOI: [10.1080/0020174X.2021.1904639](https://doi.org/10.1080/0020174X.2021.1904639)

To link to this article: <https://doi.org/10.1080/0020174X.2021.1904639>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 30 Mar 2021.



Submit your article to this journal [↗](#)



Article views: 1045



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Frankfurt cases, alternative possibilities and agency as a two-way power

Helen Steward

School of Philosophy, Religion and History of Science, University of Leeds, Leeds, UK

ABSTRACT

In this paper, I argue that having ‘leeway’ is part and parcel of what it is to be the agential source of an action, so that embracing source incompatibilism does not, by itself, absolve the incompatibilist of the need to find Frankfurtian agents to be possessors of alternate possibilities. I offer a response to Frankfurt-style counterexamples to the *Principle of Alternate Possibilities*, based on the idea that Frankfurt’s Jones exercises the two-way power of agency when he acts – a power whose exercise intrinsically implies the possibility of having done otherwise. I then show how to respond to the objection that the alternative possibility noted is not sufficiently ‘robust’ to ground his moral responsibility. I also distinguish my own argument for the claim that source incompatibilism is not truly independent of leeway incompatibilism from an argument for the same conclusion which has been offered previously by Kevin Timpe, and suggest that my own version has the dialectical advantage that it does not automatically assume from the outset that sourcehood requires indeterminism, and hence is in line with the traditional idea that the alternate possibilities requirement on moral responsibility is the common property of compatibilists and incompatibilists alike.



KEYWORDS Free will; moral responsibility; incompatibilism; *Principle of Alternate Possibilities*; Frankfurt cases; two-way power

1. Introduction

Are alternative possibilities required for moral responsibility? Frankfurt (1969) argued that they were not. The *Principle of Alternate Possibilities* was formulated by Frankfurt as follows:

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

and the main claim of his article is that (PAP) is false, shown to be so by a range of counterexamples. If the counterexamples work, then there can

CONTACT Helen Steward  H.Steward@leeds.ac.uk  School of Philosophy, Religion and History of Science, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, UK

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

be agents who could not have done other than they did who nevertheless are morally responsible for what they have done. And with this argument, one might have thought, compatibilism about moral responsibility and determinism ought to be within easier reach. If determinism is a problem for moral responsibility, isn't that at least partly because it might be thought to rule out alternative possibilities?

One influential way of denying that Frankfurt's counterexamples to (PAP) have significant consequences for the compatibilism vs incompatibilism debate has been to argue that the most serious difficulties that determinism presents for moral responsibility and freedom are not really about alternative possibilities (leeway) at all; they are really about the fact that determinism does not permit agents to be properly understood as the *source* of their actions (Pereboom 1995, 2003; Kane 1996; 2002; Clarke 2003). If an action has a cause (or set of causes) which deterministically produce it, and those causes are in turn themselves deterministically caused, and so on, the source of actions seems traceable back well beyond the agent. How then, if determinism is true, can the agent be regarded as anything other than a helpless puppet, caught up in the inevitable machinations of the natural world? How can the agent be the true source of anything that happens? In this paper, I want to acknowledge the power of the intuitions which generate the position which has come to be known as source incompatibilism – but I will argue that it is a mistake to suppose that source incompatibilism is truly a distinct incompatibilist *alternative* to its more traditional cousin, leeway incompatibilism. Rather, I shall claim, a certain kind of leeway is in any case required for sourcehood.

The case for a connection between source incompatibilism and leeway incompatibilism has been made before, by Timpe (2007). While agreeing with Timpe that sourcehood and leeway are connected, I shall try to suggest, however, that Timpe's explanation of that connection is one which necessarily denies alternative possibilities one of their important traditional dialectical roles in the free will debate, as something the existence of which might constitute an independent premise in a potential *argument* for incompatibilism. Roughly speaking, the structure of the kind of argument I have in mind is this

P1 Alternative possibilities are required for free will/moral responsibility.

P2 Alternative possibilities are inconsistent with causal determinism.

So

C Free will/moral responsibility is inconsistent with causal determinism.

In arguments of this traditional kind, notice, P1 is a premise that could in principle be accepted either by compatibilists or incompatibilists – it is supposed to be, in David Hunt’s words, ‘the common property of compatibilists and incompatibilists alike’ (Hunt 2000, 196), although compatibilists who accepted it would of course be likely to give a very different account of what the relevant alternative possibilities amounted to from that given by their incompatibilist opponents. For the purposes of this paper, I shall call this argument the ‘Common Property Argument’ (CPA), in order to highlight the importance of this feature of its first premise. I shall try to show that Timpe’s treatment of the relationship between source and leeway incompatibilism cannot permit alternative possibilities to play their own independent role in an argument of this kind. I shall try to develop the connection between source and leeway incompatibilism, by contrast, in such a way that the alternative possibility requirement can retain this independent role and continue to function as the first premise in a Common Property Argument.

Given that my position entails that alternative possibilities remain necessary for moral responsibility even if one is a source incompatibilist, some response to Frankfurt’s alleged counterexamples is of course required. I shall argue that although Frankfurt may successfully have shown that *one* kind of alternative possibility we might have thought was required for moral responsibility is not in fact needed – the sort that might be thought to be connected with fairness requirements on blameworthiness – there is another important kind which is essential to moral responsibility, because (and only because) it is essential to agency. Such possibilities, I shall maintain, remain available to Frankfurtian agents and help ground our conviction that they are indeed agents of the actions they undertake. This crucial kind of alternative possibility, I shall argue, is intrinsic to the kind of sourcehood that is encoded in the concept of agency. The source incompatibilist is right, therefore, that Frankfurt has not succeeded, by way of his refutation of (PAP), in showing that compatibilism is within our grasp – but wrong if she supposes that Frankfurt *has* nevertheless shown that the availability of alternative possibilities is not crucial to the free and morally responsible agent.

2. The principle of alternative possibilities as a principle of fairness

Why might alternative possibilities be important for moral responsibility? One intuition to this effect – perhaps the main one that has been in play in

the debate that philosophers have had about moral responsibility and its relation to determinism – stems from a thought about *fairness*. The thought is that it is unfair to hold someone morally responsible – and in particular, unfair to blame or punish a person – for something that they could not have avoided doing. This principle may be regarded as something we bring to bear in the practice of everyday moral judgement. If you knock into me on the pavement and hurt me, I won't hold you morally responsible if you've done it only because you were jostled by someone else and were unable to regain your balance; if someone steals food because they're starving, we may decide they were compelled by hunger and consider their act as something for which they are entirely non-blameworthy; if I slap someone while under the influence of hypnotic suggestion, I am unlikely to be held fully accountable for my act. I am not to blame for what I have done, the thought goes, if I couldn't (at any stage)¹ have avoided doing the thing that I have done, since it would then be unfair to hold me responsible. But if there are no alternative possibilities, it might seem that *everything* an agent does is something she couldn't help doing, and thus that holding people morally responsible for what they have done is *always* unfair if determinism is true.

It seems to be true that there is a *prima facie* important principle at stake here, concerning the fairness of holding people morally responsible when they can't do other than they do. But two important questions arise about this principle, in the present context. The first is whether it is really best formulated by means of (PAP) – for if it is not, Frankfurt's verdict that (PAP) is false immediately becomes much less significant. Provided there is *some* principle which (i) appears similarly to be justified by fairness considerations; (ii) specifies that it is a necessary condition of an agent's being morally responsible (or blameworthy) for something that the agent have some kind of capacity to do otherwise of a kind which would be ruled out by determinism; and (iii) resists refutation by Frankfurt cases, it will not be important to the general case for incompatibilism if Frankfurt is right that (PAP) itself is false.² Some philosophers have argued that there are indeed

¹'At any stage', of course, to avoid the generation of counterexamples based on cases in which it might be thought that one is compelled to act in a certain way only because one has knowingly put oneself in that predicament by, for example, taking a drug or subjecting oneself to brainwashing or hypnosis at an earlier stage.

²Cf van Inwagen (1983, 164):

The Principle of Alternate Possibilities is probably false [...]. What follows? It does not follow that we might be morally responsible for our acts even if we lacked free will; it follows only that the usual argument for the proposition that moral responsibility entails free will has a false premise [...].

such principles (Wyma 1997; Otsuka 1998; Widerker 2000).³ For the purposes of this paper, I do not wish to take a stand on the question whether these philosophers are right that there are alternative fairness-based principles which stand up to Frankfurt-style examples better than (PAP), but if there are, it is not clear that it matters terribly if (PAP) itself has been found wanting. Indeed, I do not want here to take a view at all on whether we should endorse any principle concerning fairness and alternative possibilities; in fact, I am quite sympathetic to the idea that so far as the fairness of such things as blame and punishment is concerned, Frankfurt is quite right that alternative possibilities are not relevant at all.⁴ But I shall not pursue that argument here – it has been pursued very effectively by a number of others.⁵ Rather, what I shall be focused on instead is the fact that there is a *second* important question to ask about any fairness-based principle connecting moral responsibility with alternative possibilities, which is whether, however one eventually formulates it, such a principle is the *only* possible basis on which one might try to justify the claim that alternative possibilities are necessary for moral responsibility. And as I argued in Steward (2009), and will now explain again here, the answer to this question is ‘no’.

The idea that only fairness intuitions could support the idea that alternative possibilities are required for moral responsibility often rears its head in the literature on Frankfurt cases in a rather oblique way, via an objection to what are sometimes called ‘flicker’ strategies. According to Fischer (1994, 134), who coined the term, flicker strategies tend to arise from the observation that although Frankfurt cases ‘do not involve alternative possibilities of the normal kind, they nevertheless involve *some* alternative possibilities’. Flicker strategies are very diverse in character, but to give a general idea of what they involve, let us consider a standard kind of Frankfurt case – one in which Jones is being monitored by the powerful Black, who wants Jones to vote for Clinton in the U.S. election.⁶ Should Jones show any sign of thinking about voting for Bush, Black

³For example, Wyma has argued for something he calls the ‘Principle of Possibly Passing the Buck’, which states that ‘a person is morally responsible for something she has done, A, only if she has failed to do something, B, such that doing B would have rendered her morally non-responsible for A’ (Wyma 1997, 59) and Otsuka for the Principle of Avoidable Blame’ which states that ‘one is blameworthy for performing an act of a given type only if one could instead have behaved in a manner for which one would have been entirely blameless’ (Otsuka 1998, 688). Widerker’s principle is the Principle of Alternative Expectations (PAE): ‘An agent S is morally blameworthy for doing A only if in the circumstances it would be morally reasonable to expect S not to have done A’ (Widerker 2000, 192).

⁴Not relevant, that is, independently of the alternate possibilities involved in *agency* the need for which I am about to defend.

⁵Notably by Fischer (1994); Fischer and Ravizza (1998); Sartorio (2016).

⁶This is based on a case described by Fischer (1994).

will immediately intervene to trigger a neurological mechanism which will ensure that Jones will decide to vote for Clinton instead⁷ – and moreover will ensure also that the intention thus formed will persist until the vote has been cast in the way Black wishes. What kind of thing might this ‘sign’ which triggers Black’s intervention be? One kind of imaginary case that has been constructed introduces the supposition that Black is in possession of the knowledge that Jones will reliably blush if he is about to begin considering voting for Bush. On seeing the blush, which is not under Jones’ voluntary control, Black would then intervene to ensure that Jones decides to vote for Clinton instead. Jones’s blushing would obviously trigger an alternative course of events from the one that took place in the actual world – and thus might be argued to represent an alternative possibility which was available to Jones in the Frankfurt scenario. But Fischer insists that this is not an alternative possibility which is relevant to the question whether Jones is morally responsible for voting for Clinton (as he in fact eventually does, without the need for intervention by Black). For, he insists, the alternative possibilities relevant to moral responsibility must surely at least be such that it is *in virtue of their availability* that the agent counts as morally responsible

I suggest that it is not enough for the flicker theorist to analyze the relevant range of cases in such a way as to identify an alternative possibility. Although that is surely a first step, it is not enough to establish the flicker of freedom view, because what needs to be shown is that these alternative possibilities *play a certain role* in the appropriate understanding of the cases. That is, it needs to be shown that these alternative possibilities *ground* our attributions of moral responsibility. And this is what I find puzzling and implausible. (Fischer 1994, 140)

What role is it, exactly, that Fischer thinks alternative possibilities must play if they are to ‘ground’ our attributions of moral responsibility? The idea which has come to figure large in the literature on Frankfurt cases is that the relevant possibilities must be ‘robust’ in the sense that they should be ones by means of freely taking which the agent could have avoided moral responsibility for the actual course of action they took. The thought is well expressed by Pereboom (2001, 1) who claims that alternative possibilities matter because ‘if an agent is to be blameworthy for an action, it seems crucial that she could have done something to avoid being blameworthy – that she could have done something to get

⁷I ignore, for present purposes, the objection that in the case in which Black intervenes, we cannot describe Jones as someone who has ‘decided’ or ‘chosen’ anything. I shall return to objections of this kind which have been pressed by Alvarez (2009) later in the paper.

herself off the hook'. Many others (e.g. Wyma 1997; Otsuka 1998; Widerker 2000) have made similar-sounding suggestions, in confirmatory approval of Fischer's insistence upon finding alternative possibilities available to Frankfurtian agents which are sufficiently 'robust' to sustain moral responsibility judgements. And the availability of possibilities which could have 'got one off the hook' are presumably relevant because it is plausibly *unfair* to blame agents who had no such possibilities available as options.

To my mind, then, the popularity of the 'robustness' defence of Frankfurt-style counterexamples against possible objections confirms that many philosophers take the alternative possibilities requirement to be one which has its basis in thoughts about the unfairness of blaming agents for acts they could not have avoided undertaking. But there is another possibility – we judge Jones responsible when he votes for Clinton of his own free will in part because he satisfies an alternative possibility requirement on responsibility which has quite another source, a source based not on the requirements of fairness, but on the requirements of *agency*. This demand is weaker, in that it does not demand that Jones have been able freely to bring about any alternative positive course of action (such as e.g. voting for Bush). What it demands, rather, is merely that Jones' voting for Clinton have been a genuine action of his – which in turn requires that Jones have been able *not* to perform it.

3. Agency-based intuitions about alternative possibilities

The idea that agency itself is deeply connected to the existence of a certain kind of alternative possibility is both ancient and ubiquitous. Aristotle makes the following claim about 'what depends on men themselves'

[...] what depends on men themselves forms a great portion of contingent matters, and men themselves are the sources of such contingent results. So that it is clear that all the acts of which man is the principle and controller *may either happen or not happen* and that their happening or not happening [...] depends on him. (Aristotle 1984, 1223a2–7)

This idea that all the acts of which human beings are the instigators and controllers may either happen or not happen is, I would claim, a very deeply-rooted thought that it is natural to have about any action. Individual (token) actions are, one might think, the sorts of things that *need not* happen – their coming into existence seems dependent upon the

exercise of a power or powers by the agent of the action which *need* not be exercised by her.

Of course, there will be many compatibilists anxious to insist that to the extent that this idea about the non-necessity of actions must be accepted, only a version of it which is consistent with the truth of determinism need be taken on board. Compatibilists, indeed, have offered their own versions of claims connecting 'free' action to alternative possibilities, which are often conditional in form; Hobbes, for example, claims that 'he is free to do a thing that may do it if he have the will to do it, and may forbear if he have the will to forbear' Hobbes (1999, 16) and Hume (1975, 95) that 'by liberty we can only mean a power of acting or *not acting* according to the determinations of the will; that is, if we choose to rest, we may; if we choose to move, we also may'. In these conditionalising formulations we see the classic compatibilist suggestion that liberty may be consistent with – or even perhaps require – the determination of action by such things as will, desire or choice. But what may require⁸ to be determined by such things as will, desire and choice if we are to own our actions is, I would suggest, merely the *kinds* of actions we perform. What (perhaps) must be thus determined are the things which may be rationalised by will, desire or choice – and these are facts about types of act, not tokens. It may be determined, for example, that I will go to Scotland rather than the Isle of Wight for my holiday because I prefer to go there; or that I will call a friend in need, because I want to help and this is currently my very highest priority. But it does not follow from this that *individual* actions likewise have to be deterministically produced by events and states associated with such things as will, desire and choice. My going to Scotland might be enacted by me in a vast variety of different ways, some of which will themselves, perhaps also be rationalised by will, desire and choice – but many of which will always remain entirely unchosen and unwilling by the agent. It is therefore implausible that we may denote *an individual action* free only if its occurrence is an event determined by the agent's will, desire or choice. That very particular action might not have occurred, for example, had the agent not moved her hands on the steering wheel as she drove to Scotland in quite the exact way that she did; or if she had set off ten minutes later than she

⁸For the purposes of this paper, I do not want to take a view on the question whether it is indeed the case that liberty requires that what I do (in the sense of what *kinds* of things I do) must be determined by such things as will, desire or choice. I only want to insist here that whether or not this is the case, that is not the same as saying that the occurrence of the *individual actions* by means of which we effect our will must similarly be determined.

did; or if there had not been a dead deer in the road which necessitated a lengthy diversion. That all these particulars must be determined by the agent's will, desire or choice if the action is to count as truly the agent's own is hugely implausible. So the compatibilist argument can show, at best, only that an agent may be determined to perform an action of a certain *kind*, quite consistently with that action's genuinely being an action. It cannot show that an agent *As freely only if her actual A-ing is a deterministically produced event*. And so the compatibilist thought that those like Hobbes and Hume wish to respect is, I would argue, consistent with a view according to which the occurrence of individual actions is *not* determined by such things as beliefs, desires, intentions, willings, etc. – and indeed, consistent with a view according to which individual actions are not deterministically produced existents *at all*.

It is one thing, though, to argue that some of the usual compatibilist reasons for thinking that individual actions need to be deterministically produced occurrences if they are to be free are unpersuasive, and quite another to argue that individual actions are *not* in fact determined. It is of course notoriously difficult to show of any given event whether or not it could have failed to occur in the radical sense that even holding fixed all the circumstances just prior to its occurrence, it might nevertheless not have happened. And many will doubtless think that certain general and allegedly scientific views of the Universe imply that most events, at any rate, are such that they are determined to happen by some total set of prior events and circumstances (though it is customary now to note that there might be certain exceptions to this rule, such as the particle emissions by way of which the spontaneous decay of radium material occurs). But for the purposes of this paper, I do not need to show that individual actions are not *in fact* determined occurrences.⁹ I only need the premise that there is, at any rate, a strong inclination to think of them as things which the agent need not have executed and which are therefore such that it is dependent upon her whether they are in fact executed or not. This is because I am not attempting to argue here either for indeterminism or for incompatibilism; I am attempting rather to illustrate the *possibility* of appealing to a purported alternative possibilities requirement on *agency* in order to make possible a distinctive kind of response to Frankfurt-style cases. This response turns on the claim that the reason we are happy to regard Jones as morally responsible for his action is that it is evident that he was its agent; and

⁹Though I do try to argue for this more radical conclusion elsewhere – see Steward (2012).

further, that for this to be so, Jones must have exercised a power to act that he need not have exercised, so that alternative possibilities are involved essentially in the agency-condition which underwrites his responsibility. And all that is required for my purposes here is that I establish the *prima facie* *appeal* of this claim, not that I defend it to the death. Indeed it is important to my purposes in this paper that in principle, we allow the compatibilist the possibility of attempting to accommodate this 'need not have executed' idea within a deterministic picture, much as many traditional compatibilists have attempted to accommodate the 'could have done otherwise' intuition within such a view. Provided it can be agreed that it is a tempting component of an untheoretically adjusted idea of action that they are conceived of as occurrences such that it is up to the agent at the moment of action whether or not they occur, and hence that she *need not* have executed any of her actual individual actions, this will suffice for the argument I would like to offer. It does not matter to the argument whether or not one supposes, in addition, that actions might indeed be such as to conform to this untheoretically adjusted picture.

How might one try to establish the plausibility of the idea that alternative possibilities are essential to agency itself? The idea will have to be that even in order to see something as an (individual) action in the first place, we need to think of it as something that *need not* have happened. Actions must be executed by their agents and execution arguably implies the simultaneously-possessed power of non-execution – otherwise, what we have, it might be said, we do not have agential execution at all, but merely the seamless arising from impersonal forces of an inevitable event. Unless the agent executes the action in conditions in which she need not have thus executed it, we might be inclined to insist, we find it impossible to think of it as her *doing* at all – it is merely something that happens in the world – a mere event and not the action of an agent.

It is however very difficult to say precisely what the alternative possibility is that must be present here. It is not the possibility of freely choosing to refrain from φ -ing, since that might well be lacking in the Frankfurt scenario – Black may well be able to prevent any active choice-event from occurring. It is not the possibility of not φ -ing, for any φ (however narrowly circumscribed), since one can always imagine circumstances in which an agent is the subject of what Frankfurt has called a 'volitional necessity' which dictates that it is completely inconceivable that the agent retain his identity as an agent and yet not φ , because the reasons favouring φ -ing are so powerful in his eyes (one need only

imagine a threat hovering over the agent such that something the agent wants at all cost to avoid will happen if he does not φ (or does not φ at this precise time, in this precise way, etc.). But it is also awkward and not quite right to say that the relevant possibility is the possibility of not performing the individual action one did in fact perform (though this is closer to the truth). One problem is that the power in question seems as though it should be something it is possible to assign to an agent *in advance* of acting – and hence assignable at a moment in time at which the individual action which might later ensue does not even exist yet to be referred to. A second issue is that putting things in this way can make it seem as though the point turns essentially on one's freedom to vary some minor detail of the action (thus ensuring that a *distinct* action might instead have been brought about) – which is not, of course, what crucially matters, however plausible it may be that we do in fact always possess such minor freedoms, even when we are constrained by a volitional necessity to φ . What crucially matters is the idea is that actions are created by the exercise of agential powers which, in being agential, are essentially such as to be powers whose exercise is always metaphysically *optional*, even if the agent's *motivation* dictates in advance that *some* action satisfying a particular description will take place (because the agent will ensure by the exercise of that very optional agential power that it will do so).

I have not found it easy to make the claim I want to make about the kind of power I believe it is plausible to suppose we exercise when we act, but it seems to me that the distinction between one-way and two-way powers can be elucidated in such a way that it may facilitate expression of the intuition at work here (though it might be that the notion of two-way power thus elucidated is not in all respects attuned to that developed within the Aristotelian tradition).¹⁰ Where an inanimate substance exercises one of its powers in circumstances which make that exercise possible, we tend to think that the exercise was not only possible, but also necessary – that nothing else could have happened, given those precise circumstances. When, for example, sodium chloride is added to water and the water dissolves it, this is what we tend to think – that the reaction *had* to happen. The power of water to dissolve sodium chloride is 'one-way', in the sense that only one thing could have arisen from its exercise in those particular circumstances. But when an agent exercises her power of agency in order to bring something about, we tend to

¹⁰Though I have argued elsewhere that seeds of the particular notion of two-way power I myself favour are in fact present in Aristotle; see Steward (2020).

think that she was free both to execute the action and also not to do so. In this case, then, we have an inclination to suppose that possibility does *not* imply necessity.¹¹ Agential power is conceived of as ‘two-way’ in the sense that we think of it as something that may or may not be exercised in the circumstances which make its exercise possible. And so if agency is indeed a two-way power, a certain kind of alternative possibility must be available to the agent of any action at the time of its occurrence – the possibility of not undertaking the exercise of agential power which constituted the particular action which in fact occurred. Unless this possibility exists, one might argue, the ‘action’ cannot really have been an action at all.

Even if Frankfurt were right, therefore, that Jones could not have done otherwise than vote for Clinton in the situation he imagines, it might still be true that the moral responsibility we feel inclined to accord Jones depends on our inclination to think that Jones was the *agent* of his actual voting action – and that our inclination to think that Jones was the agent of his voting action depends in turn on the idea that Jones need not have exerted the agential power he in fact exerted when he actually voted. This strategy of course requires it to be the case that the alternative possible world in which Black intervenes is an alternative world in which Jones did *not* exert his agential power – but surely that is very plausible. One way of arguing the case would be via the suggestion that the action Black produces by means of neurological fiddling is a different token action from the one that actually occurs when Jones votes – but really, it is more in line with the strategy I am suggesting to insist that no action of Jones’s occurs *at all* in the case in which Black intervenes, since Jones could not, by hypothesis, have *refrained* from executing the ‘action’ Black produces.¹² Therefore, on the view being discussed, it is not an action at all – merely an event in which Jones’s brain and body are made to participate by Black’s neurological interventions. Black, and not Jones, is the agential source of what happens.

If the line of thought we have been following is at all plausible, then, the possibility arises that there may be a crucial kind of alternative

¹¹In my experience, many compatibilist philosophers are inclined to deny not only that this intuition could correspond to reality but also that they have the intuition at all. However, there is in fact quite a lot of empirical evidence that the idea may be part of folk psychological thinking. See Nichols (2004) as well the range of suggestive work in developmental psychology about the specialisation inherent in our cognitive processing of animal motion (Gelman 1990; Premack 1990; Leslie 1994; Gelman, Durgin, and Kaufman 1995; Spelke, Phillips, and Woodward 1995).

¹²Cf Alvarez (2009, Abstract): ‘Frankfurt-style cases all require a counterfactual mechanism that *could* cause an agent to perform an action that he cannot avoid performing. I argue that, given our concept of what it is for someone to act, this requirement is inconsistent’.

possibility which we might take to be required for moral responsibility because, and only because we take it to be required for agency. The kind of alternative possibility in question is not the power to choose freely another course of action, which is the only kind of power Fischer considers sufficiently 'robust' to underwrite a principle like PAP. Rather, it is just the power not to have executed the action one did in fact execute – which is a requirement, on this view, of being the agent of that action in the first place, rather than merely a place within which and in the vicinity of which certain events occur and give rise with inevitability to other events. And Frankfurt's counterexamples do not seem to be convincing, construed as counterexamples to a principle of alternative possibilities whose justification lies in the requirements of agency rather than of fairness. Jones *could* have not undertaken the particular exercise of agential power he did in fact undertake (though had he not done so, of course, Black would then have intervened to ensure that Jones brain and body went through the motions of 'deciding' to vote and then 'voting' for Clinton). But these things would not have been Jones' actions. And this, one might argue, is the reason we feel happy to regard him as morally responsible for it, despite the fact that he could not freely have chosen to do something (positive) entirely different instead.

4. Timpe on 'narrow' and 'wide' source incompatibilism

A point somewhat like the one I have just developed is made by Kevin Timpe (2007), in connection with a distinction Timpe develops in that paper between what he calls 'narrow' and 'wide' Source Incompatibilism. The Narrow Source Incompatibilist thinks that 'an agent's being morally responsible for some action A is a matter of the agent's being the proper source of A, and that being the proper kind of source doesn't require alternative possibilities at all' (Timpe 2007, 148).¹³ The Wide Source Incompatibilist, on the other hand, believes that although ultimacy or origination is the most fundamental condition of moral responsibility, alternative possibilities are also required. And one particular variety of Wide Source Incompatibilism involves the idea that 'the most plausible understanding of the ultimacy condition will involve alternative possibilities of some sort' (Timpe 2007, 148). Source incompatibilism and leeway incompatibilism therefore need not be mutually exclusive alternatives, according to Timpe. The source incompatibilist may in fact think

¹³Timpe cites Stump (2003), Zagzebski (2000) and Hunt (2000) as philosophers who take this view.

that leeway incompatibilism follows from her view – and that it does so precisely because the existence of certain alternative possibilities is essential to the agent's *being* the source of an action in the way necessary for the action to count as an action in the first place.

Timpe argues in his paper that the Wide Source Incompatibilist's position that the existence of alternative possibilities is a necessary condition for sourcehood is a plausible one. But why is it plausible? On Timpe's view, it is because alternative possibilities are a necessary precondition for self-determination 'insofar as they mean that the action in question is not causally determined' (Timpe 2007, 149). Timpe here takes from the outset the avowedly incompatibilist view that the presence of any alternative possibilities is a sufficient condition for the falsity of causal determinism, and hence that the need for alternative possibilities follows from the need for the agent to be a true source of what happens in action, because sourcehood *requires* the falsity of causal determinism. But this defence of the alternative possibility requirement already *presupposes* that alternative possibilities are inconsistent with causal determinism. There is no true *derivation* here of an incompatibilism based on the idea of leeway from the sourcehood requirement in Timpe's account – for the incompatibility of alternative possibilities with causal determinism is already presupposed in order to secure the result that source incompatibilism indeed requires alternative possibilities. The alternative possibilities are only required, it turns out, *because their existence is implied by the falsity of causal determinism* – which is itself required by the sourcehood condition, according to Timpe's source incompatibilist. What we do *not* get via Timpe's route between source incompatibilism and leeway incompatibilism is any premise concerning alternative possibilities which might serve to play a role in any kind of Common Property Argument – an argument with a first premise of the sort that might be the common property of compatibilist and incompatibilist alike – and whose justification does not already presuppose the inconsistency of alternative possibilities with causal determinism.

My suggestion is that it would be better to think about the relation between source incompatibilism and leeway incompatibilism in a rather different way from Timpe. My alternative proposal is that the source requirement does indeed imply the need for alternative possibilities, but that that is because being the source of some action requires one, at the very least, to be its agent¹⁴ – and being the agent of some

¹⁴It might also of course require more, depending on one's precise version of source incompatibilism – but I need here only the premise that sourcehood requires agency *at the very least*.

action in turn may be thought to imply that one need not have executed it. Nothing is assumed, in this defence of the need for agential alternative possibilities, about causal determinism – it remains possible, for all that has so far been said, that accommodation of the idea that individual actions are always such that the agent need not have executed them within a deterministic picture might be attainable (just as many compatibilists have always thought that one can accommodate the idea that the agent ‘could have done otherwise’ within a deterministic picture). And so it remains possible that alternative possibilities might figure in an ‘common property argument’ of the sort described at the beginning of this paper – that their necessity to the existence of agency might be agreed upon before the compatibilist vs. incompatibilist arguments even start. This was always the hope of the proponent of the traditional Principle of Alternative Possibilities – that the compatibilist might be brought on board with the premise in *some* sense, and might then be invited to show how the principle could indeed be interpreted in such a way as to be compatible with causal determinism. In my view, it would be dialectically advantageous for any proponent of a leeway condition on agency to attempt to justify the principle without reliance on the idea that the provision of leeway in one’s metaphysics is of course going to rule out causal determinism. Timpe is right that sourcehood and leeway are connected, but it is important to say why that might be in a way that does not already beg too many questions against compatibilism – even if one is oneself an incompatibilist.

5. Conclusion

Where does this leave us with respect to Frankfurt? I believe that Frankfurt did indeed show that the Principle of Alternative Possibilities was false. He did not, however, show that alternative possibilities were not required for moral responsibility, because he did not show that they were not required for *agency* (nor that agency is not required for moral responsibility)¹⁵ – and this possibility makes it tempting to suppose that an alternative version of the Principle of Alternative Possibilities might be constructable, one to which Frankfurt’s Jones, and agents in similar predicaments, do not constitute any kind of counterexample. Jones can be construed as being in possession of the only alternative possibility he really needs – that

¹⁵It would take more argument than there is room to provide here to secure the premise that moral responsibility requires agency, but it strikes me as highly plausible.

implicit in his possession of the two-way power of agency, which implies that he had, at the time of action, the capacity *not* to execute the individual action he in fact executed.

How might this alternative principle be formulated? Unfortunately, it is not at all easy to construct a principle that avoids the ugly and in my view philosophically unsatisfactory need for talk of ‘execution’ or ‘performance’ of actions – philosophically unsatisfactory because of course we do not bear anything like the same relation to our actions as we do to the range of things we normally say we ‘execute’ (plans, strategies) and ‘perform’ (plays, musical works, roles); and because talk of ‘performance’ or ‘execution’ of actions invites the awkward question whether the performance or execution of an action is itself a further action. It is not easy because one needs somehow to achieve, within the formulation of whatever principle one adopts, a focus on the *individual* action which in fact took place – indeed, a focus upon what philosophers have come to call ‘the actual sequence’ – but it is not obvious how to do this without introducing a term for a relation between agent and action. In previous work (Steward 2009), I suggested that the relevant principle might be this:

PPN (Principle of Possible Non-Performance): A person is morally responsible for a particular action only if he could have refrained from performing it.

But I have never been very happy with this formulation, which speaks problematically of the ‘performance’ of individual actions and reifies those actions in a way that is undesirable. It would be better if it were possible somehow to achieve the wanted focus on the actual sequence without the need to single out the particular action it contains.

How might this be done? As mooted above, I suspect the key to the wanted formulation may lie in the idea of the exercise of a two-way power. That the actual sequence is indeed what matters is I think in some ways a correct conclusion from reflection on Frankfurt cases. But in my view, the right way to think about the actual sequence in a case in which it involves agency is as something whose correct characterisation itself involves modality – in that an actual sequence which constitutes the action of an agent needs to involve the exercise of a power which is genuinely *two-way* – and hence implies alternative possibilities. Of course, the alternative possibilities themselves are, precisely, *alternative* – so they cannot themselves be part of any actual sequence description of what goes on. But they can be *implied* by an actual sequence

description of what goes on, if any action must be the exercise of a two-way power. We might, then, say this

PTWP (Principle of Two-Way Powers): A person is morally responsible for what she has done only if in doing it, she exercised a two-way power.

The notion of an exercise of a two-way power might thus hold out the promise of being able to accommodate the idea that what matters is the actual sequence, while simultaneously making available the idea that what happens in the actual sequence depends upon what kinds of powers are exercised there, and hence introduces modality into the essential characterisation of what is nevertheless actual.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Alvarez, Maria. 2009. "Actions, Thought-Experiments and the 'Principle of Alternate Possibilities'." *Australasian Journal of Philosophy* 87: 61–81.
- Aristotle. 1984. "Eudemian Ethics." In *vol. 2 of The Complete Works of Aristotle*, edited by Jonathan Barnes, 1922–1981. Princeton: Princeton University Press.
- Clarke, Randolph. 2003. *Libertarian Accounts of Free Will*. Oxford: Oxford University Press.
- Fischer, John M. 1994. *The Metaphysics of Free Will*. Oxford: Blackwell.
- Fischer, John M., and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, Harry. 1969. "Alternative Possibilities and Moral Responsibility." *Journal of Philosophy* 66: 829–839.
- Gelman, Rochel. 1990. "First Principles Organize Attention to and Learning About Relevant Data: Number and the Animate-Inanimate Distinction as Examples." *Cognitive Science* 14: 79–106.
- Gelman, Rochel, Frank Durgin, and Lisa Kaufman. 1995. "Distinguishing Between Animates and Inanimates: Not by Motion Alone." In *Causal Cognition: A Multidisciplinary Debate*, edited by Dan Sperber, David Premack, and Ann James Premack, 151–184. Oxford: Oxford University Press.
- Hobbes, Thomas. 1999. "Treatise of Liberty and Necessity." In *Hobbes and Bramhall on Liberty and Necessity*, edited by Vere Chappell, 15–42. Cambridge: Cambridge University Press.
- Hume, David. 1975. *Enquiry Concerning Human Understanding*. Ed. Lewis A. Selby-Bigge. Oxford: Oxford University Press.
- Hunt, David. 2000. "Moral Responsibility and Unavoidable Action." *Philosophical Studies* 97: 195–227.
- Kane, Robert. 1996. *The Significance of Free Will*. Oxford: Oxford University Press.

- Kane, Robert. 2002. "Free Will: New Directions for an Ancient Problem." In *Free Will*, edited by Robert Kane, 222–248. Malden: Blackwell Publishers.
- Leslie, Alan M. 1994. "ToMM, ToBY and Agency: Core Architecture and Domain Specificity." In *Mapping the Mind*, edited by Lawrence A. Hirschfeld and Susan A. Gelman, 119–148. Cambridge: Cambridge University Press.
- Nichols, Shaun. 2004. "The Folk Psychology of Free Will: Fits and Starts." *Mind and Language* 19: 473–502.
- Otsuka, Michael. 1998. "Incompatibilism and the Avoidability of Blame." *Ethics* 108: 685–701.
- Pereboom, Derk. 1995. "Determinism al Dente." *Noûs* 29: 21–45.
- Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- Pereboom, Derk. 2003. "Source Incompatibilism and Alternative Possibilities." In *Moral Responsibility and Alternate Possibilities*, edited by David Widerker and Michael McKenna, 185–199. Aldershot: Ashgate.
- Premack, David. 1990. "The Infant's Theory of Self-Propelled Objects." *Cognition* 36: 1–16.
- Sartorio, Carolina. 2016. *Causation and Free Will*. Oxford: Oxford University Press.
- Spelke, Elizabeth S., Ann T. Phillips, and Amanda L. Woodward. 1995. "Infants' Knowledge of Object Motion and Human Action." In *Causal Cognition: A Multidisciplinary Debate*, edited by Dan Sperber, David Premack, and Ann James Premack, 132–161. Oxford: Oxford University Press.
- Steward, Helen. 2009. "Fairness, Agency and the Flicker of Freedom." *Noûs* 43: 64–93.
- Steward, Helen. 2012. *A Metaphysics for Freedom*. Oxford: Oxford University Press.
- Steward, Helen. 2020. "Agency as a two-way Power: A Defence." In *Abilities and Powers: The Monist 103:3*, edited by Helen Beebe and Ann Whittle, 342–355.
- Stump, Eleonore. 2003. "Moral Responsibility Without Alternative Possibilities." In *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, edited by Michael McKenna and David Widerker, 138–158. Aldershot: Ashgate.
- Timpe, Kevin. 2007. "Source Incompatibilism and its Alternatives." *American Philosophical Quarterly* 44 (2): 143–155.
- van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- Widerker, David. 2000. "Frankfurt's Attack on the Principle of Alternate Possibilities: A Further Look." *Philosophical Perspectives* 14: 181–201.
- Wyma, Keith. 1997. "Moral Responsibility and Leeway for Action." *American Philosophical Quarterly* 34: 57–70.
- Zagzebski, Linda. 2000. "Does Libertarian Freedom Require Alternative Possibilities?" *Philosophical Perspectives* 14: 231–248.