Philosophical Studies (2026) 183:531–550
https://doi.org/10.1007/s11098-025-02456-z

Inability, fallibility, and the positive case for PAP

David Storrs-Fox[1]

Received: 26 March 2025 / Accepted: 12 November 2025 / Published online: 4 December 2025

Abstract
According to the Principle of Alternative Possibilities (PAP), someone is morally responsible for something she has done only if she could have done otherwise. Since Harry Frankfurt's seminal article on PAP, the literature has mostly concerned whether Frankfurt-type cases are counterexamples to the principle. There is still no consensus on that. The positive case for PAP has received much less attention. This article addresses a source of support for PAP that appears frequently in the literature, but is rarely discussed at length. This source of support involves cases where it seems someone is not morally responsible because she is unable to do otherwise. I argue that these cases provide poor support for PAP. My argument relies on the thought – developed recently by Alfred Mele, Romy Jaster, Chandra Sripada, and David Storrs Fox – that our abilities are often fallible. The challenge for PAP's adherents is to provide a better positive case for their principle.

Keywords Abilities · Fallibility · Moral responsibility · Principle of alternative possibilities · Reliability

1 Introduction

The Principle of Alternative Possibilities (PAP) says that a person is morally responsible for something she has done only if she could have done otherwise. Since the publication of Frankfurt (1969), the literature on PAP has mostly concerned whether there are successful counterexamples to the principle, along the lines of the examples Frankfurt suggests. There is still no consensus on that (Ekstrom, 2005; Fischer, 2005, 2011). Much less attention has been given to the positive case *for* PAP. My view is

✉ David Storrs-Fox
david.storrs-fox@philosophy.ox.ac.uk; dstorrsfox@gmail.com

[1] University of Oxford (Jesus College, Institute for Ethics in AI), Turl Street, Oxford OX1 3DW, UK

that even if arguments like Frankfurt's do not defeat PAP, the reasons to accept the principle are weak. More than one article is needed to justify that claim fully. Here I focus on *one* source of support for PAP, which involves what I will call *inability cases*. More on those in a moment. First, some setup is required.

On the most common understanding, to be morally responsible for something is to be a proper object of blame, praise, or other "reactive attitudes" in virtue of that thing (Robb, 2020; Strawson, 1962/2008). An agent "could have done otherwise" in the sense relevant to PAP just in case she had the ability to do otherwise: that is, it was in her power to do otherwise. More exactly, an agent could have done otherwise just in case she had the *specific ability* to do otherwise (Fischer, 2018, 220–222; Whittle, 2010; Capes, 2012, 3–4). Suppose Rafael Nadal was in bed sick all day yesterday. It would have been true (in some sense) to say yesterday that Nadal was able to serve faster than 110mph, even though (in another sense) his sickness made him unable to serve at all. As the point is commonly put: Nadal had the *general* ability yesterday to serve faster than 110mph, but not the *specific* ability (i.e., the ability on that specific occasion). In the sense relevant to PAP, he could *not* have served faster than 110mph – and so according to PAP, was not responsible for not doing so.[1]

It is natural to read PAP as requiring that the agent was able *at the time of action* to do otherwise. But as others have pointed out, PAP is implausible read in this way. If I jump into a pit at 8:55am to avoid helping you at 9am, I am unable at 9am to help you. But that inability does not suffice to get me off the hook for not helping you.[2] Similar problems arise over far shorter timescales. Suppose Laura punches Al, and her fist makes contact with his face at some instant t. Ten milliseconds before t, it is already too late for Laura not to punch Al. Her fist has considerable momentum by then, and it would take some time (say, 200 milliseconds[3]) for a decision not to punch Al to affect her movements. Laura is therefore unable at t not to punch Al. However, that inability is not enough to get her off the hook for punching him.

Mele (2006, 84–86; 2009, 165) provides a simple approach to this problem. We can understand PAP as requiring that the agent was able at *or before* the time of action to do otherwise. I was able at 8:55am to help you at 9am, and Laura was able (say) 300 milliseconds before t not to punch Al. In the sense relevant for PAP, both Laura and I were able to do otherwise than we did.[4] However, it is presumably false that an ability at *any* past time to do otherwise is enough to satisfy PAP's require-

[1] For comments on earlier drafts, thanks to Jonathan Baker, Justin Capes, Cian Dorr, John Martin Fischer, Ishtiyaque Haji, Arden Koehler, Andrew Law, Chanwoo Lee, Mallory Medeiros, Yining Nie, Derk Pereboom, Jim Pryor, and David Velleman, as well as audiences at New York University, Seoul National University, the University of California, Riverside, the University of Colorado Boulder, the University of Calgary, the Chinese Academy of Sciences, and the American Philosophical Association Pacific Division Meeting (2021). Thanks also to two anonymous reviewers for *Philosophical Studies*.

For the distinction between specific and general or "generic" abilities, see Mandelkern et al. (2017); Maier (2015).

[2] This case is adapted from Ekstrom (2005), 310. See also Mele (2006), 84–86; Mele (2009), 165.

[3] See Schultze-Kraft et al., 2016.

[4] Mele (2006) also discusses a second approach (86–87). Cf. Ekstrom (2000), 210–212. For my purposes, the differences between these approaches do not matter. The important point is that PAP's necessary condition can be satisfied not only by an ability possessed at the very time of action, but also by an ability possessed before the time of action.

ment. Consider a driver who cannot, in the moment, avoid hitting a pedestrian who unexpectedly runs in front of her car. At some earlier time (call it $t_{earlier}$), the driver could have decided not to take that route. She therefore was able at $t_{earlier}$ not to hit the pedestrian later. But PAP's proponent might reasonably deny that this earlier ability to do otherwise satisfies PAP's requirement.

In what follows, I will take PAP to say that someone is morally responsible for something she has done only if there is some *relevant* time, at or before the time of action, at which she could have done otherwise. But why would an earlier time not be relevant? Most obviously, because of the agent's epistemic situation at that time (cf. Ekstrom, 2000, p. 211). At $t_{earlier}$, the driver could not reasonably have foreseen that her taking the ill-fated route would or might (with sufficient likelihood) result in her hitting the pedestrian. More generally, suppose we are assessing someone's moral responsibility for *F*-ing. PAP's proponent can say an earlier time t is irrelevant if the agent could not, at t, reasonably have foreseen that her actions at t would or might (with sufficient likelihood) result in her *F*-ing.[5]

I will leave it to PAP's proponents to specify exactly which times are relevant for their principle. For example, they might amend the epistemic condition above to require *actual foresight* rather than reasonable foreseeability (Miller, 2017). For my purposes, the important point is that times prior to the time of action will often be relevant. As the cases of the pit and the punch show, an inability to do otherwise at the very time of action often should not be sufficient for PAP to count the agent as non-responsible.

There are some arguments for PAP that I will set aside for the present article. Perhaps most prominently, David Copp and others have argued for a version of PAP from the maxim that 'ought' implies 'can', or from closely related claims (Copp, 1997; Widerker, 2003; van Inwagen, 1983, 161–162). More recently, Alvarez (2009) and Steward (2008, 2009) have defended PAP via the claim that an agent has not even *acted* if she could not have done otherwise. I set aside these arguments not because they are unpersuasive (although I think they are), but because they have already been extensively critiqued in the literature on PAP (e.g. Zimmerman, 1993, 1988, 1997; Haji, 2002, 167–176; Capes, 2010, 2012).

The present article focuses on a different source of support for PAP, which involves what I call *inability cases*. Inability cases appear frequently in the literature on PAP, and they plausibly capture an important part of the intuitive case for the principle. However, they have yet to receive sustained critical reflection. This article aims to remedy that. In an inability case, someone is unable to avoid doing something. It seems to many that *because* the agent is unable to avoid doing the thing, she is not morally responsible for doing it. (Alternatively, someone is unable to *do* something – and apparently because of that inability, is not responsible for *not* doing it.) For example: Carl Ginet supposes that someone slips a reaction-slowing drug into a driver's coffee, without the driver's knowledge. As a consequence, the driver is unable to avoid hitting a pedestrian. It seems the driver is not morally responsible for hitting

[5] '*F*' here stands in for a verb phrase that picks out an action type. Miller (2017) provides helpful discussion of similar conditions in the literature on "tracing." Also see Fischer and Tognazzini (2009).

the pedestrian. And that is plausibly so *because* she is unable to do otherwise (Ginet, 2003, p. 89).

It is a common thought in the literature that such examples support PAP (McKenna & Widerker, 2003, 2–3; Widerker, 2006, p. 164; Spencer, 2013, p. 149; Mayr, 2019; Robb, 2020). Philosophers rarely explain in detail how inability cases support PAP, but the basic idea is well expressed by Copp (2003, 265):

> If a person could not have done otherwise than she did, we are strongly inclined to absolve her of blame for what she did … PAP captures the intuitively plausible idea that the inability to do otherwise excuses a person from blame. If we deny PAP, we need to rethink the familiar excuses.

The thought is that PAP provides a good explanation of why agents in inability cases are not morally responsible for what they have done: they are not responsible because they could not have done otherwise. Further, the "familiar excuses" suggest that we intuitively accept that explanation: it is natural to excuse the driver for hitting the pedestrian by citing her inability to do otherwise. PAP gains support (the thought goes) because it makes good sense of such cases.

The present article seeks to undermine that support for PAP, by appealing to the claim that (specific) abilities are often *fallible* to some degree (Mele, 2003, 2017; Jaster, 2020; Sripada, 2018, 2019; Storrs-Fox, 2025).[6] A golfer might have the ability on some occasion to sink a short putt, yet still miss it (Austin, 1979). I might be able to take a sip of my coffee now, yet pour it down my shirt instead. These abilities are not sure-fire: possessing them does not guarantee successful performance. That is, they are fallible.

It is plausible that *most* of our abilities are fallible. Nearly everything we do (including mental actions) is implemented by an extremely complex sequence of cognitive steps. Many of these steps are carried out by subpersonal systems over which we have little control, and which are subject to error. Errors might arise because of the stochasticity of neural processes, the misallocation of limited cognitive resources, or other factors (Sripada, 2018, 12–14; Redish et al., 2008). Such subpersonal errors can cause higher-level tasks to go awry. For example, a subpersonal error could cause a tennis player to hit a ball slightly too hard and miss a serve (Sripada, 2018, 9–14).

Importantly, I am claiming that *specific* abilities are often fallible. The golfer might miss the putt even though she had the ability on this specific occasion. As J.L. Austin says, imagining himself as the golfer:

> It is not that I should have holed it if conditions had been different: that might of course be so, but I am talking about conditions as they precisely were, and asserting that I could have holed it. (Austin, 1979, fn. to 218)

Some authors reject Austin's judgment that he had the specific ability to hole the putt, saying instead that he merely had the *general* ability to hole putts on similar occa-

[6] 'Fallibility' is Sripada's term. Mele and Jaster discuss the same phenomenon using the term 'reliability.' An ability is fallible just in case it is (in Mele and Jaster's terminology) not perfectly reliable.

sions (Berofsky, 2011, p. 155; Berofsky, 2012, 75–76; White, 1968, 87–89). Perhaps these authors would deny that specific abilities are ever fallible. However, many others accept Austin's judgment concerning the golf case (Mele, 2003, p. 449; 2017, 65–66; Vihvelin, 2004; Fara, 2008, p. 847, 862; Thalberg, 2004, 126–129; Alvarez, 2013; Jaster, 2020, 54–55). Moreover, a number of recent authors agree that specific abilities are often fallible (Mele, 2003, 2017, 63–89; Jaster, 2020, 54–55; Sripada, 2018). Instead of seeking to resolve the dispute about cases like Austin's, the present article simply assumes that specific abilities are indeed often fallible. As such, PAP's defenders might respond to my argument by opposing that assumption.

A final point about fallibility: we should distinguish an ability's fallibility from the degree of effort required to exercise it. When I am playing darts, my ability to hit the bullseye is highly fallible. But a successful exercise of that ability does not require significant effort, any more than it requires significant effort to roll a six when throwing dice. Conversely, it might require a great deal of effort for you to cook a three-course meal, but your ability to do so need not be especially fallible.

To undermine the support inability cases provide for PAP, I employ what I call *fallibility cases*. In each of these cases, someone is able to do something. Her ability is fallible, but only to a very small degree. Even so, the ability's fallibility is manifested, and the agent does not manage to do the thing. In Sect. 1, I argue that the fallibility cases reveal something that is sufficient to make the agents in those cases non-responsible: they are (as I will say) *impelled* to act as they do. The cases also show that impulsion does not require that the agent be unable to do otherwise. I then argue that agents in the *inability* cases – the ones used to support PAP – are also impelled to act as they do, and that impulsion is sufficient to make the agents in *those* cases non-responsible.

Of course, PAP's adherents can acknowledge that something other than the inability to do otherwise is sufficient to make agents in inability cases non-responsible. I do not claim to present counterexamples to PAP. However, I will argue that the observation about impulsion undermines the *positive support* that inability cases provide for PAP. The point is that PAP is explanatorily redundant in the inability cases. Impulsion is sufficient to make the agent non-responsible, so there is no need also to appeal to the inability to do otherwise.

Section 2 rejects two responses PAP's defenders might make to my argument. Both responses rely on the thought that impulsion and the inability to do otherwise are connected in a way that supports PAP. Section 3 addresses the "familiar excuses" in the inability cases, which seem to appeal to the agent's inability to do otherwise. I argue that we can make sense of these excuses by appealing to impulsion, without any need for PAP. Section 4 concludes by advertising my own view of what impulsion consists in, which fits naturally with so-called "valuational theories" of moral responsibility.

My argument challenges PAP's significant role in the contemporary discussion of moral responsibility. Even if the literature contains no decisive counterexample to PAP, it also (in my view) offers insufficient reason to believe in it. I make part of that case here. The challenge for PAP's adherents is to provide a better positive argument for their principle.

2 Fallibility and impulsion

I will work with four inability cases, which are representative of the cases used to support PAP in the literature:

Coffee: Unbeknownst to Krishani, someone has slipped a reaction-slowing drug into her coffee. Consequently, she is unable to avoid hitting a pedestrian who runs in front of her car. Because of her drug-slowed reactions, she ploughs into him.[7]

Thief: A thief is sprinting off with the jewels. Roger is out jogging and does not intend to interfere. However, as the thief approaches he trips and falls into her, knocking her over. He is unable to avoid knocking her over. The thief is apprehended because of Roger's intervention.[8]

Seizure: Adelle suffers a seizure that causes her to hit Ed with a crowbar. Because of the seizure, she is unable to avoid hitting Ed.[9]

Legs: Both of Timo's legs are broken, making him unable to get up and leave the room. He strains to get up and leave, but his broken legs cause him to stay put. As a consequence:

Version (a) *Timo becomes a witness to a crime, at great risk to himself.*

Version (b) *Timo does not help a child who is in distress in the neighboring room.*[10] *Based on Spencer (2013, 149).*

It seems plausible that Krishani is not morally responsible for hitting the pedestrian, Roger is not morally responsible for knocking over the thief, Adelle is not morally responsible for hitting Ed, and Timo is not morally responsible for staying put. The inability cases appear to support PAP, because it supplies a natural explanation of why the agents are not responsible: they are not responsible because they could not have done otherwise. More specifically: they are not responsible because there is no time relevant for PAP at which they could have done otherwise. Perhaps Krishani could have avoided driving that day, and could thereby have avoided hitting the pedestrian. However, we can suppose that she could not have foreseen (and did not foresee) earlier in the day that the pedestrian would run in front of her. In that case, it seems reasonable for PAP's proponent to count her earlier ability to do otherwise as irrelevant for PAP. In the PAP-relevant sense, she could not have done otherwise. Similar remarks apply to the other inability cases.

Inability cases do not merely support versions of PAP that concern responsibility for actions, rather than omissions (cf. Alvarez, 2009; van Inwagen, 1983, 162–166). For example, PAP appears to explain why Timo is not responsible for staying put. The support the cases provide is also not restricted to versions of PAP that concern specific reactive attitudes (like blame: cf. Copp, 1997, p. 453; Widerker, 2003), rather

[7] Based on Ginet (2003, 75); cf. Spencer, 2013, p. 149).

[8] Ginet (2003)'s quarterback example (75).

[9] This case (names included) is from McKenna and Widerker (2003, 2). There are similar cases involving hypnosis, kleptomania, tumors, and other such conditions (Robb, 2020; Mayr, 2019, p. 107).

[10] Based on Spencer (2013, 149).

than conditions for moral responsibility in general. For example, Roger and (in version (a) Timo are presumably candidates for praise or credit, not blame.

Before proceeding with my argument, it is worth noting two features that inability cases tend to share – features that undergird their support for PAP. Firstly, the factors that render the agent unable to do otherwise also cause her action (or omission). For example, Krishani's ingesting of the drug both makes her unable to avoid hitting the pedestrian, and causes her to hit him. Roger's tripping both makes him unable to avoid knocking over the thief, and causes him to knock her over. This feature matters, because PAP appears not to gain similar support from cases in which the inability-producing factors do *not* cause the action (or omission). Consider a case in which the drug makes Krishani unable to avoid hitting the pedestrian, without causing her to hit him. It is far less clear in this case that Krishani is not responsible for her action. Indeed, the amended case contains the ingredients for a Frankfurt-style (purported) counterexample to PAP. We need only add that Krishani hits the pedestrian for her own malicious reasons, and it will seem to many that Krishani *is* responsible for her action.

I take no stand on whether there are any successful Frankfurt-style counterexamples to PAP. My point is simply that if we amend inability cases by removing the feature described above, they are unlikely to provide convincing support *for* PAP. That is because in the amended cases, it is unclear that the agent lacks moral responsibility.

The second feature inability cases tend to share is that the agent does not act (or omit to act) *on the belief* that she is unable to do otherwise. It is not that Adelle believes she is unable to avoid hitting Ed, and for that reason decides to hit Ed. Rather, her seizure causes the hitting in some other way. McKenna and Widerker (the originators of Adelle's case) do not explicitly mention this feature, but it would take an unnatural reading of the case to suppose that Adelle acts on such a belief.

Moreover, there is good reason for PAP's defenders to ascribe the second feature to inability cases. It might well be that the agent who acts on such a belief is not responsible – plausibly, that will partly depend on whether the belief is justified. However, it is not clear in such cases that *actual* inability to do otherwise makes the agent non-responsible, rather than the agent's acting on the belief that she is unable to do otherwise. After all, it is plausible that agents are not responsible for acting even on the *false* (but justified) belief that they are unable to do otherwise. Suppose I believe that I am unable to catch the thief. That belief is well justified, because I have always been a terrible runner. However, without my knowledge you have laced my tea with a drug that makes me incredibly fast. I stay put in the false belief that I am unable to catch the thief. Plausibly, I am not responsible for not catching her.

None of this shows that PAP is false. Rather, the point is that cases without the second feature do not provide compelling support *for* PAP. That is because it is unclear why we should think actual inability makes the agent non-responsible, as opposed to the agent's acting on the belief (true or not) that she is unable to do otherwise.

In order to undermine the support inability cases provide for PAP, I will employ amended versions of the four cases above. I call the amended versions *fallibility cases*. Here is the first:

Coffee-F: Unbeknownst to Krishani, someone has slipped a reaction-slowing drug into her coffee. As in Coffee, a pedestrian runs in front of her car. The drug

does not remove Krishani's ability to avoid hitting the pedestrian, but it does make that ability slightly more fallible. There is a small risk that the drug will slow her reactions enough to cause her to hit the pedestrian. Unfortunately, that risk is realized. She ploughs into him.

How could the drug produce such a result? Here is one story. As is plausibly the case for most of our actions, suppose there are multiple alternative sequences of subpersonal cognitive steps that could implement Krishani's slamming on the brakes. Many of these sequences differ only minimally, and Krishani has little control over exactly which sequence implements her action. We can suppose the drug would not affect most of these sequences. However, a small proportion would (were they actualized) take significantly longer because of the drug, with the result that Krishani would react too slowly to avoid the pedestrian.

I said above (drawing on Sripada, 2018) that a tennis player might hit a ball slightly too hard because of a subpersonal error, and so miss a serve. Nevertheless, the player could have made the serve. Similarly, Krishani might react too slowly because of the subpersonal selection of a drug-affected sequence, and so hit the pedestrian. Nevertheless, she could have avoided hitting the pedestrian. If the drug would affect only a small proportion of the relevant sequences, it would plausibly make her ability fallible only to a small degree.

Here are the other fallibility cases:

Thief-F: A thief is sprinting off with the jewels. Roger is out jogging and does not intend to interfere. Just as the thief approaches, Roger trips and begins to fall. He is able to prevent himself from falling – and so to avoid knocking over the thief – but his ability is (slightly) fallible. As it happens, that fallibility is manifested. He does not manage to prevent the fall. Roger knocks over the thief, and she is apprehended.

Seizure-F: Adelle suffers a seizure, which reduces but does not eliminate her control over her movements. She remains able to avoid hitting Ed with the crowbar, but the seizure makes that ability slightly more fallible. Unfortunately, that fallibility is manifested. The seizure causes her to hit Ed with the crowbar.

Legs-F: Both of Timo's legs are injured. He remains able to leave the room, but his injuries make that ability slightly more fallible. A small error in how he distributes his weight could cause his legs to buckle. In the event, this fallibility is manifested. His injuries cause him to stay put.

The agents in these cases are like Austin's golfer, who misses a short put. They are able do otherwise, but that ability is fallible. In the event, the golfer misses the putt, Roger knocks over the thief, Adelle hits Ed, and Timo stays put. Just as in the inability cases, the agents in the fallibility cases are plausibly not morally responsible for what they have done: Krishani for hitting the pedestrian, Roger for knocking over the thief, Adelle for hitting Ed, Timo for staying put.

Importantly, in the fallibility cases it is *not* because the agents could not have done otherwise that they are non-responsible. After all, they could have done otherwise. Of course, at some point each of these agents *becomes* unable to do otherwise. There is nothing remarkable about that. Once you have done something, you always become unable not to have done it (assuming you were able not to do it in the first place). You cannot change the past. Moreover, you often lose the ability to do otherwise shortly *before* you have done the thing in question. Laura loses the ability not to punch Al

shortly before she punches him. And *any* driver who hits a pedestrian at an instant t loses the ability to avoid the pedestrian shortly before t, given the car's momentum and the driver's reaction time.

In the fallibility cases, the agents initially are able (albeit fallibly) to do otherwise. But once the fallibility manifests, they are no longer able to do otherwise.[11] For example: when the pedestrian first runs in front of the car, Krishani is able to avoid him. At this point, the presence of the drug simply makes her ability to avoid him slightly more fallible. But once the drug-affected subpersonal sequence is selected (instead of one of the many unaffected sequences), she loses the ability to avoid the pedestrian. PAP's proponent might spot an opening. In the sense relevant to PAP (they might say), the agents in the fallibility cases could not have done otherwise. Although Krishani in Coffee-F is able to avoid the pedestrian when he initially runs in front of her car, her ability to do otherwise at *this* time (they might say) is irrelevant for PAP.

However, it is unclear on what basis they could claim that. As I said above, inability at the very time of action is not sufficient for PAP to count someone as non-responsible: earlier times are often relevant too. Unlike their counterparts in the inability cases, each of these agents could have done otherwise on the occasion described. For example, it is not simply that Krishani could have avoided driving that day. She could have avoided the pedestrian even for some interval of time when he was in front of the car. We cannot exclude this interval from relevance by appealing to the epistemic condition mentioned above. Krishani (we can suppose) can and does foresee during that interval that her not braking in time will result in her hitting the pedestrian.[12] It is

[11] A reviewer suggests that both Jaster and Sripada would deny that someone could lose a fallible specific ability when its fallibility manifests. Jaster (2020) sometimes expresses the thought that specific abilities are fallible by saying they can be "masked," in a sense drawn from the dispositions literature. When a disposition is masked, the disposition itself remains. As such (the reviewer wonders), should we not say that when an ability is masked, the ability remains? I acknowledge that Jaster's talk of "masking" invites the reviewer's question, and I do not use that terminology myself. However, Jaster takes a specific ability to be maskable when a failed attempt does not "[show] that the agent *did not have the ability in the first place*" (2020, 55, my emphasis; see also the bullseye example on p. 54). She does not claim that an agent with a fallible ability must retain that ability even once the failed attempt has occurred. Rather, the claim is that an agent might *have had* the ability to *F* even if (as things turned out) she did not manage to *F*.

Concerning Sripada, the reviewer points out that he takes fallibility to involve a diminishment rather than a loss of control. If that is right (the reviewer wonders), how could fallibility manifest in loss of control? In response, consider Sripada (2018)'s central illustration of fallibility. In the Stroop Task, a subject might (for example) be presented with the word 'red' in green text. To succeed, they must respond with the color of the text (green) rather than the color to which the word refers (red) (Sripada, 2018, 12–13). Sripada says that a typical person "has very high [but not perfect] levels of control over producing the right response on each individual trial" (2018, 17). He does *not* say that such a person retains her control in an individual trial *even once* she has selected the incorrect response. That would be an odd claim. I have substantial control over whether I respond with the correct color to the next question. However, once I have answered incorrectly, I surely *no longer* have control over whether I answered that question correctly. I cannot change the past.

[12] A reviewer suggests that this interval might count as irrelevant for PAP because during this interval, Krishani does not know that she has ingested the drug. However, some story is needed about why this ignorance would make the earlier time irrelevant for PAP. It does not prevent her from foreseeing that she will hit the pedestrian if she does not brake in time (which she is able, albeit fallibly, to do). Moreover: at *most* times we are ignorant of many of the subpersonal factors that might cause our actions to go awry, but I take it these times are not for that reason irrelevant for PAP.

hard to see why her ability in this interval would be irrelevant for PAP. Similar comments apply to the other fallibility cases.

But suppose PAP's proponent insists that in the fallibility cases, the agents' last-moment inability to do otherwise *is* sufficient for PAP to count them as non-responsible. They then face the following challenge. They need a way to count a last-moment inability to do otherwise as sufficient for non-responsibility in *these* cases, without counting such a last-moment inability as sufficient for non-responsibility in *other* cases, like the one in which Laura punches Al. I am skeptical that they can meet this challenge. But even if they can, meeting the challenge will require a more precise account of what makes a time relevant or irrelevant for PAP. My argument will then have played a constructive role, by helping to precisify the principle.

In my view, PAP's proponent should concede that the agents in the fallibility cases are relevantly able to do otherwise. But of course, these agents are not at any time *infallibly* able to do otherwise. As such, we might wonder whether PAP should be understood (or strengthened) to say a person is morally responsible for something she has done only if she was *infallibly* able to do otherwise. However, it is implausible that a tiny degree of fallibility in the ability to do otherwise is enough to get someone off the hook. Suppose Amelie suffers a seizure similar to Adelle's in Seizure-F, and her ability to avoid hitting Ed is fallible to the same (possibly very small) degree as Adelle's. But in Amelie's case, that fallibility is not manifested. The seizure does not cause her to hit Ed. Instead, she hits him of her own accord. Amelie is surely morally responsible for hitting Ed even though her ability to do otherwise is fallible, and fallible to the same degree as Adelle's. We can construct similar cases for Coffee-F, Thief-F and Legs-F.

More generally (as I said above), it is plausible that most of our abilities are somewhat fallible, given that even simple actions are implemented by complex and error-prone cognitive sequences. If you are morally responsible for doing something only if you were infallibly able to do otherwise, you are plausibly hardly ever responsible for what you do. In my view, a small degree of fallibility in our abilities does not justify such a revisionist picture.

The agents in the fallibility cases are not responsible for what they do, but that is *not* because of a PAP-relevant inability to do otherwise. So why are they not responsible? In each case, there are causal factors that seem central to the answer: Krishani's drug, Roger's tripping, Adelle's seizure, Timo's injuries. It appears that each of these factors brings about the action (or omission) in such a way that because of that factor, the agent is not morally responsible. I will use the term 'impulsion' (and its verb form 'impel') to denote the relation (or relations) between causal factor and action that is on display in the fallibility cases, in virtue of which the agents in those cases are not responsible. The drug impels Krishani to hit the pedestrian. Roger's tripping impels him to knock over the thief. Adelle's seizure impels her to hit Ed. Timo's injuries impel him to stay put.

In Sect. 4, I will propose that a person's action or omission is impelled just in case it does not manifest what she cares about. A causal factor impels a person's action or omission (on my view) just in case it is partly because of that factor that the action or omission is impelled. However, for present purposes I will not assume this or any other specific account of what impulsion consists in. The important point

is that whatever exactly impulsion is, it appears to be sufficient to make the agent in the fallibility cases non-responsible *even* in the absence of a PAP-relevant inability to do otherwise.[13]

An alternative view to mine is that a person's action or omission is impelled just in case she *tries* to do otherwise, and fails. I doubt that this view will cover all cases of impulsion: for example, Krishani's drug could prevent her even from trying to avoid the pedestrian, and still (even: thereby) impel her to hit him (cf. Chisholm, 1964; Lehrer, 1968). However, I have no need to press that claim here. The important point is that impulsion does not require that the agent be unable to do otherwise. If Krishani does try to avoid the pedestrian, Coffee-F shows that an agent might try and fail to do something *even though* she is able (albeit fallibly) to do it.

Returning to the inability cases, I make two claims. Firstly: if the drug impels Krishani to hit the pedestrian in Coffee-F, it presumably also impels her to hit him in Coffee. The causal story in these cases is very similar: a drug slows her reactions, and thereby causes her to hit the pedestrian. In both cases, the drug appears to make Krishani non-responsible for her action. Similar comments apply to the other inability cases: in all these cases the agent is impelled, just as in the fallibility cases.

Secondly: if impulsion is sufficient to make Krishani non-responsible in Coffee-F, it is presumably also sufficient to make her non-responsible in Coffee. Again, it is unclear why the differences between the two cases would make any difference to the upshot of impulsion for her responsibility. In the sense relevant for PAP, Krishani could have done otherwise in Coffee-F. It therefore seems that impulsion is sufficient to make her non-responsible *whether or not* (in the PAP-relevant sense) she could have done otherwise. Similar comments apply to Thief and Thief-F, Seizure and Seizure-F, and Legs and Legs-F.

These points about impulsion undermine the support inability cases provide for PAP, because it seems the inability to do otherwise is not needed to explain why the agent is not responsible. Impulsion is enough. PAP's proponent might claim that Krishani's non-responsibility in Coffee is overdetermined: that is, impulsion is sufficient to make her non-responsible, but also (independently) so is her inability to do otherwise. But in making that claim, PAP's proponent would admit that the PAP-relevant inability to do otherwise is not required to explain why Krishani in Coffee is not responsible. Inability cases are supposed to provide *positive support* for PAP, not simply to be consistent with it. If there is something present in those cases (i.e., impulsion) that is sufficient to make the agent non-responsible without any PAP-relevant inability to do otherwise, the inability cases do not provide good reason to believe in PAP.

PAP's proponent might respond that we can simply see or intuit that (for example) Krishani's inability to avoid hitting the pedestrian in Coffee makes her non-responsible for hitting him, so we should conclude that her non-responsibility is indeed overdetermined. On this view, our intuitive judgments about the inability cases still

[13] In a way, the fallibility cases reverse the structure of Frankfurt (1969)'s (supposed) counterexamples to PAP. In Frankfurt's cases, some factor makes the agent unable to do otherwise without causing her action. In the fallibility cases, some factor causes the agent's action without making her relevantly unable to do otherwise. Still, that factor appears to make the agent non-responsible in a way that parallels the inability cases.

provide positive support for PAP, even though the inability to do otherwise is not needed to explain the agents' non-responsibility in any of these cases.

Given the extensive disagreement about PAP, it seems likely that judgments will differ on this point. For my part, I grant that there is some initial plausibility to the claim that Krishani in COFFEE is non-responsible because unable to do otherwise. But I lack any clear intuition that this claim is true once I notice the similarity with COFFEE-F, and accept that in the latter case Krishani is relevantly able to do otherwise. I am open to an argument for the claim – and so for the view that Krishani's non-responsibility is overdetermined – but I am not moved simply by the appeal to intuition.

That said, I acknowledge that some readers might have sufficiently clear and compelling intuitions that Krishani is non-responsible because unable to do otherwise, even if they accept that she is also (independently) non-responsible because impelled. These readers might conclude that her non-responsibility is overdetermined. Even if they are right, the support the inability cases provide for PAP would surely be significantly stronger if the inability to do otherwise *were* needed to explain why the agents are not responsible. In admitting that it is not needed, PAP's proponent concedes makes these cases far less useful for supporting her position than they would otherwise be.

3 Impulsion and the inability to do otherwise

Instead of claiming overdetermination, PAP's proponents might claim that impulsion and the inability to do otherwise are connected in such a way that my claims about impulsion still support PAP. The present section rejects two responses to my argument that develop that thought.

Firstly, PAP's defender might claim that someone is impelled to do something only if she *tries her hardest* to do otherwise, and that someone tries her hardest to do otherwise only if she is unable to expend greater effort to do otherwise. This view would threaten my argument if true, because it entails that impulsion requires that the agent be (in some sense) unable to do otherwise – at least, if expending greater effort to do otherwise counts as doing otherwise.

However, nothing in the descriptions of the fallibility cases entails that the agents are unable to expend greater effort to do otherwise. We can suppose (for example) that Krishani is able to expend greater effort to avoid the pedestrian – but *that* ability is fallible, and its fallibility is manifested. For example, the drug might cause her to expend less effort than she is able to. Despite her ability to expend greater effort, her action is impelled. Because someone might be impelled even if she does not try her hardest to do otherwise, the first response on behalf of PAP's proponent fails.

The second response relies on (what I call) the Principle of Non-Impulsion (PNI).

PNI: A person is morally responsible for something she has done only if she was not impelled to do it.

I am not committed to PNI. My argument above requires only that impulsion be sufficient in the fallibility and inability cases to make the agent non-responsible,

rather than in all cases. Nevertheless, PNI is highly plausible. And PAP's proponents could combine PNI with the following principle:

IEI: If a person was unable to do otherwise than she did, then she was impelled to do what she did.

('IEI' stands for 'Inability Entails Impulsion.')

PNI and IEI together entail PAP. PAP's defender could therefore acknowledge that impulsion is sufficient in the inability cases to make the agents non-responsible. But instead of claiming that the agents' responsibility is overdetermined, she could instead say that the sufficiency of impulsion supports PAP via IEI.

However, it is unclear why we should believe in IEI. Indeed, IEI will face potential counterexamples similar to the ones PAP faces. For example:

Coffee-IEI: Unbeknownst to Krishani, someone has slipped a reaction-slowing drug into her coffee. As a consequence, she is unable to avoid hitting a pedestrian who runs in front of her car. However, before ingesting the drug she had vowed to seize any opportunity to kill this person. When she sees him, she decides to act on her vow and ploughs into him. The drug plays no role in causing her action.

Plausibly, Krishani is not impelled to hit the pedestrian. The fallibility cases suggest that a factor impels someone to do something only if it causes her to do it, so it seems the drug does not impel Krishani. And it is unclear which other factor(s) in Coffee-IEI would impel her.

Coffee-IEI resembles a Frankfurt-style case. We can expect that many of the responses PAP's defenders make to Frankfurt-style cases will apply here. Most obviously, IEI's defenders might insist that *if* the drug does not cause (and so does not impel) Krishani, she must retain some ability to do otherwise. For example, she is able to (begin to) *try* to avoid hitting the pedestrian, even if she was unable to avoid hitting him.[14] In my view, the discussion of putative counterexamples to IEI would unfold similarly to the discussion of Frankfurt-style cases. I will not argue here that IEI is false. Rather, I claim that examples like Coffee-IEI suggest that IEI will likely be contentious. As such, I am pessimistic that PAP's defender will recover the lost support from inability cases by appealing to IEI.

The challenge for PAP's defenders is to explain why the inability cases support their principle – even though impulsion is sufficient to make the agents in those cases non-responsible whether or not they are able to do otherwise. The two responses I have considered here fail to meet that challenge.

4 The familiar excuses

IEI does not express the only possible connection between impulsion and the inability to do otherwise. In fact, the two clearly are connected in the inability cases – but not, in my view, in a way that supports PAP. (Keep in mind that PAP is a universal generalization: it says that *whenever* an agent could not have done otherwise, she is not morally responsible for what she did.) This section uses the connection between

[14] This defense of IEI resembles the "flickers of freedom" defense of PAP. For overviews of the literature on Frankfurt-style cases, see Ekstrom (2005); Fischer (2005); Fischer (2011).

impulsion and the inability to do otherwise to explain what Copp (2003, 265) calls "the familiar excuses." I take these excuses to be utterances that appear to cite someone's inability to do otherwise, in order to argue that she is not morally responsible for something she has done. Consider the following sentences:

(1) Adelle could not have avoided hitting Ed.
(2) Timo lacked the ability to leave the room.
(3) Krishani could not avoid hitting the pedestrian.
(4) Roger was unable to avoid knocking over the thief.

In an appropriate context, an utterance of any of these sentences could be among the "familiar excuses." (It is admittedly unnatural to call an utterance of (4) an "excuse," because Roger is a candidate for praise rather than blame. However, our practice of (apparently) appealing to inability to argue that someone is non-responsible is not limited to cases where they would otherwise be blameworthy.) In my view, the sufficiency of impulsion for non-responsibility in the inability cases is enough to explain utterances like these. No appeal to PAP is needed.

It will help to begin with the fallibility cases, before returning to the inability cases. In each of the fallibility cases, fallibility is connected with impulsion: the factors that impel the agent also make her ability to do otherwise more fallible. For example, in Seizure-F Adelle's seizure impels her to hit Ed, and makes her ability to avoid hitting him more fallible. This connection makes it plausible to claim that if Adelle's ability to avoid hitting Ed were not fallible, she would not have hit him. In that counterfactual scenario, she presumably would not have had the seizure. Moreover, Adelle might reasonably respond to a complaint from Ed by pointing out that her ability was fallible. (An ordinary English speaker might not say 'fallible.' Perhaps Adelle could make the same point by saying, "I didn't have total control over my actions," or, "my ability to avoid hitting you was limited.")

As I said above, it is implausible that an agent is morally responsible for something she has done only if she was *infallibly* able to do otherwise. Still, the connection between fallibility and impulsion in Adelle's case makes it relevant to cite the former when excusing her action. Similar points apply to the other fallibility cases. We need not conclude from such excuses that a fallible ability to do otherwise is sufficient in general to make someone non-responsible.

Now consider the inability cases. In each of those cases, *inability* is connected with impulsion: the factors that impel the agent to act as she does also render her unable to do otherwise. For example, in Seizure Adelle's seizure impels her to hit Ed, and renders her unable to avoid hitting him. This connection makes it plausible to claim that if Adelle had been able to avoid hitting Ed, she would not have hit him. Moreover, Adelle might reasonably respond to Ed's complaint by pointing out that she could not have avoided hitting him.

The comparison with the fallibility cases suggests we need not conclude from such excuses that *whenever* someone was unable to do otherwise, she is not morally responsible for what she did. The connection between inability and impulsion in the inability cases makes it reasonable to utter sentences like (1) to (4) to argue that someone is not responsible – even if PAP is false.

Before moving on from the excuses, it is worth noting an ambiguity in English. There are plausibly true utterances of sentences (3) and (4) *even* for the fallibility cases, where the agents are able to do otherwise. And that is so even though we could also utter these sentences truly, concerning the fallibility cases:

(5) Krishani was able to avoid hitting the pedestrian, but she did not manage to do so.
(6) Roger had the ability to avoid knocking over the thief, but his ability was fallible.
(7) Krishani could have avoided the pedestrian, but she got unlucky.

Consider again the tennis player who misses a serve. Plausibly, he could have made the serve, but (in some sense) he was unable to make it. What is going on here?

The answer lies in work on the semantics of modal terms. Rajesh Bhatt has shown that the English phrase 'was able to' is ambiguous between two readings. One reading can be paraphrased using the phrase 'managed to' (the *managed reading*), and the other using the phrase 'had the ability to' (the *ability reading*) (Bhatt, 1999; cf. Thalberg, 2004, 121–122).[15] To use an example from Bhatt (1999, 79–80), someone might say:

(8) John was able to sit through the Chinese opera.

An utterance of (8) could express that John *managed* to sit through the Chinese opera, which entails that he *did* sit through it. ("We're proud of John: for the first time, he was able to sit through the Chinese opera.") That is the managed reading. Alternatively, an utterance of (8) could express that John simply had the ability to sit through the Chinese opera, which does not entail that he sat through it. ("John was able to sit through the Chinese opera, but he rudely left halfway through.") That is the ability reading.

It is easy to extend Bhatt's point to negated versions of his examples:

(9) John was unable to sit through the Chinese opera.

An utterance of (9) on the managed reading says that John did not manage to sit through the Chinese opera. An utterance of (9) on the ability reading says that John lacked the ability to sit through the Chinese opera. It is harder to distinguish between these readings in the negated sentence (9), because both readings (plausibly) entail that John did *not* sit through the opera. Still, the two readings available for (8) should lead us to expect two corresponding readings for (9).

Fallibility can help to distinguish the two readings of (9). Suppose John has a mild seizure like Adelle's in Seizure-F, which reduces but does not eliminate his control over his movements. He is able to remain seated (and so sit through the opera), but the seizure makes that ability slightly more fallible. Unfortunately, the fallibility is manifested, and the seizure causes him to fall off the chair. In this case, (9) might be uttered truly on the managed reading, or falsely on the ability reading. On the managed reading: John was unable to sit through the Chinese opera, because the seizure

[15] Hacquard (2021) discusses the subsequent linguistics literature on Bhatt's observation.

threw him off his chair. On the ability reading: John was *able* to sit through the Chinese opera. He had the ability to remain seated, but that ability was fallible. Similarly: if Nadal has just missed a serve, an utterance of 'Nadal was unable to make the serve' will have a true managed reading and (usually) a false ability reading.

Return to PAP, and sentences (3) to (7). Firstly, PAP surely involves an ability reading of 'could have done otherwise,' not a managed reading. It is clearly false that someone is morally responsible for doing something only if she managed not to do it. (Indeed, I doubt that a managed reading is even available for utterances of 'could have done [otherwise].' By contrast, a managed reading *is* available for utterances of 'could do [otherwise],' concerning a past action. Similarly, I doubt that a managed reading is available for utterances of sentences (1) and (2).)

Secondly, I claim for Coffee-F and Thief-F (fallibility cases) that the true utterances of (3) and (4) involve managed readings, and the true utterances of (5) to (7) involve ability readings. I will focus on Coffee-F and (3), but similar comments apply to Thief-F and (4). If there were a true ability-reading utterance of (3), it would presumably say Krishani lacked the relevant ability *at or shortly before* the time of her collision with the pedestrian. If that claim were true, it should have been possible to utter (10) truly at or shortly before the collision time:

(10) Krishani is unable to avoid hitting the pedestrian.

But that is not so. We might have uttered (11) truly before the collision:

(11)Krishani will be unable to avoid hitting the pedestrian.

However, the true utterance of (11) surely demands a managed reading: it is naturally paraphrased by (12) rather than (13) (cf. Thalberg, 2004, p. 121).

(12) Krishani will not manage to avoid hitting the pedestrian.
(13) Krishani will lack the ability to avoid hitting the pedestrian.

The true reading of (3) for Coffee-F involves a managed reading. As such, there is surely also a true reading of (3) for Coffee (the inability case) that involves a managed reading. After all, it is true in both cases that Krishani did not manage to avoid hitting the pedestrian.

Return to the excuses. In both Coffee and Coffee-F it is natural to excuse Krishani by saying that she did not manage to avoid hitting the pedestrian – which we could do by uttering (3). Of course, someone might not manage to do something and yet be responsible for not doing it. It seems that what matters is *why* they did not manage. Suppose John does not manage to sit through the Chinese opera, because he finds it unpleasant. He had earlier promised that he would sit through it. He might well be morally responsible for sitting through the opera: the fact that he did not manage does not automatically get him off the hook. By contrast, in both Coffee and Coffee-F it is because of an impelling drug that Krishani does not manage to avoid the pedestrian. Given that context, it seems natural to excuse Krishani (in part) by

uttering (3) on a managed reading. And Krishani might reasonably excuse herself by saying, "I slammed on the brakes, but didn't manage to avoid him."

Bhatt's distinction between managed and ability readings explains why we could truly utter (3) and (4) even in Coffee-F and Thief-F. It also provides an additional resource for PAP's opponents to explain the familiar excuses. For utterances like (3) and (4) on a managed reading, the appearance that they cite the agent's inability to do otherwise is illusory. I conclude that we can make sense of the familiar excuses in the inability cases without appealing to PAP.

5 Conclusion: valuational theories

I have focused on four inability cases, which are representative of those in the literature. However, I do not intend simply to undermine the support *those* inability cases provide for PAP. Rather, my argument provides a recipe for responding to other inability cases PAP's defender might employ.

For any such case, the recipe instructs us to create a parallel fallibility case with the following features. Instead of being unable to do otherwise, the agent in the fallibility case is able to do otherwise – albeit (to a small degree) fallibly so. And that fallibility is manifested. The agent in the fallibility case performs the same action or omission as in the inability case. The causes of the agent's action or omission in the fallibility case are the same as (or as similar as possible to) those in the inability case. We can then use the fallibility case to argue that the inability to do otherwise is not needed to explain why the agent in the inability case is non-responsible, as I argued above.

I will close with a brief advertisement for my own views about impulsion. In my view, a person's action or omission is impelled just in case it does not manifest what she cares about. A causal factor impels a person's action or omission just in case it is (partly) because of that factor that the action or omission is impelled. And an action or omission manifests what the agent cares about only if it is (in part) causally explained by what she cares about. Finally, I believe that if a person is impelled to do something, she is not morally responsible for doing it.

Taken together, these claims fit naturally with *valuational theories* of moral responsibility, according to which (roughly) someone's responsibility for her actions (or omissions) depends on those actions' being caused in the right way by her values or cares (Murray et al., 2019; Doris, 2015; Sripada, 2016, 2019; Smith, 2005). In Coffee and Coffee-F, Krishani's hitting the pedestrian does not (in my view) manifest what she cares about, because what she cares about does not causally explain her action. And it is because of the drug that her action does not manifest what she cares about, because it severs the connection between cares and action. Therefore (on my view), the drug impels Krishani to hit the pedestrian. Similar comments apply to all the inability and fallibility cases. The fallibility cases illustrate that an action might fail to manifest what an agent cares about *even* if she was able to do otherwise.

There is plenty more to say about valuational theories and impulsion, but I will not say it here. This article's chief aim is to undermine the support that inability cases provide for PAP. PAP has occupied considerable attention in contemporary discus-

sions of free will and moral responsibility. Many philosophers deny that Frankfurt-style cases provide counterexamples to the principle. But even if that is so, in my view the positive reasons to accept PAP are weak. The present article forms part of the argument for that claim, by countering the use of inability cases to support PAP. The challenge for PAP's adherents is to present a compelling case for their principle. Without such a case, our attention might be better directed elsewhere.

Declarations

Competing interests I have no competing interests or funding to declare.

References

Alvarez, M. (2009). Actions, thought-experiments and the 'Principle of alternate possibilities'. *Australasian Journal of Philosophy*, *87*, 61–81. https://doi.org/10.1080/00048400802215505

Alvarez, M. (2013). Agency and Two-Way powers. *Proceedings of the Aristotelian Society*, *113*, 101–121. https://doi.org/10.1111/j.1467-9264.2013.00347.x

Austin, J. L. (1979). Ifs and cans. *Philosophical papers* (pp. 205–232). Oxford University Press.

Berofsky, B. (2011). Compatibilism without Frankfurt: Dispositional Analyses of Free Will. In R. Kane (Ed.), *The Oxford Handbook of Free Will (2nd Edition)* (pp. 153–174). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399691.003.0008

Berofsky, B. (2012). *Nature's Challenge to Free Will*. Oxford University Press.

Bhatt, R. (1999). Ability Modals and Their Actuality Entailments. In K. Shahin, S. Blake, & E.-S. Kim (Eds.), *The Proceedings of the Seventeenth West Coast Conference on Formal Linguistics* (pp. 74–87). Stanford: CSLI Publications.

Capes, J. A. (2010). The W-defense. *Philosophical Studies*, *150*, 61–77. https://doi.org/10.1007/s11098-009-9397-3

Capes, J. A. (2012). Action, responsibility and the ability to do otherwise. *Philosophical Studies*, *158*, 1–15. https://doi.org/10.1007/s11098-010-9662-5

Chisholm, R. M. (1964). J.L. Austin's *Philosophical papers*. *Mind*, *73*(289), 1–26.

Copp, D. (1997). Defending the principle of alternate possibilities: Blameworthiness and moral responsibility. *Noûs*, *31*, 441–456.

Copp, D. (2003). Ought' Implies 'Can', Blameworthiness, and the Principle of Alternate Possibilities. In D. Widerker, & M. McKenna (Eds.), *Moral responsibility and alternative possibilities* (pp. 265–300). Ashgate.

Doris, J. M. (2015). *Talking to our selves*. Oxford University Press.

Ekstrom, L. (2000). *Free will : A philosophical study*. Westview.

Ekstrom, L. (2005). Libertarianism and Frankfurt-style Cases. In R. Kane (Ed.), *The Oxford Handbook of Free Will, 1st Edition* (pp. 309–322). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195178548.003.0013

Fara, M. (2008). Masked abilities and compatibilism. *Mind*, *117*, 843–865. https://doi.org/10.1093/mind/fzn07

Fischer, J. M. (2005). Frankfurt-Type examples and Semi-Compatibilism. In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 281–308). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195178548.003.0012

Fischer, J. M. (2011). Frankfurt-Type Examples and Semi-Compatibilism: New Work. In R. Kane (Ed.), *The Oxford Handbook of Free Will, 2nd Edition* (pp. 242–265). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399691.003.0012

Fischer, J. M. (2018). The freedom required for moral responsibility. In D. O. Brink, S. Sauvé, Meyer, & C. Shields (Eds.), *Virtue, Happiness, knowledge* (pp. 216–233). Oxford University Press.

Fischer, J. M., & Tognazzini, N. A. (2009). The truth about tracing. *Noûs*, *43*, 531–536.

Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy*, *66*(23), 829–839.

Ginet, C. (2003). In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing. In D. Widerker, & M. McKenna (Eds.), *Moral responsibility and alternative possibilities* (pp. 75–90). Ashgate.

Hacquard, V. (2021). Actuality entailments. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmerman (Eds.), *The wiley blackwell companion to semantics*. Wiley. https://doi.org/10.1002/9781118788516.sem052

Haji, I. (2002). *Deontic morality and control*. Cambridge University Press.

Jaster, R. (2020). *Agents' Abilities*. De Gruyter.

Lehrer, K. (1968). Cans without Ifs. *Analysis*, *29*(1), 29–32.

Maier, J. (2015). The agentive modalities. *Philosophy and Phenomenological Research*, *90*(1), 113–134. https://doi.org/10.1111/phpr.12038

Mandelkern, M., Schultheis, G., & Boylan, D. (2017). Agentive modals. *Philosophical Review*, *126*(3), 301–343. https://doi.org/10.1215/00318108-3878483

Mayr, E. (2019). Alternative possibilities and asymmetry. *Synthese*, *196*, 105–125. https://doi.org/10.1007/s11229-017-1503-z

McKenna, M., & Widerker, D. (2003). Introduction. In D. Widerker, & M. McKenna (Eds.), *Moral responsibility and alternative possibilities* (pp. 1–16). Ashgate.

Mele, A. R. (2003). Agents' abilities. *Noûs*, *37*, 447–470.

Mele, A. R. (2006). *Free will and luck*. Oxford University Press.

Mele, A. R. (2009). Moral responsibility and agents' histories. *Philosophical Studies*, *142*, 161–181.

Mele, A. R. (2017). *Aspects of agency*. Oxford University Press.

Miller, D. J. (2017). Reasonable foreseeability and blameless ignorance. *Philosophical Studies*, *174*, 1561–1581.

Murray, S., Murray, E. D., Stewart, G., Sinnott-Armstrong, W., & De Brigard, F. (2019). Responsibility for forgetting. *Philosophical Studies*, *176*, 1177–1201. https://doi.org/10.1007/s11098-018-1053-3

Redish, A. D., Jensen, S., & Johnson, A. (2008). A unified framework for addiction: Vulnerabilities in the decision process. *Behavioral and Brain Sciences*, *31*, 415–487. https://doi.org/10.1017/S0140525X0800472X

Robb, D. (2020). Moral Responsibility and the Principle of Alternative Possibilities. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2020 Edition)*.

Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Görgen, K., Dähne, S., et al. (2016). The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences*, *113*, 1080–1085. https://doi.org/10.1073/pnas.1513569112

Smith, A. M. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics*, *115*, 236–271.

Spencer, J. (2013). What time travelers cannot *not* do (but are responsible for anyway). *Philosophical Studies*, *166*, 149–162. https://doi.org/10.1007/s11098-012-0029-y

Sripada, C. (2016). Self-expression: A deep self theory of moral responsibility. *Philosophical Studies*, *173*, 1203–1232. https://doi.org/10.1007/s11098-015-0527-9

Sripada, C. (2018). Addiction and fallibility. *Journal of Philosophy*, *115*, 1–19.

Sripada, C. (2019). The fallibility paradox. *Social Philosophy & Policy*, *36*(1), 234–248. https://doi.org/10.1017/S0265052519000293

Steward, H. (2008). Moral responsibility and the irrelevance of physics: Fischer's Semi-Compatibilism vs. Anti-Fundamentalism. *The Journal of Ethics*, *12*, 129–145. https://doi.org/10.1007/sl0892-008-9027-x

Steward, H. (2009). Fairness, agency and the flicker of freedom. *Noûs*, *43*, 64–93.

Storrs-Fox, D. (2025). Graded Abilities and Action Fragility. *Erkenntnis, 90*.

Strawson, P. F. (1962/2008). Freedom and resentment. *Freedom and resentment and other essays* (pp. 1–28). Routledge.
Thalberg, I. (2004). *Enigmas of agency*. Routledge.
van Inwagen, P. (1983). *An essay on free will*. Clarendon.
Vihvelin, K. (2004). Free will demystified: A dispositional account. *Philosophical Topics*, *32*(1&2), 427–450.
White, M. (1968). On what could have happened. *The Philosophical Review*, *77*(1), 73–89.
Whittle, A. (2010). Dispositional Abilities. Philosophers' Imprint, *10*(12), 1–23. http://hdl.handle.net/2027/spo.3521354.0010.012
Widerker, D. (2003). In D. Widerker, & M. McKenna (Eds.), *Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities* (pp. 53–74). Ashgate.
Widerker, D. (2006). Libertarianism and the philosophical significance of Frankfurt scenarios. *The Journal of Philosophy*, *103*, 163–187.
Zimmerman, M. J. (1988). *An essay on moral responsibility*. Rowman & Littlefield.
Zimmerman, M. J. (1993). Obligation, responsibility and alternate possibilities. *Analysis*, *53*, 51–53.
Zimmerman, M. J. (1997). A plea for accuses. *American Philosophical Quarterly*, *34*(2), 229–243.